THE EFFECTS OF THE YUBA METHOD ON THE VOCAL PITCH ACCURACY OF INACCURATE ELEMENTARY SINGERS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI'I IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

MUSIC

MAY 2004

By

Karen Ann Miyamoto

Dissertation Committee:

Barbara McLain, Chairperson
Arthur Harvey
Grant Okamura
Karen Kennedy
Richard Green

DEDICATION

This project is dedicated to the one whose "name is above every name" (Philippians 2:9)-- Jesus Christ, my Lord and Savior. It is through Him that "all things are possible" (Luke 18:27). May all call upon Him. "That if you confess with your mouth, 'Jesus is Lord,' and believe in your heart that God raised him from the dead, you will be saved. For it is with your heart that you believe and are justified, and it is with your mouth that you confess and are saved" (Romans 10: 9-10).

ACKNOWLEDGMENTS

I would like to thank first of all, the members of my committee, Dr. Barbara McLain, Dr. Arthur Harvey, Dr. Karen Kennedy, Grant Okamura, and Dr. Richard Green for their participation and assistance in completing the dissertation. This dissertation would not have been possible without their invaluable contributions.

Dr. Barbara McLain chaired this project in a manner that guided and supported in a very positive way. Her recommendations and assistance in the writing of all phases of the project were invaluable. She has been a driving force to help push this project along. With gratitude I want to acknowledge her generous support and encouragement during my entire doctoral program. She helped to make pursuing the degree a gratifying challenge and experience.

Special thanks goes to Dr. Richard Green for his many hours of assistance concerning the experimental design and analysis for this project. His many years of expertise in this area have helped this paper immeasurably.

Thank you to Dr. Russell Yost, expert statistician, who also spent many hours assisting with the analysis of the data using the SAS statistical analysis program. His willingness to help and to open up his statistical laboratory and share his knowledge is greatly appreciated.

I would like to thank Toru Yuba for his contributions to this project. He made himself available in person, and by correspondence whenever I needed assistance

on his method. He is a truly dedicated individual to the music education profession.

Faculty at the participating school also contributed to this project. Their eagerness to cooperate and to arrange for the participation of the children is gratefully acknowledged. They are truly some of the best. Thank you to principal Patricia Dang, vice-principal Lillian Uyeda, sixth grade teachers Lenn Uyeda, Roy Morigawara, Yuri Toyama, Jan Takasaki, fifth grade teachers Rhonda Masuda, Ellen Nishioka, Cheryl Oumi, Jill Yoshimoto, fourth grade teachers Glenn Fujita, Vera Ing, Christine Kamimura, and Shelly Kaya. Thank you to Mary Pascual and Novena Godinez for allowing me to use their room during the treatment sessions.

Thank you to Shawn Yacavonne, who generously volunteered of his time to translate the Yuba Method into English. His assistance in this matter was invaluable.

Thank you to my brother, Michael Lim, for his mathematical genius. He generously gave advice on mathematical computations whenever it was requested of him.

Thank you to my mother, Sharon Lim, for her constant love and support. For the many sacrifices made to provide music lessons, many musical instruments, encouragement, and positive experiences throughout the years.

Thank you to my father, Adam Lim, for his unending support and encouragement throughout the years and during the course of pursuing this degree.

Special thanks to my beloved husband, Dan T. Miyamoto. This project has indeed become a "family affair." Thanks to his invaluable assistance on the computer and technology without which this degree would not have been possible. Thanks to his loving and unending support.

Thanks to my precious four children, with whom I've been very blessed, for their support for "Mommy." To Colby, Keane, Carissa, and Kristyn. May they grow up knowing the full joys of making music.

Lastly, thanks to the many children who I have had the privilege of teaching and knowing throughout the years. They have all been a great inspiration. Teaching music is a great honor. It is hoped that the results of this study may help other teachers to assist their students so that all students may know and experience the joy of singing.

ABSTRACT

A pretest-posttest design was utilized on a population (N = 320) to determine the effects of the Yuba Method on inaccurate elementary singers. Testing was completed using the *Sona-Speech* software program. Inaccurate singers (N = 168) from the population were divided into three subgroups and a random sample of subjects selected for a treatment group (N = 30) and a control group (N = 30).

The Yuba Method, a treatment program devised by Toru Yuba which specifically targets training of the cricothyroid muscle, was administered to treatment subjects (N = 30) in a single forty-five minute session within a three week period. The effect of treatment was found to be highly significant at the $p < .0001$ significance level. There were also significant differences among groups at the $p = .002$ significance level. Some subgroups benefited from the treatment more than others. The treatment "high" subgroup benefited the most, followed by the treatment "middle" subgroup, and lastly the treatment "low" subgroup.

The Yuba Method was found to be highly successful in treating inaccurate singers in this study. Further research is recommended to determine the permanence of the effects of the Yuba Method and its ability to affect elementary age inaccurate singers across a larger population.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Singing has played an important role in the preservation of cultures and societies. As far as it is known, every human society has music (Titon, 1996). Singing is a basic human skill. Singing has been valued by music educators for its aesthetic effects, the development of reading skills, the development of pitch discrimination, tonal memory, and for the physical and mental health benefits (Joyner, 1969; Rutkowski, 1986). Research suggests that accurate singing may be a motivational factor in music education and conversely, inaccurate singing may be a deterrent to participation in music activities. Greenberg (1970) in discussing the effects of inaccurate singing, concluded that a child that knows something is wrong with his or her singing withdraws from most phases of the music program. The National Standards for Music Education developed by the Music Educators National Conference (1994) in anticipation of the *Goals 2000: Educate America Act*, specify that all children are to be taught to sing. Content Standard One stated that students must experience "Singing, alone and with others, a varied repertoire of music" (26). This implies that all children, and therefore all human beings, have singing potential that must be developed.

Student incidence of inaccurate singing has been reported in various studies (Bentley, 1968; Gould, 1968; Roberts & Davies, 1975; Franklin & Franklin, 1983, 1988; Goetze, 1985; Gackle, 1987; Aaron, 1991; and Cooper, 1993). Differences in reported percentages of inaccurate singers were attributed to the differing criteria and repertoire used to measure pitch deviation and the ages being tested (Bentley, 1968; Gould, 1968; Roberts & Davies, 1975; Franklin & Franklin, 1983, 1988; Goetze, 1985; Gackle, 1987;

Aaron, 1991; and Cooper, 1993). The *National Assessment of Educational Progress* for the years 1971-1972 indicated that fifty percent of the nine-year-olds, forty-five percent of the thirteen-year-olds, thirty-five percent of the seventeen-year-olds, and thirty percent of the adults were unable to sing the song *America* with acceptable pitch. Goetze (1985) reported that seventy percent of her kindergarten, first and third grade subjects were inaccurate singers. Aaron (1991) reported that sixty-nine percent of the fourth, fifth, and sixth graders in his study were inaccurate singers.

The occurrence of the inaccurate singer has been an ongoing problem since the inception of singing in the elementary school music curricula over a century ago (Birge, 1955). Goetze (1985) reported that while inaccurate singers were given special drills in music classes, they were often asked to remain quiet during group singing. The music educators investigated stated that they did not want inaccurate singers to "mar" the class singing and potentially influence accurate singers to form poor singing habits. She also noted that despite new methods, the practice of eliminating inaccurate singers from group singing and performance and having them mouth the words without singing is still in existence today. Goetze (1985) stated that anecdotes in informal conversation with adults, years after their singing experiences in childhood, suggested that they suffered from exclusion and embarrassment and have never recovered an enthusiasm for singing. She recommended that more information is needed to provide teachers with an understanding of singing instruction.

Gordon (1985) discovered that music educators used teaching methods, that were unproven in terms of research and that proven methods remained relatively unused. An example of an unproven method used was seating an inaccurate singer next to an accurate

singer. An example of a method Gordon found to be proven and seldom used, was the use of a vertical keyboard to help visualize higher and lower pitches.

Students who are inaccurate singers often reach adulthood unable to successfully participate in the most basic singing activities expected of the general population. This may result in embarrassment, humiliation, and even societal maladjustment (Yuba, 1998). Left uncorrected, inaccurate singers are denied participation in the most personal form of music expression available to human beings.

It is essential to note that various terminologies have been used over the years to indicate singers who are unable to sing consistently in tune. Some commonly used terms have been: "defective singers" (Culpepper, 1961), "monotones" (Bentley, 1968; Joyner, 1969), "non-singers" (Eikum, 1963), "out-of-tune singers" (Humpherys, 1967; Moss, 1973), "problem singer" (Gould, 1969; Jones, 1979), "singing impaired" (Blount, 1982), "tone-deaf" (Yuba, 1998), "tune deaf" (Kalmus & Fry, 1980), and "uncertain singers" (Cobes, 1970; Franklin & Franklin, 1983, 1988; Jones, 1979; Klemish, 1974) and "untuned singer" (Greenberg, 1970). Each of the above terms carried an author-devised definition.

Joyner (1969) defined the monotone as someone who consistently failed to produce the tonal configuration of a melody in a recognizable manner. Jones (1979) described the uncertain singer as having the inability to "carry a tune." Jones further defined the uncertain singer as referring to students who are problem singers, non-singers, or monotones. There is no extant literature stating which of the above is pedagogically correct. The present study will use the term "inaccurate singer" as did Goetze (1985), Reuter (1956), Richner (1976), and Zwissler (1972). Those studies used the term

"inaccurate singer" to denote that person who sings "off pitch," "who cannot correctly reproduce a given melody at a given pitch level, and a term for "one who sings out-of-tune in both individual and group situations."

PURPOSE OF THE STUDY

The purpose of the study is to determine the effect of the Yuba Method on the vocal pitch accuracy of inaccurate elementary school singers in grades four, five and six.

DELIMITATIONS OF THE STUDY

1. This study will be limited to students in grades four, five, and six in one public elementary school in Honolulu, Hawaii.
2. Data generated is limited to scores on the Pretest and Posttest Singing Stimuli.
3. Pitch will be the only criteria used to measure singing accuracy.

DEFINITION OF TERMS

Accurate Singer. Subjects with a vocal pitch accuracy (VPA) score within a semitone (under 100 cents) as determined by analysis of the *Pretest Singing Stimulus* four criterion pitches utilizing the *Sona Speech Model 3600* software program.

Arytenoid Cartilages. A gristly tissue which resembles three-sided pyramids. Their base presents a surface on which this cartilage rocks back and forth and sideways. The vocal cords are attached to these cartilages to tense the vocal cords, making the different pitches that we sing or speak.

Arytenoid Muscles. One of the two muscle groups responsible for the phenomenon of registration. The group itself is comprised of the transverse arytenoids, the oblique arytenoids, and the cricoarytenoids. When tensed, these are responsible for yielding tone qualities commonly recognized as "chest voice."

Break. An interruption of the tonal flow in vocal singing production that results from any of several functional conditions, such as the following:

1. A developmental discrepancy between the two registers at the point of their juncture;
2. A predominance of chest register in the tonal area above its natural limits;
3. An attempt to sing too loudly in the lower range of either the falsetto or head voice;
4. The collapse of an open-throated resonance adjustment, and
5. The sudden release of constrictor tension which occurs when the throat-opening muscles become operative.

Breath Management. The institution of volutional muscular controls over the respiratory function whose design and purpose is to steady the tone and regulate breath expenditure.

Breath Support. The exercise of direct control over the respiratory muscular system for the purpose of ensuring tonal steadiness and regulating breath expenditure; largely synonymous with breath control. The dominant muscle groups involved in the practice of tonal support are the intercostals, the diaphragm, and the abdominals.

Changing Voice. A voice whose function is being modified by the physiological effects of puberty.

Chest Voice. A pedagogic term, commonly used to refer to the lower and heavier of the two register mechanisms.

Closers of the Vocal Folds. The internal thyroarytenoid (vocalis muscles) and interarytenoid muscles, which participate in approximating the vocal cords, and thereby narrow the glottal slit. These muscles work in conjunction with the cricothyroids and arytenoids, antagonists which hold the vocal cords in balanced tension for phonation.

Cricoid Cartilage. Lies below the thyroid and attaches to the first cartilage of the trachea or breathing tube. Muscles lie between the thyroid and cricoid cartilages which, by contraction, can aid in tensing the vocal cords.

Cricoarytenoid Muscles. Part of the complex muscular system which causes the arytenoids to rotate upon the cricoid.

Cricoid Cartilage. That cartilage which forms the base of the larynx; familiarly known as the "ring" cartilage because of its similarity to a signet ring. The posterior cricoarytenoid muscles originate in the cricoid cartilage, and that cartilage also serves as a point of attachment for the longitudinal fibers of the esophagus.

Cricothyroid Muscle. The muscle that controls the vocal cords. Its main function is to act as a tensor, tilting the thyroid cartilage forward and downward, lengthening the vocal folds and making them thinner, resulting in raised pitch. Its relaxation lowers the pitch.

Cycles Per Second (cps). The number of times a complete sound wave (comprised of an alternating compression and rarefaction phase) occurs within the designated time span. Also the same as Hertz.

Epiglottis. A flap-type valve, which closes the larynx as food passes to the esophagus. A thin, yellowish cartilage at the base of the tongue which folds over the glottis during peristaltic movement.

Expiration. The discharging of inspired breath from the lungs. Physically, breath is expired by the contraction of a complex series of abdominal muscles. Their natural opponents, the inspiratory muscles, contract for inspiration.

External Thyroarytenoid Muscle. One of the muscle bands comprising the vocal folds. The thyroarytenoid originated below the thyroid notch and inserts into the base of the arytenoids cartilages. By drawing the arytenoids cartilages forward toward the thyroid cartilage, the thyroarytenoid muscles shorten the vocal folds. Acting against its antagonist (the cricothyroid muscle), the contracted thyroarytenoid firms and tenses the vocal folds. Acting alone, this contraction shortens the vocal folds.

Extrinsic Laryngeal Muscles. Muscles which extend from the exterior of the larynx to other parts of the body. Many of the extrinsic laryngeal muscles attach to such structures as the hyoid bone, the mandible, the sternum, the clavicle, and the shoulder blades, while others, like the inferior and middle pharyngeal constrictors, form the pharyngeal wall. The extrinsic laryngeal muscles are responsible for controlling the vertical movements of the larynx and shaping the pharyngeal resonating system, and, because the sternothyroids hold synergetically against the contraction of the cricothyroids, they play an important role in determining the ratio of registration.

Falsetto. The upper of the three vocal registers. When only the inner edges of the vocal folds vibrate to create the upper register pitches. Tone in which the cricothyroids are tensed without engaging a counterpull or synergetic "hold" on the part of the arytenoids.

Gain Score. The difference between the *Pretest Singing Stimulus* VPA Score and the *Posttest Singing Stimulus* VPA Score obtained by subtracting the *Posttest Singing Stimulus* VPA Score from the *Pretest Singing Stimulus* VPA Score. Positive numbers reflect an increase in singing accuracy. Negative numbers reflect a decrease in singing accuracy.

Glottis. The space between the vocal cords at the upper part of the larynx.

Head Register. A term used synonymously with head voice to denote falsetto-dominated tone qualities.

Head Voice. Tone qualities produced through the coordinate activity of both register mechanisms, the chest register and the falsetto, but with the falsetto strongly dominant; so named because the sensations of vibration evoked as a result of this physical arrangement appear to be concentrated in the head cavities.

Heavy Mechanism. The chest register.

Hertz. A synonym for one cycle per second (cps).

High Register. The head voice.

Inaccurate Singer. A singer whose VPA (Vocal Pitch Accuracy) score is 100 cents or greater as determined by analysis of the *Pretest Singing Stimulus* four criterion pitches utilizing the *Sona Speech Model 3600* software program.

Inspiration. The act of taking air into the lungs and the animation influencing or guiding of a student by those aspects of personal commitment that move the intellect or emotions.

Involuntary Muscle. A muscle whose contraction is not stimulated by volition or other overt means.

Laryngeal. Having to do with the larynx function.

Larynx. Upper part of the trachea. The musculo-cartilaginous structure, lined with mucous membrane and housing the vocal folds, that forms the upper part of the trachea or windpipe; the vocal generator, or voice box.

Mean Cent Deviation. The *Vocal Pitch Accuracy Score*. The average cent deviation of the four criterion pitches in either the *Pretest Singing Stimulus* or the *Posttest Singing Stimulus* as determined by calculation and measurement using the *Sona-Speech Model 3600* software program.

Monotone. A singer unable to produce a discernable melodic contour when singing.

Mutational Chink. A small, triangular opening between the arytenoids which occurs during phonation.

Nasal Cavity. That portion of the pharynx bounded at either end by the posterior nares and the nostrils, and divided by the septum into two fossae.

Nasopharynx. One of the three divisions of the pharynx. It lies above the soft palate. It begins at the rear nasal opening. It opens below into the oropharynx. The soft palate acts as a valve to open and close the nose and nasopharynx from the oral cavity and oropharynx.

Natural Voice. Also referred to as the "chest voice." The lower of the three vocal registers and when the vocal folds vibrate to the full width and length to form the lower or chest voice.

Octave Labeling System. The octave labeling system for this study will use the system where middle C is C4, or the 28th key on the piano keyboard, having the frequency of 261.63 hertz. C above middle C would subsequently be called C5.

Oral Cavity. That cavity commonly known as the mouth. The mouth is bounded anteriorly by the teeth and lips, laterally by the interior walls of the cheeks, and posteriorly by the constricted aperture, or isthmus, formed by the pillars of the fauces, that leads into the oropharynx. Its roof is formed by the hard and soft palates, and its floor, by the tongue muscles and mucous membranes.

Oropharynx. The space between the soft palate and the epiglottis. It has three openings: (1) An opening to the nasopharynx above; (2) An opening in the oral cavity in front (mouth); and (3) An opening to the hypopharynx below.

Pharynx. The common passageway for both air and food from the nose and mouth to the esophagus and larynx. It has three divisions: the nasopharynx, oropharynx, and hypopharynx.

Posttest Singing Stimulus. The singing test given at the end of the experimental treatment period. Used against the *Pretest Singing Stimulus* to calculate *Vocal Pitch Accuracy* gain scores.

Pretest Singing Stimulus. The singing test given at the beginning of the study to classify accurate and inaccurate singers. Used against the *Posttest Singing Stimulus* to determine *Vocal Pitch Accuracy* gain scores.

Register. A group of like sounds or tone qualities whose origin can be traced to a special kind of mechanical (muscular) action.

Register Break. The point of separation between two register mechanisms.

Respiration. The life-sustaining physical process in which an organism takes in oxygen, distributes and utilized it in oxidation, and releases carbon dioxide or other waste.

Soft Palate. Acts as a valve to open and close the nose and nasopharynx from the oral cavity and Oropharynx. The soft palate is a muscular extension of the hard palate.

Sona-Speech. A software-only product for the personal computer. It is the same software used in the *Visi-Pitch III* software/hardware system but can be utilized on a laptop computer. Applications for the software include: visual feedback in therapy, articulation training, voice disorders such as excessive muscular tension in the head and neck regions, inappropriate amplitude levels caused by head trauma, hearing impairment, and esophageal speakers, or treatment via the establishment of an "optimal" speaking pitch, correction of situations in which the client speaks without taking a full breath of air and thus cannot produce efficient voicing, correction of inappropriate voice onset or glottal attach resulting in edema, contact ulcers, or vocal nodules, correction of the glottal attack in instances of unilateral vocal cord paralysis, fluency, English as a second language, and motor speech disorders such as dysarthria, cerebral palsy, neurogenic disorders (such as Parkinson's or multiple sclerosis), and head trauma. Instead of using the hardware of *Visi-Pitch III*, *Sona-Speech* relies on standard sound cards for data acquisition and playback. Real-time visual feedback of important speech/voice parameters and quantitative measurements to track subject performance are unique features. *Sona-Speech* extracts acoustic parameters (e.g., pitch, amplitude, and spectral characteristics) during speech/voice production and presents these in real time, providing clear, intuitive visual displays.

A key program of *Sona-Speech* is Real-Time Pitch, which displays fundamental frequency and relative intensity in real time. Stress, timing, and intonational patterns, as well as target pitch and/or amplitude values during running speech, can be seen as they

are said by the subject. This is the component used in analysis of the *Pretest Singing Stimulus* and *Posttest Singing Stimulus*. Auditory Feedback Tools is a module included for critical listening feedback (using headphones) and various types of auditory feedback for achieving desired speech/voice behaviors. Included are high-fidelity amplification, looped playback of selected speech tokens (e.g., word, phrase, or sentence level), a limited range of Delayed Auditory Feedback, white-noise masking, speech-rate modification (e.g., prolonging an acquired utterance without altering its pitch on playback), and a metronome pacer.

Thyroarytenoid Muscle. One of two complex muscles covered by the vocal folds.

Thyroid Cartilage. The large anterior laryngeal cartilage commonly known as the "Adam's apple."

Tone Deafness. An inability to aurally distinguish differences in pitch. A by-product of this condition may be the inability to sing accurately.

Trachea. The main trunk of a muscular membranous tube that connects the lower part of the larynx with the bronchi and through which air passes into and out of the lungs; known in the vernacular as the "windpipe."

Transverse Arytenoid. A muscles whose contraction approximates the two pyramids to which the arytenoids muscles are attached.

Upper Register. The mechanism responsible for tonal qualities generally recognized as head voice or falsetto.

Vocal Cords. Also called vocal folds. The edges of the folds are of a different texture than the rest of the fold, and that edge is what is called the vocal cords. The lower portion of the thyroarytenoid muscles, which projects into the cavity of the larynx and whose

movement creates the pressure variations responsible for pitch. Theorists now substitute the word "folds" for cords, since the older usage suggests that a parallel exists between the vibratory characteristics of the vocal folds and the strings of the violin; such a suggestion is both untrue and misleading.

Vocal Fry. In phonetics, a rough, squeezed vocalized sound produced by an upward movement of the larynx accompanied by a corresponding lowering of the epiglottis; in singing, a tonal quality resembling a death rattle that is heard only in the lowest pitch range of the voice.

Vocal Pitch Accuracy Score. The average cent deviation of the four criterion pitches in the *Pretest Singing Stimulus* or *Posttest Singing Stimulus* as determined by calculation and measurement on the *Sona-Speech Model 3600* software program.

NULL HYPOTHESIS

The Yuba Method vocal training program will have no effect on the vocal pitch accuracy of inaccurate fourth-grade, fifth-grade, and sixth-grade students.

CHAPTER II

REVIEW OF LITERATURE

THE ACT OF SINGING

The act of singing involves the physiological process of motor coordination and the psychological processes of pitch perception and memory. Singing is a highly complex skill. Sataloff (1998; 22) described voice production as the following:

> The production of speech or song, or even just a vocal sound, entail a complex orchestration of mental and physical actions. The idea for making a sound originates in the cerebral cortex of the brain, in the speech area. The movement of the larynx is controlled from the voice area and is transmitted to the larynx by various nerves. As a result, the vocal folds vibrate, generating a buzzing sound. It is the resonation of that sound throughout the area of the vocal tract above the glottis—an area that includes the pharynx, tongue, palate, oral cavity and nose—that gives the sound the qualities perceived by a listener. Auditory feedback and tactile feedback enable the speaker or singer to achieve fine-tuning of the vocal output (22).

According to Murphy (2002), voice production is dependent on a power source (respiration), a vibrator (larynx), and a resonator (nasal cavity, oral cavity, pharyngeal cavity, and trachea) (3). This chapter will briefly summarize these three main elements of voice production as they relate to the topic of this paper.

Respiration

Murphy (2002) explained that voice production is dependent on a power source, also known as respiration. Appelman (1985) defined respiration as "the will to breathe" (9). Sataloff (1998) said that during voice production, respiration forces air toward the larynx, also known as the vibrator. Sataloff stated that when in an adducted position, the vocal

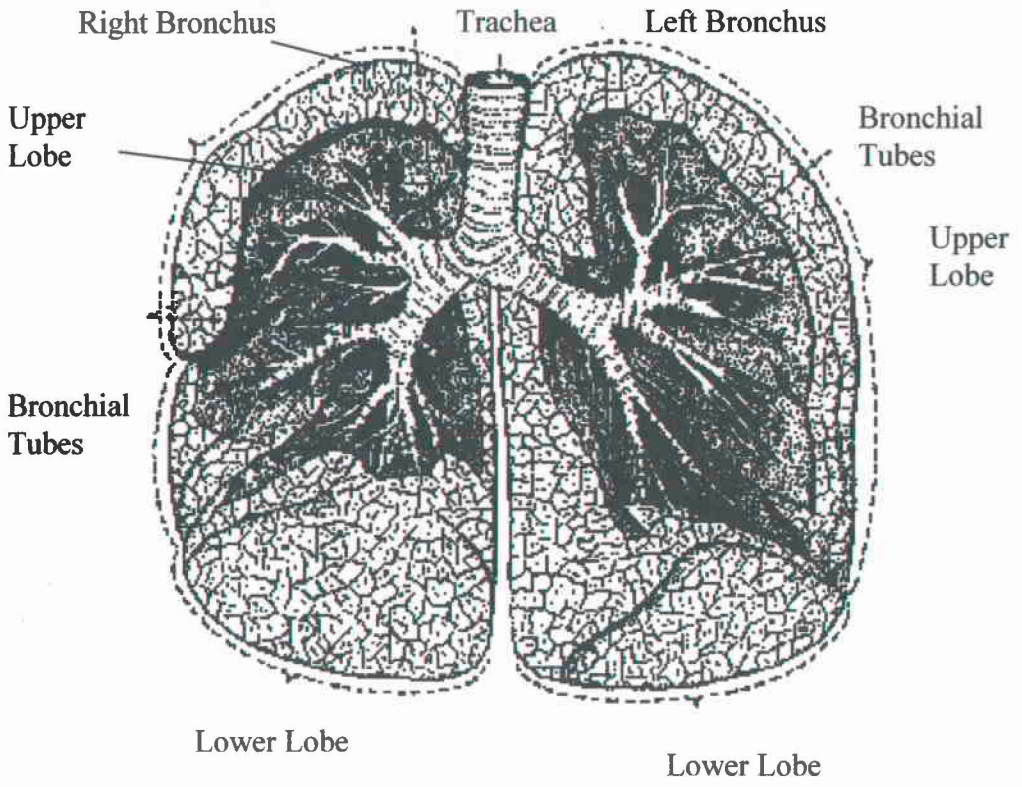
folds are forced open by air pressure and closed by elasticity and the Bernoulli effect (21).

Appelman (1985) said that the lungs are the primary organs of respiration. He added that their movement within the respiratory act depends upon the pressures exerted on them by the surrounding musculature (25).

The anatomy of the lungs was described as the following by Appelman (see Figure 1):

1. Each of the two lungs is divided into lobes. The left lung has two lobes, the right has three.
2. The lungs follow the outline of the thoracic cage. The base of the lungs follows the conformation of the domes of the diaphragm.
3. The porous, spongy tissue of the lungs is made up of millions of tiny air sacs, called alveoli, through which oxygen is passed to the blood stream.
4. During expiration the air is passed from the alveoli into larger sacs called the tabules, then to the bronchioles, to the bronchi, the trachea, and then the mouth (25).
5. The lungs are enclosed by a delicate membranous sac called the pleura. This enclosure makes the air pressure within the lungs very responsive to forces exerted upon it by the action of the thoracic cage and the abdominal diaphragm (26).

Figure 1. The Lungs (Appelman, 1985; 26).



Sundberg (1987) explained that the vocal folds require an overpressure of air in the lungs from the respiratory mechanism during singing, which he referred to as a subglottic pressure. Sundberg said the way in which the vocal folds vibrate at a given subglottic pressure is significant for the amplitude and also, to some degree, for the frequency of phonation (25).

Phillips (1996) explained the proper motion of breathing for regular respiration cycles, or what he described as the process of inhalation-exhalation. He described the correct breathing motion for singers as follows:

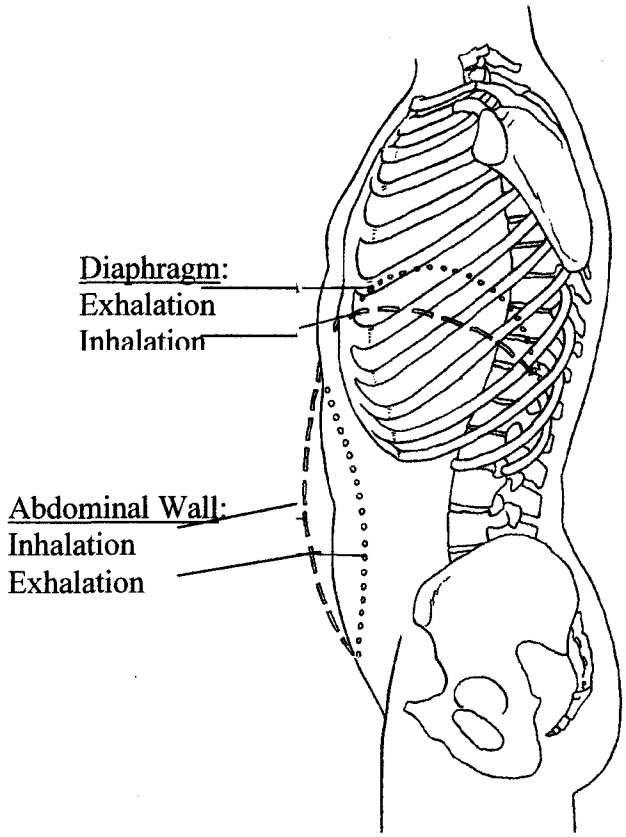
1. Inhalation: The diaphragm descends, or contracts, and the lower ribs expand outward, with a corresponding enlargement of the body around the waistline.
2. Exhalation: The diaphragm ascends, or relaxes, and the lower ribs contract inward, with a corresponding contraction of the body around the waistline (195).

Phillips (1996) said that the diaphragm is the major muscle of inhalation. He described it as having a double-dome shape and serving to separate the thoracic cavity from the abdominal cavity. He explained that it is a thin muscle like a large floor separating the torso into halves (196).

Phillips further explained the diaphragmatic action in respiration as the following (see Figure 2):

> The diaphragm is an "unpaired" muscle, that is, it has no paired muscle to reverse its action. Thus, the diaphragm only contracts during inhalation. In the breathing cycle, the diaphragm relaxes upward upon exhalation. The contraction of the abdominal muscles and muscles between the ribs, or internal intercostals, exert pressure upon the diaphragm during exhalation. The diaphragm then returns to its dome-shaped position under the lungs, decreasing the size of the thoracic and pleural cavities and thereby causing air to flow out of the lungs (198).

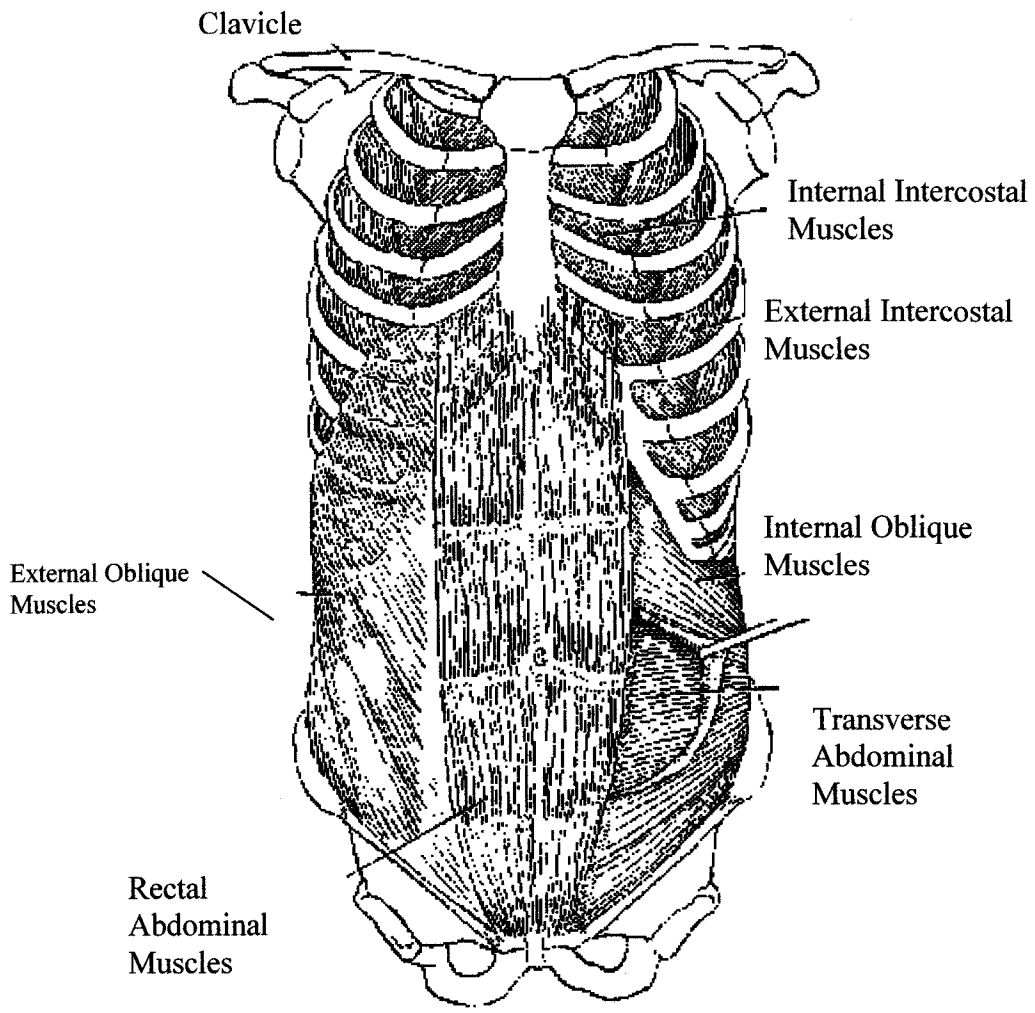
Figure 2. Abdomen-Diaphragm Displacement in Inhalation. (Phillips, 1996; 196).



Phillips (1996) identified two muscle groups that facilitate respiration (see Figure 3):

1. External intercostals: Muscles that help in the inhalation phase of the breathing cycle. When these contract, they move the rib cage upward, allowing for expansion of the thoracic cavity and inhalation of air. Acting together, the diaphragm and the external intercostals serve as the two basic muscle sources for the inhalation phase of the breathing motion.
2. Internal intercostals: The internal intercostals serve to pull the ribs inward and downward, contracting the thoracic cage and producing exhalation (199).

Figure 3. Abdominal Muscles of Support. (Phillips, 1996; 198)



The Larynx

Murphy (2002) explained that voice production is also dependent upon a vibrating source, which is the larynx (3). Gates (2003) clarified that the larynx is the human instrument for singing, but that it also plays three roles in the body—the principle role not being to produce sound. The three roles described were the following:

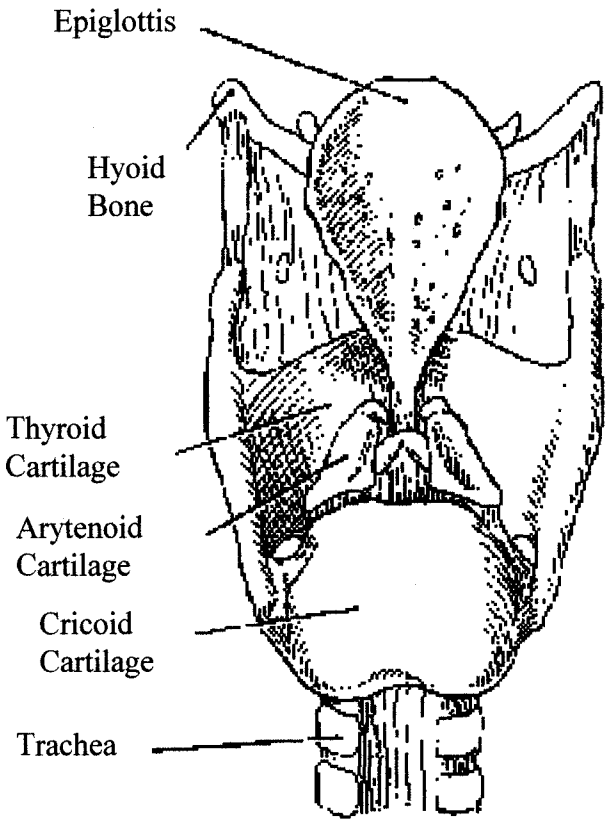
1. The larynx exists primarily to prevent anything from entering the lungs except air. A special type of mucous membrane called stratified squamous epithelium lines the larynx and has ultra sensitive sensors that cause the larynx to spasm and the epiglottis to snap shut over the windpipe, or trachea, if any substance other than air touches them. The spasm causes laryngeal muscles to contract, and the vocal folds to pull together to form a protective shield over the trachea.
2. The larynx allows you to hold your breath. Your vocal folds contract over the trachea and prevent air from escaping.
3. The larynx produces sound. In order to produce sound, the laryngeal musculature contracts to adduct the vocal folds. While they are adducted, air travels up from the lungs through the trachea. Air pressure builds in the trachea, under the adducted folds, and blows the folds apart. The folds then draw back together. Because moving air has less density than stagnant air (known as the Bernoulli Effect), the folds are sucked together until the air pressure builds again and the process repeats itself, producing vibrations that create sound (2).

Sataloff (1998) explained that the larynx can be thought of as comprised of four anatomic units which consist of the skeleton, mucosa, intrinsic muscles, and extrinsic muscles. He elaborated that the glottis is the space between the vocal folds (5).

Sataloff (1998) described the skeleton of the larynx as the following (see Figure 4):

> The most important parts of the laryngeal skeleton are the thyroid cartilage, cricoid cartilage, and the two arytenoid cartilages. Intrinsic muscles of the larynx are connected to these cartilages. The laryngeal cartilages are connected by soft attachments that allow changes in their relative angles and distances, thereby permitting alterations in the shape and tension of the tissues extended between them. The arytenoids are capable of complex motion (5).

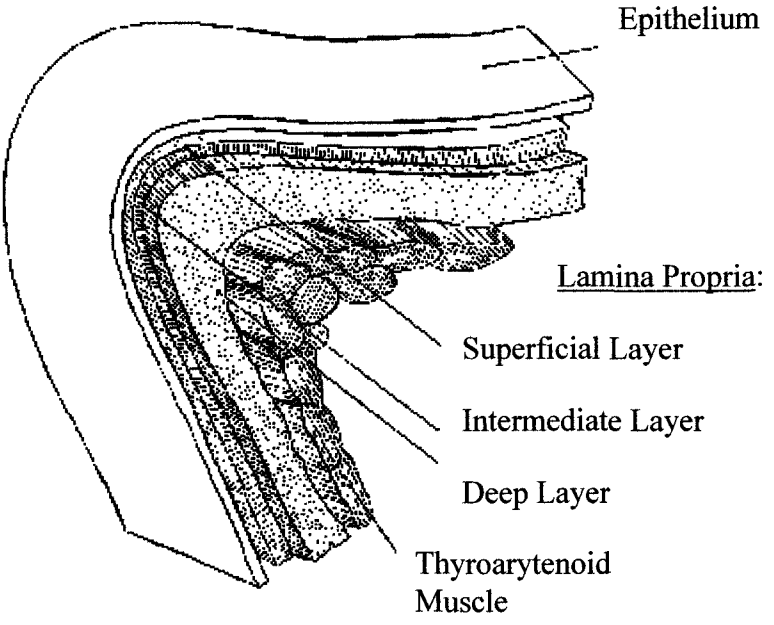
Figure 4. The Cartilages of the Larynx, Posterior (Phillips, 1996; 227).



Sataloff (1998) described the mucosa of the larynx as follows (see Figure 5):

> The vibratory margin of the vocal folds consists of five layers (epithelium; lamina propria: superficial layer, intermediate layer, and deep layer; and the thyroarytenoid muscle) (5). Functionally, the five layers have different mechanical properties and may be thought of as somewhat like ball bearings of different sizes in allowing the smooth shearing action necessary for proper vocal fold vibration (9).

Figure 5. The Mucosa of the Larynx (Phillips, 1996; 233).



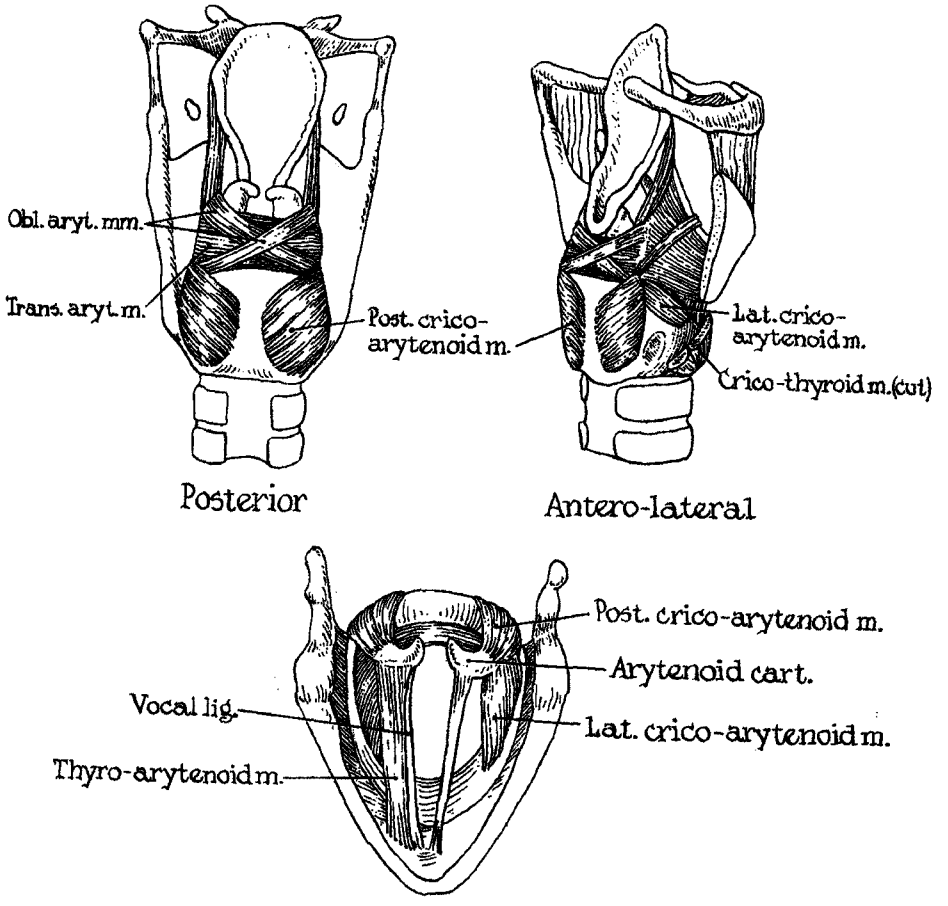
Sataloff (1998) described the intrinsic muscles of the larynx as follows (see Figure 6):

> Intrinsic muscles are responsible for abduction, adduction, and tension of the vocal folds. The intrinsic muscles of primary functional importance (in singing) are the thyroarytenoid, posterior cricoarytenoid, lateral cricoarytenoid, and cricothyroid. The thyroarytenoid muscle adducts, lowers, shortens, and thickens the vocal fold, rounding the vocal fold edge. It tends to lower vocal pitch. The thyroarytenoid is the third largest intrinsic muscle of the larynx. The thyroarytenoid muscle is divided into two compartments (medial and lateral). The medial compartment is also known as the vocalis muscle. It contains a high percentage of slow twitch muscle fibers. The lateral compartment has predominantly fast twitch muscle fibers (10).

He further elaborated on how these muscles interact:

> The posterior cricoarytenoid muscle abducts, elevates, elongates, and thins the vocal fold by rocking the arytenoids cartilage posterolaterally. When all layers are stiffened, the edge of the vocal fold is rounded. It is the second largest intrinsic muscle. The lateral cricoarytenoid muscle is a small muscle which adducts, lowers, elongates, and thins the vocal fold. When all layers are stiffened, the vocal fold edge takes on a more angular or sharp contour. The interarytenoid muscle primarily adducts the cartilaginous portion of the vocal folds. The cricothyroid muscle moves the vocal folds. It also lowers, stretches, elongates, and thins the vocal fold, stiffening all layers and sharpening the vocal fold contour. It is the largest intrinsic laryngeal muscle. The cricothyroid muscle is largely responsible for longitudinal tension, a very important factor in control of pitch. Contraction tends to increase vocal pitch (12).

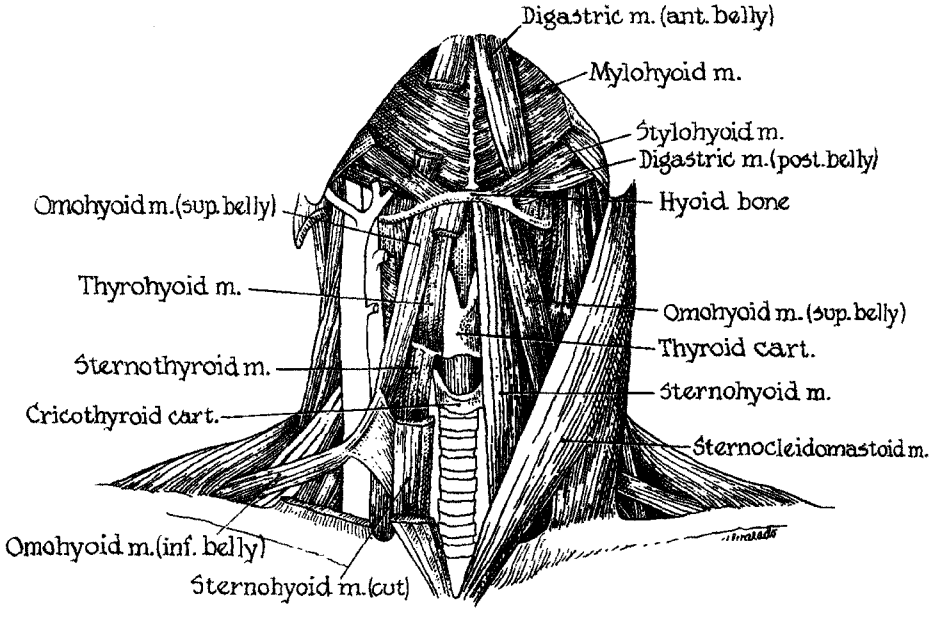
Figure 6. The Intrinsic Muscles of the Larynx (Phillips, 1996; 229).



Sataloff (1998) described the extrinsic muscles of the larynx as follows (see Figure 7):

> Extrinsic laryngeal musculature maintains the position of the larynx in the neck. The extrinsic muscles are critical in maintaining a stable laryngeal skeleton so that the delicate instrinsic musculature can work effectively. In the Western classically trained singer, the extrinsic muscles maintain the larynx in a relatively constant vertical position throughout the pitch range. Training of the intrinsic musculature results in vibratory symmetry of the vocal folds. This contributes to what the listener perceives as a "trained" sound. The extrinsic muscles may be divided into those below the hyoid bone (infrahyoid muscles) and those above the hyoid bone (suprahyoid muscles). The infrahyoid muscles include the thyrohyoid, sternothyroid, sternohyoid, and omohyoid. The suprahyoid muscles include the digastric, mylohyoid, and stylohyoid (17).

Figure 7. The Extrinsic Muscles of the Larynx (Phillips, 1996; 230).

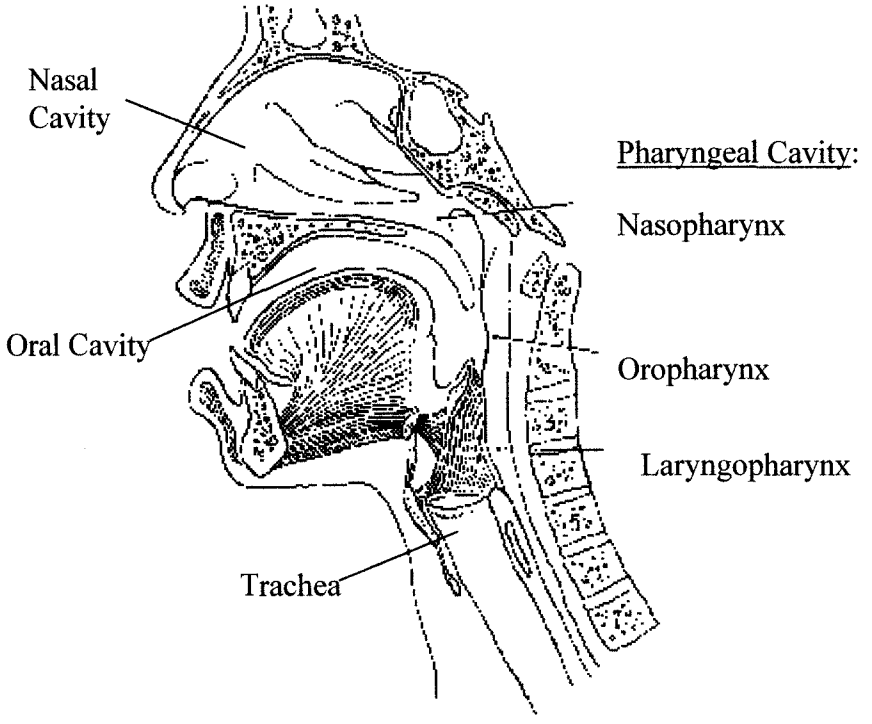


The Resonating System

Murphy (2002) stated that the third main component in voice production is the resonator which includes the nasal cavity, oral cavity, pharyngeal cavity, and trachea (3). Appelman (1985) similarly explained that resonance in singing depends upon four cavities that are resonators and upon the characteristics of those cavities to vary greatly in their ability to change in size, shape, aperture, and length of neck of the aperture. Appelman listed them as follows (see Figure 8):

1. The nasal cavity.
2. The oral cavity.
3. The pharyngeal cavity and its subdivisions, the nasopharynx, the oropharynx and the laryngopharynx.
4. The trachea (74).

Figure 8. The Resonating System (Phillips, 1996; 258).



Appelman (1985) described the nasal cavity as the following:

1. It extends from the floor of the cranium to the roof of the oral cavity.
2. The septum divides this large cavity into two separate cavities called fossae, which act as resonators in the production of nasal sounds.
3. The nasal fossae are not amenable to change, and their conformation cannot be changed during phonation at any pitch or intensity.
4. The nasal cavity has two orifices, the anterior nares at the front of the nose and the posterior nares at the opening into the orypharynx. He said that neither of these orifices is subject to control (74).

Appelman (1985) described the oral cavity as the following:

1. The narrow passage from the mouth to the pharynx which is the most amenable to change.
2. The shape and size of the oral cavity may be varied by the movement of the mandible, the tongue, and the lips (74).

Appelman (1985) described the pharynx as the following:

1. Approximately four and one-half inches in length, extending from the base of the skull to the level of the sixth cervical vertebra.
2. It is subdivided into three separate cavities (the nasopharynx, the oropharynx and the laryngopharynx); each contributing a quality component to the tonal spectrum (74).
   a. The nasopharynx:
      i. Extends vertically from the base of the skull to the velum anterior-posteriorly from the posterior nares to the pharyngeal wall.
      ii. Is a closed resonator and does not serve a primary function during the production of most vowels and consonants because the uvula is pressed firmly against the pharyngeal wall, thus closing the entrance to the nasopharynx (74).
      iii. Serves as a resonator in the production of the nasal consonants "n," "m," "n" (77).
   b. The oropharynx:
      i. Extends vertically from the velum and uvula to the tip of the epiglottis.
      ii. Anterior boundary is the postdorsum and the root of the tongue; the posterior border is the pharyngeal wall.
      iii. Is most amenable to change through the movement of the larynx and its transverse dimension is altered by the action of the pharyngeal constrictor muscles and muscles of articulation (77).

c. The laryngopharynx:
   i. Is an important resonator in the coupled phonatory system.
   ii. Extends from the tip of the epiglottis to the superior surface of the vocal fold. Its lateral boundaries are the aryepiglottic folds which completely enfold it.
   iii. Superior orifice is the epiglottis, which covers it as a lid during the act of swallowing.
   iv. Inferior orifice is the glottis.
   v. Has within its musculature the ventricular folds, or false vocal folds, which rise and nearly approximate during phonation (77).

Appelman (1985) described the trachea as the following:

1. A cartilaginous membranous tube extending from the subglottic larynx to the division of the bronchi.
2. This muscle contracts the tracheal ring and, thereby, decreases in diameter.
3. The relaxation of the trachea enlarges the ring and permits the passage of large volumes of air during forced respiration (79).
4. Tomograms revealed a tracheal alteration—a bulbous enlargement below the base of the cricoid cartilage—during the production of high pitches and increased vocal force (80).

Theories of Registration

The concept of vocal registers is related to the physiological act of singing. The previous sections of this chapter addressed the basic elements of voice production. This section will present how the various nuances of the basic elements of voice production interact to form and produce the different vocal registers which affect the singing voice. Various investigators have suggested theories to explain vocal registration. Following is a review of some of the extant literature.

The theory of vocal registers was advanced in the 1800s by Manuel Garcia. His invention of the laryngoscope made possible the viewing of the vocal folds in action. Garcia (1894) recognized vibratory patterns of the vocal folds, which he noted as being

responsible for three vocal registers, "upper," "middle," and "lower." He said "Each of the three registers has its own extent and sonority which vary according to the gender of the individual and the nature of the organ" (7). Garcia reported that only the inner edges of the vocal folds vibrate in the upper register and that the folds vibrate to the full width and length for the lower, or chest voice. He said these two different mechanical principles overlapped in the middle vocal register (7).

Berg (1980) theorized that registration involves more than the vocal ligaments. He said the various vocal responses of the human larynx are most commonly divided into the three main registers: "chest," "head" or "mid," and "falsetto register," and two lesser registers at the extreme low and high sides: "strohbass" and "flute" or "whistle" register (147).

Berg explained that when the laterals are active but the transverses inactive, a triangular opening is seen between the arytenoids, the vocal processes contact each other, but the posterior parts at the apex do not contact each other. He said this is true provided that the vocal folds are not stretched, as stretching of the vocal ligaments abducts the vocal processes, as the ligaments try to become oriented along the shortest line between their insertion on the thyroid and the articulation of the arytenoids on the cricoid. He said this adjustment is used for the flute or whistle register, up to 2500 cps, and weak tones at small flows of air and small subglottic pressures. He said the tone originates by the formation of eddies in the triangular opening and subsequent cavity resonance. He further elaborated that at large flows the subglottic pressure increases to the extent that the vocal folds are thrown into vibration. He said the very high whistle register is not an

extreme of the high falsetto register, but an entirely different register, which is found in some female voices (154).

Berg said when the transverses are active, but the laterals inactive, the posterior parts of the arytenoids at the apex contact each other, but the vocal processes do not contact each other or the contact is very slight. He said this adjustment is only suitable for the Strohbass voice, with its very low pitches, at small flows of air and with a vibrational pattern characterized by large amplitudes of the vocal folds, but no or only a short closure during the vocal cycle and, consequently, a small number of partials are produced. He said at large flows of air, a low-pitched chest voice results, as the closure time increases progressively with increasing flow (154).

Berg said the main registers are produced with active transverses and with various degrees of activity of the laterals. He said his experiments indicated that a gradual increase of the activity resulted in a gradual increase of the pitch, up to a certain level. He said this can be explained by considering that an increase of the medial compression of the vocal processes results in a decrease of the effective length of the glottis, up to the moment when the vocal processes are pressed so firmly against each other that a further increase becomes meaningless. He added that during this increase, the vibrational type remains essentially the same, i.e., the response remains in the same register. He said however, as the mid voice and the falsetto voice require progressively larger minimal degrees of stretching of the vocal ligaments, they require progressively larger minimal degrees of stretching of the vocal ligaments, they require also certain minimal medial compressions of the vocal processes. Berg said a complication arises when the larynx is made to respond in the regions when the main registers overlap each other. He said in

increase of the medial compression beyond a certain limit may then result in a register transition, thus, that the response drops to the next lower register, i.e., from falsetto to mid, or from mid to chest voice (154).

Berg said the chest voice is characterized by large amplitudes of the vocal folds, and these are only possible when the vocal ligaments are slack, i.e. at small longitudinal tensions in the vocal ligaments. He said when the longitudinal tension in the vocal ligaments is gradually increased, the larynx will stop phonating at a very small tension, somewhere between zero and twenty grams. He said this stop can be delayed and even be prevented by gradually increasing the medial compression of the vocal processes and/or the flow of air, but then the larynx will change into a mid or falsetto voice response at a longitudinal tension of about twenty to fifty grams (155).

Berg said a mid voice is obtained at longitudinal tensions in the vocal ligaments somewhere between ten and fifty grams. Berg stressed the point that the mid voice is not a really independent register but a mixture of chest and falsetto register. Berg pinpointed that a falsetto voice originates at longitudinal tensions in the vocal ligaments larger than about ten grams, the limit being higher, up to about fifty grams (155).

Sundberg (1987) stated that there is no generally accepted clear definition of the term "register." He defined it as the following:

1. A register is a phonation frequency range in which all tones are perceived as being produced in a similar way and which possess a similar voice timbre.
2. A vocal register is a totally laryngeal event; it consists of a series or a range of consecutive voice frequencies which can be produced with nearly identical phonatory quality; there will be little overlap in fundamental frequency and…the operational definition of a register must depend on supporting perceptual, acoustic, physiologic and aerodynamic evidence (50).

Sundberg (1987) explained that a register covers a certain phonation frequency range, but the various registers overlap, so that a person may phonate at a given phonation frequency in different registers. He said the range of overlap between male modal and falsetto registers is in the vicinity of 200 to 350 Hz, or approximately G3-F4. He added that in the female voice, the ranges of overlap are found in the neighborhood of the following phonation frequencies: chest—middle, 400 Hz, or pitch G4, and middle-head, 660 Hz, or pitch E5. He said those ranges of register overlap, and the register boundaries vary substantially among individuals (51).

Sundberg (1987) added that it is easy to determine the register from the voice timbre. He noted that a classic aim of singing pedagogy is to reduce or even eliminate, timbral variation between registers and that it is generally regarded as optimal that shifts into a different register be accompanied by the smallest possible timbral differences. He hypothesized that clearly audible register shifts can be eliminated by training. He noted that under such conditions the differences in the register of a skilled singer would be hard to define perceptually, although they may still exist at a laryngeal level (51).

Sundberg noted that there are differences in the muscular behavior between the female middle and head registers, as well as between the male modal and falsetto registers. He said when phonation frequency is raised beyond a certain point, it seems that the entire control apparatus is in some way reset, so that the values used for the bottom range of the lower register can again be used for the bottom tones of the next higher register, almost as in shifting the gears of a car (53).

Sundberg said the role of the lateral cricoarytenoid muscles in this connection would be to prevent the vocal folds from abducting. He said the abducting force may stem from

the contraction of some laryngeal muscles engaged in the pitch-raising maneuver. He added that this force must be taken care of by means of adductor muscles, such as the lateral cricoarytenoid muscles (54).

Hirano (1991) said there are three major vocal registers: "falsetto" or "light," "modal" or "heavy," and "vocal fry." He said falsetto is characterized by the absence of complete glottal closure. He added that the modal, or heavy register is accompanied by complete glottal closure for each vibratory cycle, and is traditionally subdivided into "head," "mid," and "chest" registers. He said vocal fry is characterized by an extremely long closed phase relative to one vibratory cycle and that, occasionally, it has two open phases during one vibratory cycle (212).

Hirano said muscular activity was compared among sustained tones phonated in different registers, but at the same pitch, by means of EMG. Hirano summarized the findings of a study of four singers:

1. The heavier the register, the greater the VOC (thyroarytenoid activity) in all four subjects.
2. LCA (lateral cricoarytenoid) tended to be more active for the heavy register, but not as consistently as VOC.
3. The difference in LCA activity was not as marked as that in VOC activity.
4. The activity of IA (interarytenoid), which was investigated in only one subject, was slightly greater for the heavier register.
5. The activity of CT (cricothyroid) did not show any consistent relationship to changes in vocal register (213).

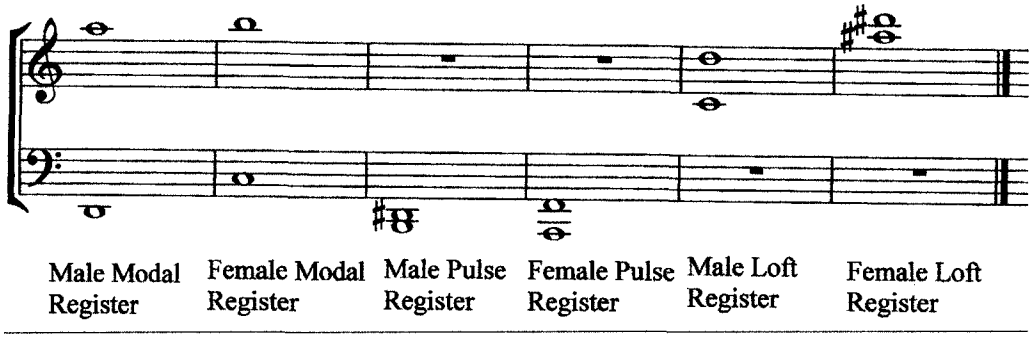
Hirano also reported results from another study in which the muscular activity pattern was investigated when the vocal register was shifted during singing scales, arpeggios, vocalizes, and parts of songs. Five singers served as the subjects for the tasks in that study (213). The results were the following:

1. VOC always presented a marked change in activity in response to register shifts.
2. Register shifts from heavy to light were accompanied by a decrease in VOC activity, whereas shifts in heavier registers were associated with VOC increases.
3. The direction of the changes in the LCA, CT, and IA activity were the same as in the case of the VOC.
4. In the LCA, changes in activity accompanied by register shifts were observed less consistently, and changes in CT activity were even less consistent.
5. IA, investigated in one subject, also showed changes in activities associated with register shifts (214).
6. As CT activity became greater relative to VOC activity, the voice changed from chest to mid, and mid to head (218).
7. The principal register agents appeared to be VOC and CT (219).
8. In falsetto, VOC contracted weakly or completely relaxed. CT was active to a significant degree (220).

Sataloff (1998) described that the acoustic impressions of vocal registers "are derived from changes in the way the vocal source signal is molded by the vocal tract as well as from differences in the vocal source signal itself" (34). While Sataloff acknowledged that many different terms have been coined to label the different registers, along with various definitions, he used the phrase "laryngeal registers" to refer to the concept of vocal registration. Sataloff suggested three distinct laryngeal registers (see Figure 9):

1. Modal register: Describes the laryngeal function in the range of fundamental frequencies most commonly used by untrained speakers (from about 75 to about 450 Hz in men; 130 to 520 Hz in women). The name, in fact, derived from the statistical term for "most common value." This register may include the musical "chest," or "low," "mid," and "head," or "high" registers, depending on how these are defined.
2. Pulse register: Occurs in the F0 range at the low end of the frequency scale (25 to 80 Hz in men; 20 to 45 Hz in women). The laryngeal output is perceived as pulsatile in nature. The term is broadly synonymous with "vocal fry," "glottal fry," or the musical term "strohbass."
3. Loft register: Is employed at the upper end of the vocal continuum (275 to 620 Hz in men; 490 to 1,130 Hz in women). The name is intended to convey a sense of "upper reaches." It corresponds to the older term "falsetto" (35).

Figure 9. Three Laryngeal Registers (Sataloff, 1998; 35).



Several different theories have been discussed in regards to vocal registration. Each researcher has also come up with a definition and a parameter for each register. These will be briefly summarized below. All investigators recognize at least three main vocal registers. Garcia (1894) recognized three registers that have their own extent and sonority which he said varies according to the gender of the individual and the nature of the organ. He defined the three registers by the part of the vocal folds which vibrate—the upper register utilizing only the inner edges of the vocal folds, the lower register utilizing the full width and length of the folds, and the middle register utilizing overlapping principals (7).

Berg (1980) outlined five different vocal registers—chest, head or mid, falsetto, strohbass, and flute or whistle. The parameters of these are defined by the activity or inactivity of the cricoarytenoid muscles and the longitudinal tensions in the vocal ligaments. He described the flute or whistle register (up to 2500 cps) as when the lateral cricoarytenoid muscles are active, but the transverses inactive, which produces a triangular opening between the arytenoids, where the vocal processes contact each other,

but the posterior parts at the apex do not contact each other. Berg described strohbass register as occurring when the transverse cricoarytenoid muscles are active and the lateral cricoarytenoid muscles inactive, and when the posterior parts of the arytenoids at the apex contact each other, but the vocal processes do not contact each other or the contact is very slight. This register is defined as having very low pitches, at small flows of air and with a vibrational pattern characterized by large amplitudes of the vocal folds (154). Berg described the chest voice as having large amplitudes of the vocal folds, said only to be possible when the vocal ligaments are slack or at small longitudinal tensions in the vocal ligaments. He described the mid voice as a mixture of chest and falsetto registers and at longitudinal tensions in the vocal ligaments between ten and fifty grams. Berg defined the falsetto voice as originating at longitudinal tensions in the vocal ligaments larger than about ten grams up to about fifty grams (155).

Sundberg (1987) explained that register covers a certain phonation frequency range, but the various registers overlap, so that a person may phonate at a given phonation frequency in different registers. He added that ranges of register overlap and that the register boundaries vary substantially among individuals. He said register can be determined from the voice timbre. Sundberg explained that there are differences in the muscular behavior between the registers. He said when phonation frequency is raised beyond a certain point, it seems that the entire control apparatus is in some way reset, so that the values used for the bottom range of the lower register can again be used for the bottom tones of the next higher register (53). Sundberg defined the registers by their overlap—male modal and falsetto overlap (G3-F4), female chest and middle overlap (G4), and female middle and head overlap (E5) (51).

Hirano (1991) recognized three major vocal registers—falsetto or light, modal or heavy, and vocal fry. He said that the falsetto or light register, is characterized by the absence of complete glottal closure. He said that the modal, or heavy register is accompanied by complete glottal closure and can be subdivided into head, mid, and chest registers. Hirano described vocal fry as being characterized by an extremely long closed phase relative to one vibratory cycle (212). Hirano described muscular activity in relation to phonation at the different registers. He summarized the following: 1) the heavier the register, the greater the thyroarytenoid activity; 2) the lateral cricoarytenoid tended to be more active for the heavy register, but not as consistently as the thyroarytenoid muscle (213); 3) register shifts from heavy to light were accompanied by a decrease in thyroarytenoid activity, whereas shifts in heavier registers were associated with thyroarytenoid muscle increases; 4) as cricothyroid muscle activity became greater relative to thyroarytenoid muscle activity, the voice changed from chest to mid, and mid to head; 5) the principal register agents appear to be thyroarytenoid muscle and cricothyroid muscle activity (219); 6) in falsetto, thyroarytenoid muscles contracted weakly ro completely relaxed and cricothyroid muscles were active to a significant degree (220).

Sataloff (1998) acknowledged three distinct registers—modal, pulse, and loft. He explained that the variant acoustic impressions derive from changes in the way the vocal source signal is molded by the vocal tract as well as from differences in the vocal source signal itself (34). He described the modal register (75 to 450 Herz in men and 130 to 520 Herz in women) as that most commonly used by untrained speakers and includes the chest, mid, and high registers. Sataloff described the pulse register (25 to 80 Herz in men

and 20 to 45 Herz in women) as being at the low end of the frequency and synonymous with "vocal fry," "glottal fry," or "strohbass." Finally, Sataloff described the loft register as being employed at the upper end of the vocal continuum (275 to 620 Herz in men and 490 to 1130 Herz in women). He said this corresponds to the term "falsetto" (35).

The theory of registration by Garcia (1894) didn't recognize the extreme lower and upper registers as does Berg (1980), Hirano (1998), and Sataloff (1998) but his definitions were in terms of the vibration of the vocal folds alone. Berg (1980) recognized five vocal registers and defined them by the activity of the muscles that control the vocal folds—namely the cricoarytenoid muscles, and the cricothyroid muscles and the longitudinal tensions of the vocal folds which he described in terms of tensions in grams. Sundberg (1987) acknowledged three vocal registers similar to that of Garcia (1894) and also agreed that vocal registers were produced by differences in the muscular behavior between the registers although those were not specified. He said that the ranges of register overlap and defined his three registers by the points of overlap for males and females. Hirano (1991) acknowledged three major vocal registers with three subregisters which amounted to five different registers similar to that of Berg (1980). His description of the registers were in terms of glottal cosure of the vocal folds and by the muscular activity of the thyroarytenoid muscles and the cricothyroid muscle. These were similar to that of Berg (1980). Sataloff (1998) acknowledged three distinct registers with three subgroupings to comprise a total of five different registers similar to that of Berg (1980) and Hirano (1991). Sataloff defined the registers in terms of frequency in Herz.

Figure 10 summarizes the theories of registration terminology by the various researchers discussed.

Figure 10. Theories of Registration: Comparison of Terminology.

Author	Register Terminology
Garcia (1894)	Upper, middle, lower
Berg (1980)	Chest, head/mid, falsetto, strohbass, flute/whistle
Sundberg (1987)	Male modal, male falsetto, female chest, female middle, female head.
Hirano (1991)	Falsetto/light, modal/heavy, vocal fry.
Sataloff (1998)	Modal, pulse, loft.

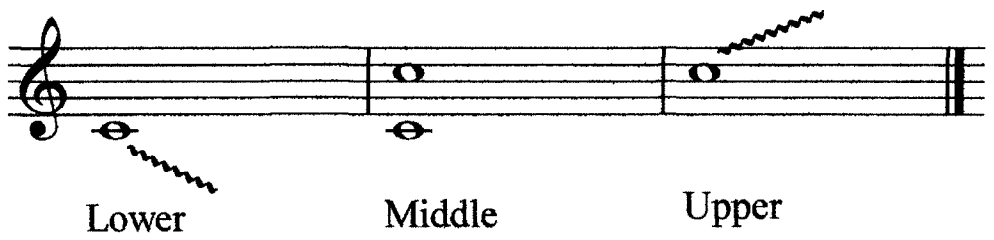
VOCAL REGISTERS OF CHILDREN

It is perceivable that the vocal registers of children are different from those of the adult female and male changed voice. Following is an overview of the vocal registration of the child.

Phillips (1996) advocated a three-register approach in working with the vocal registers of children. Phillips labeled these three registers "upper," "middle," and "lower." He recommended that the pure lower voice, which utilizes vibration of the full length and width of the vocal folds, should be used only from C4 and lower. He said C4 is the pitch where children will traditionally shift into the chest voice, if permitted. He added that the pure upper voice, which utilizes vibration of the inner edges of the vocal folds, begins at C5 and extends upward. Phillips explained that between the two pure "upper" and

"lower" registers is the "middle" voice, which falls between the pitches C4 and C5. Phillips said the "middle" voice is a combination of both lower and upper registers (see Figure 11) (43).

Figure 11. Vocal Registers of Children (Phillips, 1996; 43).



Phillips said the middle register, or shared adjustment, bridges the transition from pure lower to pure upper voice, and results in a fifty-fifty balance of the registers at approximately the pitch F#4. He said the rationale behind the three-register approach is to establish a healthy pedagogy for child and adolescent vocal training (44).

According to Phillips (1996), children lacking vocal instruction gravitate to the chest voice for singing. He said this is the speaking-voice register, with which they are most comfortable. Phillips explained that because this lower register in the child's voice is quite elastic, it can be used to sing far above C4, resulting in a harsh sound and strained technique, which is potentially damaging to the vocal folds. Phillips said that such singing, if sustained over a long period of time, can lead to permanent vocal damage. He

said this is due to the fact that the vocal folds make far greater physical contact when singing in the lower adjustment (44).

Phillips (1996) recommended a top-down approach to registration for working with pitches between C4 and C5. He explained that when the lower register is permitted to join with the upper, the sound maintains it ring and robust quality. He explained that the joining must be accomplished from the top downward; that is, vocalizes begin in the upper voice and move downward into the middle register (C5 to C4). He said this approach permits the gradual sharing necessary for a balance of vocal-fold vibration in the middle voice. Phillips said early attempts to blend these registers with vocalizes from lower to upper will prove difficult because students inevitably do not shift out of the chest voice as they ascend, and can only do so once the balance is understood aurally and kinesthetically. Phillips said this technique of blending registers is best learned from the top down (46).

Phillips (1996) pointed out that the pitch F#4, which is evenly divided between the lower and upper registers, is the pivotal note in balancing the middle register. He said pitches beginning at G4 or higher will gradually have more and more upper register quality, when vibrations of the vocal folds are in the process of moving to the inner edges, until the quality becomes pure upper at pitch C5, when the vocal folds vibrate only on the inner edges. He further explained that the pitches beginning at F4 or lower will gradually have more lower quality, which include vibrations of the vocal folds moving to the full width and length, until the quality becomes pure lower at C4 (46).

VOCAL REGISTERS OF ADOLESCENTS

The adolescent age is typically the period between childhood and adulthood. The exact age varies from person to person and between genders. Kahane (1978) noted that pubertal changes occurred in female and male subjects ranging in age from nine to nineteen years (11). Various physiological changes occur during this period of development. The singing voice moreover undergoes various changes. This age range is included in the study and is therefore relevant to this paper. An overview of the vocal registers of adolescents follows.

Phillips (1996) outlined that the adolescent period in human development is generally considered to begin in the seventh grade and continue through the twelfth grade. He defined it as "a time of passage from childhood to adulthood" (75). Phillips said students in the adolescent years undergo a voice-change process that involves registration problems not encountered in children. He said the adolescent female does not face the radical voice change of the male, but does experience a vocal adjustment. Phillips mentioned that adolescent males must learn to deal with a range that is lowered at least an octave (47).

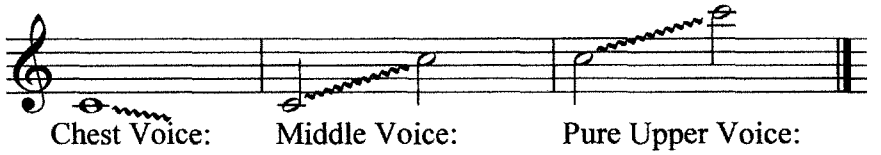
The Adolescent Female Voice

Kahane (1978) said the dimensions of the pre-pubertal female larynx are closer to adult size and weight than their male counterparts, and "thus, the pre-pubertal female larynx requires less growth per unit time to reach maturity" (18). Phillips (1996) said, however, that the female larynx does increase in size, more laterally (width) than anteroposteriorly (front to back). He said this, along with a slight increase in vocal-fold length and

thickening, results in a period of adjustment in which the voice may become breathy or sound congested (47). Collins (1981) said that the period of change for the adolescent female voice is between grades four and nine (12).

Phillips (1996) explained that the vocal registers of adolescent females is similar to that of children with slight changes. He said the chest voice begins in its pure form at C4 and extends lower as the voice matures. He explained that the middle voice register (C4 to C5) is similar to that of the child in that it is a sharing of lower and upper registers. Phillips said the traditional treble voice break between the chest and upper registers, occurs at approximately pitch A4 for the female adolescent singer. He said the pure upper voice begins at C5 and extends upward an octave to C6 (47) (see Figure 12).

Figure 12. Vocal Registers of Adolescent Females (Phillips; 1996; 47).



Phillips (1996) said that adolescent girls seem to sing in one register—mixed or middle. He further explained that as they venture to sing above pitch C5, they fail to make the necessary adjustment of removing all of the lower vibratory pattern from the voice. He said this added weight of the lower voice prohibits the ease of production in the upper voice, the range of which extends to around F5. Phillips said that this is compounded by an elevated larynx, which interferes with the lengthening and thinning of the vocal folds.

Phillips advocated that all females, including altos, should be vocalized in a light manner from C5 to C6. He said this will help them to learn the feeling of moving in an inner-edge-only vibratory pattern of the vocal folds, thus eliminating all lower-voice production (48).

Gackle (1991) prescribed a model for classifying characteristic stages of development in the adolescent female voice. The stages and various levels follow:

> Stage I: Prepubertal: Ages 8-10
> Singing voice: Light, flutelike quality; no apparent register breaks; soprano quality; flexible, able to manage intervallic skips; much like male voice at same age with the exception that the female voice is lighter in "weight" because the volume potential is generally not as great. Depending on other physiological changes (i.e., breast development, menarche) this stage could continue through age 12 or 13.
>
> Stage IIA: Pubescence/Pre-Menarcheal: Ages 11-12
> First signs of physical maturation begin.
> Singing voice: Breathiness in the tone due to appearance of mutational chink, an inadequate closure of the vocal folds as growth occurs in the laryngeal area; register break appears between G4 and B4; if not using lower or chest voice, there is apparent loss of lower range—around C4. Some girls have trouble producing chest voice at this time.
> Symptomatic signs: Difficulty or discomfort with singing; difficulty in achieving volume (especially in middle and upper range); breathy tone throughout upper range (head voice); fuller tone in lower/chest range; obvious "flip" into breathy, childlike, fluty voice at transition from lower to upper registers.
>
> Stage IIB: Puberty/Post-Menarcheal: Ages 13-14
> Peak of Maturation
> Singing voice: Very critical time; after the Stage IIA, tessituras can move up or down or, sometimes, can narrow at either end, yielding basically a five- or six-note range of comfortable singing. Register breaks still apparent between G4 and B4, and also at D5 to F#5. At times, lower notes are more easily produced, yielding an illusion of an alto quality; singing in this range may be easier and can be recommended for short periods of time; singing only in the lower range for an indefinite period of time can be injurious to the young unsettled voice because of the tendency to over use the lower (chest) register.
>
> Vocalization should occur throughout the vocal range, always striving to avoid any unnecessary strain in the lower or upper range. Because the changes during this stage are sporadic and unpredictable, it is necessary to listen to individual voices frequently in order to assess vocal development.

Symptomatic signs: Hoarseness without upper respiratory infection; voice cracking; difficulty or discomfort with singing; lack of clarity in the tone.

Stage III: Young Adult Female/Post-Menarcheal: Ages 14-15
Singing voice: Overall range capabilities increase. (At times, range does not decrease during the time of mutation. One characteristic of a quality singing voice is that it encompasses a large range. This does not imply that any voice is an alto at age 15-16 simply because those tones are within the young singer's capability.) Greater consistency occurs between registers; voice breaks are more apparent at passaggio D5 to F#5 (more typical of adult voice). Breathiness appears to decrease. Tone, though not as full as mature adult, is deeper and richer. Ease returns in the singing process. Vibrato appears in the voice. Volume, resonance, and vocal agility increase (22-33).

The Adolescent Male Voice

The adolescent male voice undergoes a greater physiological change than the adolescent female voice (Kahane, 1978). Cooper (1963) said that about ninety percent of boys experience the voice change sometime between entering grade seven and the end of grade nine. He said that the male adolescent years encompass grades four through twelve (5). The greatest period of change in the adolescent male voice has been documented between the ages of twelve to fifteen years of age (Cooksey, 1992).

The adolescent male voice change was documented by Kahane (1978). Kahane noted that by puberty, the male larynx was significantly larger than in the female. Kahane explained that the thyroid eminence, or Adam's apple was also more clearly prominent in the male. Kahane also explained that the vocal folds in both sexes reached essentially their adult length by puberty; however, the absolute length of the male vocal folds had increased by over two times that of the female. Kahane said the significantly greater growth of the male vocal folds compared with female partially explained the structural bases for the dramatic drop in fundamental frequency in the male voice during puberty (11, 18).

According to Alderson (1979) the boy singer who enters adolescence begins to lose the ability to sing in his pure upper register for the pitches C5 to C6, as his range in the lower register begins to expand downward. He said this does not mean that adolescent boys lose the ability to sing in the pure upper voice. He added that the upper register of the pubertal boy must remain in use in what was his middle register (C4 to C5) if the new top of his range is to be successfully handled for singing. He stated, "While the boy's voice is changing he should rely on head voice techniques for his highest notes" (234).

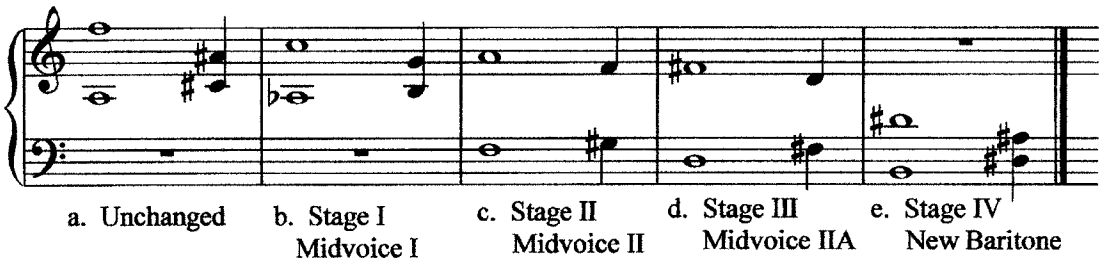
Cooksey (1992) said that voice change in adolescent boys is a predictable, sequential, but sometime erratic process which generally takes place over a period of one to two years. He explained that voices do not change overnight, but develop over time as vocal folds thicken and lengthen, ligament and laryngeal cartilages develop, and vocal tract expansion takes place. He said the most active phase of change occurs on the average between the ages twelve-and-a-half to fourteen (9).

Cooksey developed a classification system that included the following stages of the male voice maturation (see Figure 13):

Stage 0 = Unchanged (premutational)
Stage 1 = Midvoice I (initial period of voice change)
Stage 2 = Midvoice II (high mutation period)
Stage 3 = Midvoice IIA (climax of mutation and key transitional period)
Stage 4 = New Baritone (stabilizing period) (postmutational development and re-expansion period) (9).

Cooksey outlined the mean ranges and tessituras for the voice change stages. He said low terminal pitches of the singing range indicated a logical progression downward by thirds (10).

Figure 13. Mean Ranges and Tessituras for the Male Voice Change Stages (Cooksey, 1992; 10).



Cooksey said the growth pattern described is highly correlated with the development of primary and secondary sexual characteristics associated with puberty. He said the onset of the male voice change, Stage 1--Midvoice I, begins as the physiological changes take place. He described this stage as the following:

1. Higher pitches are lost (usually between C5-F5) and the singing range decreases by about four half steps.
2. While high notes may still be produced for a while, there is increased constriction and breathiness. There is also an overall decrease in the richness of the tone.
3. Begins between the ages of twelve to thirteen, but can occur sooner and may last for several months or continue for more than a year. Individual growth ages vary, so no one can predict how long adolescent singers will remain in this stage (11).

Cooksey described Stage 2--the Midvoice II stage as the following:

1. Lower pitches begin to appear in the male adolescent's range while higher notes become much more unstable.
2. Primary and secondary sexual characteristics are more evident.
3. The lower pitch plateau is established at the E3-F3 level below C4, high terminal pitches, A4-C5, begin to fluctuate and the falsetto register begins to emerge.

4. Due to transition adjustments between the two registers, modal and falsetto, some coordination can be lost. On the other hand, some voices have no trouble with this and can even produce notes in the whistle register (11).
5. Male voice quality is distinctive—thicker, darker in color, and less resonant.
6. The most stable pitch areas are between A3-F4 (around middle C).
7. This stage lasts for about twelve to thirteen months, and begins closer to age thirteen, although there are many exceptions (11).

Cooksey described the male voice change in Stage 3--Midvoice IIA as the following:

1. This is the time when the voice is most vulnerable to abuse.
2. Low terminal pitch descends to D3 and average high terminal pitch occurs at about F4.
3. The falsetto register continues above this point, but in some cases is not easily produced—particularly in the area just above the highest pitch in the modal range.
4. The most comfortable part of the singing range is limited to G3-C4/D4.
5. The quality of the voice is huskier than Midvoice II, and has a tendency towards breathiness. He added that there is also a loss of agility, and boys sometimes want to "push" or force the tone, especially in the lower and upper pitch extremes.
6. Midvoice IIA is a transition stage and may last for only four weeks, or in some cases two to ten months.
7. The majority of Midvoice IIA's are found in the eighth grade, and that the normal age span of boys in this classification is thirteen to fourteen years (12).

Cooksey described Stage 4--the New Baritone stage as the following:

1. Represents the beginning of more stability in the voice change process since the most dramatic period of maturation is over.
2. The range is stable (Bb2/B2-C4/D4), and the voice quality is clearer but light and somewhat thin.
3. The sound does not approximate adult-like quality.
4. The new baritone has less pitch agility, but usually produces the falsetto register fairly easily.
5. In some cases, there is a blank spot (C4-F4) where no notes can be produced at all.
6. Some may experience difficulty in finding the falsetto while others can sing falsetto very easily above F4, but cannot produce pitches below that point without pushing or forcing the tone in the modal register.
7. A large percentage of new baritones can be found in both the eighth and ninth grades.
8. The normal mean age for this classification is fourteen years, but can vary between ages thirteen to fifteen (12).

INACCURATE SINGING

Inaccurate singing, also known as "monotone," "tone deaf," "untuned," "out-of-tune," "off-pitch," "off-key," "untuned," "uncertain," and other terms, is difficult to define. There is no known standardized definition of inaccurate singing today. Definitions have been dependent upon the type of assessment, subject age, testing conditions, or other variables.

Moss (1973) defined the out-of-tune singer as "a child whose vocal performance is characterized by an inability to sing the melodic intervals of familiar songs with acceptable pitch accuracy" (3). Roberts & Davies (1975) defined "monotone" subjects as those who are "always completely untuneful with little variation in pitch" (228). Aaron (1991) used the term "inaccurate singer" to denote that person who sings "off pitch." Goetze (1985), Reuter (1956), Richner (1976) and Zwissler (1972) also use this term for designating this condition. Zwissler (1972) defined the inaccurate singer as that child "who cannot correctly reproduce a given melody at a given pitch level" (7). Richner (1976) defined the "inaccurate singer" as one who sings out-of-tune in both individual and group situations.

Rating Scales

Various types of rating scales have been developed and used to classify the different types and stages of singing accuracy. Smith (1961) constructed one of the first rating scales used to evaluate the young child's singing voice. Smith specified criteria for each of the four scoring levels he used to evaluate the performance of intervals. The scale was

intended to measure intonation of intervals rather than use of the singing voice. The rating scale follows:

4 = Complete accuracy in reproducing an interval.
3 = Complete accuracy but with a tendency to slide into either the first or second interval tone.
2 = Child was able to sing only one of the two tones.
1 = Complete lack of tone matching ability (137).

Eikum's (1963) study used two elementary music teachers and one elementary classroom teacher to rate recorded samples of the students' singing (N=1040). The scores "four," "three," "two," "one," and "zero" conformed, respectively, to the terms "excellent," "good," "average," "fair," and "poor." Students who made teacher rating scores of "three" or "four," were classified as singers. Students who deviated from those classified as singers conformed with the following list:

1 = Sing in the speaking range only and in a very limited range (may include those who appear to sing an octave lower than much of the song melody).
2 = Sing consistently higher than the correct melody but seldom, or not at all, lower than the melody.
3 = Follow the general direction of the melodic line but do not sing the correct note at the proper time.
4 = Have trouble matching even one tone correctly.
5 = May imitate only two or three, or a very small number of tones or part of a phrase.
6 = Can sing with the group but not alone.
7 = Can sing alone but get confused when singing with the group.
8 = Do not differentiate between the speaking and singing voice but say the words in rhythm with no tune or noticeable variation in pitch.
9 = Possess similar characteristics to number 8 but who pronounce words infrequently and in a sporadic nature with relation to the rhythm of the song (18).

Boardman (1964) was also one of the first investigators to use an instrument to measure children's use of singing voice. The instrument was a seven-point scale. One point was given if a student failed to respond at all, and seven points were scored for a perfect response. There was no criteria provided by the author for points two through six.

Dittemore (1969) used a seven-point rating scale. The scale was limited to songs containing four melodic patterns only. Although points could be added for any additional patterns, that alteration prevented comparisons of scores if songs of varying length were used. The rating scale is as follows:

1 = No correct response.
2 = Student displays only a general sense of direction.
3 = Student displays a general sense of tonal direction and comprehends tonality.
4 = ...And correctly completes any one melodic pattern.
5 = ...Any two melodic patterns.
6 = ...Any three melodic patterns.
7 = ...Any four melodic patterns (29).

Joyner (1969) developed rating scales to determine accurate and inaccurate singers. The main singing test was the performance of the *British National Anthem*, initially in its usual key of G major, and then in a series of stepwise downward transpositions to G major, one octave lower. Each performance received a subjective grading of "A," "B," or "C" according to the following rubrics:

A = no major pitch errors or departures from the melodic outline.
B = generally erratic in pitch, in spite of moments of tunefulness.
C = no hint whatsoever of melodic outline or pitch coincidence (115).

Joyner then developed different levels of tuneful singing. The four divisions included:

1. Normal Singers: Able to sing in tune at both pitches.
2. Grade A Monotones: Tuneful at low pitches, untuneful at usual one.
3. Grade B Monotones: Erratic at both pitches, slightly better at low one.
4. Grade C Monotones: Always completely untuneful (117).

Pedersen and Pedersen (1970) used the following rating scale to judge vocal pitch accuracy of subject:

4 = Perfect vocal production of the standard pitch or pitches.
3 = Vocal production of the standard pitch or pitches one octave higher or lower.
2 = Vocal production of the standard interval pitches or standard sequence pitches in another key but with proper pitch relationships (this did not apply to Series A).

1 = Correct vocal production of one out of two pitches in Series B or two out of three pitches in Series C.
0 = Unsuitable vocal response (268).

DeYarman (1972) revised the Dittemore scale by changing the phrase "sense of tonality" to "comprehends tonality" to help clarify the criterion. Since the measurement of melodic patterns was not conducted, the rating scale could be used with any song materials. The rating scale follows:

1 = No correct response.
2 = No, or very poor sense of tonality, but general sense of direction.
3 = Poor sense of tonality, general sense of direction.
4 = Fair, moderately good, sense of tonality, good sense of direction.
5 = Good sense of tonality, very good sense of direction.
6 = Very good sense of tonality.
7 = Excellent tonal performance (12).

Roberts and Davies (1975) used a five-point scale similar in format to the Boardman scale of 1964. Criteria were not established for the scoring levels and considerable subjectivity was utilized. The scale was as follows:

0 = Tune completely unrecognizable.
1 = Part of tune recognizable.
2 = (Unspecified)
3 = (Unspecified)
4 = Correct performance (24).

Jarjisian (1981) developed a rating scale to determine vocal pitch accuracy. A five-point rating scale was employed by two judges, the investigator and another music educator, to evaluate the performances of the criterion songs. The judges rated the performances independently of each other. Following is the rating scale used:

1 = Use of singing voice.
2 = Maintenance of pitch center or general sense of melodic direction.
3 = Maintenance of pitch center and general sense of melodic direction.
4 = Accuracy in singing adjacent intervals or leaps.
5 = Accuracy in singing adjacent intervals and leaps (22).

Phillips (1983) also developed a five-point rating scale to determine singing accuracy. The rating scale is as follows:

5 = Superior Accuracy: Subject can sing the musical phrase correctly with direction and accurate matching of pitch (98%).
4 = High Accuracy: Subject can sing the musical phrase correctly with direction and accurate matching of most pitches (90%).
3 = Medium Accuracy: Subject can sing the musical phrase correctly with direction and accurate matching of at least half the pitches (50%).
2 = Low Accuracy: Subject can sing the musical phrase with general direction and accurate matching of less than half the pitches.
1 = No Accuracy: Subject cannot sing the phrase with direction or accurate pitch, or no response (214).

Small and McCachern devised a pitch-matching test and rating scale for their 1983 study. All subjects with perfect pitch matching scores on both pretests, (N=8) were eliminated from further experimental participation.

Feierabend (1985) developed a five-point rating scale to judge singing accuracy. The rating scale was as follows:

5 = The tonal pattern was accurately reproduced with good intonation.
4 = The tonal pattern was correctly reproduced but with some uncertainty.
3 = Melodic direction was evident but some tones were incorrectly reproduced.
2 = Melodic direction was evident but no tones were correctly produced.
1 = Reproduction of the tonal pattern was not recognizable (57).

Graham Welch (1986), developed a "continuum of singing ability" which was characterized by five stages from out-of-tune (stage 1) to in-tune singing (stage 5). These stages were:

Stage 1: The words of the song appear to be the initial center of interest rather than the melody. Often there is little variation in sung pitch, perhaps because some children find it impossible to attend to more than one parameter of the song at any one time, and words are, for them, the dominant feature. In response to a pitch stimulus, children appear to choose a comfortable vocal pitch rather than attempt to match the target. There is some evidence, that the comfortable pitch is frequently consonant with the pitch target.

Stage 2: Some variation in sung pitch may occasionally coincide with the target. There is a growing awareness that vocal pitch can be a conscious process and that changes in pitch are controllable.
Stage 3: A more active attempt is made to control vocal pitch by making the voice jump intervals toward the target. More individual pitches are matched correctly. Melodic outline follows the general contours of the target melody. Vocal range continues to expand.
Stage 4: Children are now able to perform some fine-tuning of pitches. The melodic shape and composite pitches are mostly correct, but some changes of tonality may occur if the pitch targets become uncomfortable or outside the still relatively limited vocal range.
Stage 5: No major pitch or melodic errors are made. There is a high level of pitch-matching ability. Vocal range is both higher and lower than previous stages (300).

Rooks (1987) developed a rating system to determine the accurate and inaccurate singer (N=53). A test of singing accuracy was constructed and used as a pretest and a posttest. The first section of both the pretest and posttest was comprised of eight one-note, two-note, three-note pitch-matching tests and four seven-note matching tests. The second part of the test included the singing of a song in both the low and high registers (35). The rating scales follow:

Single Pitch Tests
4 points = Sung pitch was within a whole step of test pitch.
3 points = Sung pitch was within a whole step of test pitch.
2 points = Sung pitch was within a major third of test pitch.
1 point = Sung pitch was within a perfect fifth of test pitch.
0 point = Sung pitch was an interval of a perfect fifth or greater from test pitch (65).

Multiple Pitch Tests
4 points = All sung pitches were within a whole step of test pitch.
3 points = All sung pitches were within a whole step of test pitch.
2 points = Sung pitches moved in same direction as test pitches, but one or more pitches were greater than a whole step from test pitch.
1 point = Sung pitches did not resemble the melodic contour of test pitches, but did move from a single pitch level.
0 point = Sung pitch maintained a single, static pitch level (65).

The *Singing Voice Development Measure* was developed by Rutkowski (1990) for studying the singing responses of kindergarten children. The five categories were very similar to those proposed by Welch (1986):

1. Pre-singers: Children who do not sustain tones; their singing response resembles chanting in the speaking voice range.
2. Speaking-range singers: Children who sustain tones and exhibit some sensitivity to pitch but remain within the speaking voice range…
3. Uncertain singers: Children who sustain tones but often waver between a speaking-voice range and a singing-voice range. When in singing voice, they utilize a range up to approximately F#4.
4. Initial-range singers: Children who have use of the singing-voice range up to the register lift, usually to A4.
5. Singers: Children who are able to sing over the register lift, Bb3-Bb4 and above, and have full use of their singing voices (92).

Atterbury and Silcox (1993) developed a rating scale to judge singing ability of young children. The taped singing of the song *Pinto Pony* by all subjects was evaluated using the following scale:

1 = Presinger: Does not sing but chants the song text.
2 = Uncertain singer: Sustains tones, uses both speaking and singing voice. when singing uses a limited range of about a third.
3 = Partial singer: Sings some phrases correctly but not entire song.
4 = Singer: Sings entire song correctly in one key (43).

Mizener (1993) developed a seven-point set of criteria for rating singing accuracy was designed for analysis of singing accuracy.

7 = Begins and ends in same tonality, with no loss of tonality within the song and no noticeably inaccurate intervals.
6 = Begins and ends in same tonality with no loss of tonality within the song but with some noticeably inaccurate intervals.
5 = Begins and ends in same tonality but with loss of tonality within the song and some noticeably inaccurate intervals.
4 = Ends in tonality different from the beginning tonality, with some noticeably inaccurate intervals and/or an abrupt shift in tonality within song.
3 = Begins and ends in same tonality but with little pitch variation around the tonal center.
2 = Ends in tonality different from beginning tonality, or ends in spoken tones, with little pitch variation around the tonal center.

1 = Has no clearly established tonal center and most intervals are inaccurate, or is chanted in spoken tones (235).

Phillips and Aitchison (1997) developed a rating system for determining vocal pitch accuracy. The investigator twice played a three-note tonal pattern in standard quarter note rhythm on an electric keyboard, and then asked the subject to repond by singing the pattern on "loo." There were four patterns using pitches within the octave of middle C4 to one octave above. The patterns were pitches "five-three-one," "five-eight-seven," "six-four-two," and "three-five-one." The investigator said that those were standard patterns found in children's song literature, and with which the children were familiar from singing in class. As the subject responded by singing the pitches of each pattern, the investigator marked on a form those pitches sung incorrectly, and totaled those sung correctly, with a possible score of twelve. The scoring was dependent upon the subjective judgement of the investigator.

Mathias (1997) developed a system for determining the accurate and inaccurate singer. Students were told that the investigator had made a video tape of herself singing. They would see her on the video singing short phrases that they would be asked to repeat, and their responses would be tape recorded. The investigator started recording on the audio tape and recorded the student's number on tape and stage of testing. The investigator then began the video. The student listened to the first example. The investigator paused the video tape but kept the audio recorder going while the student sang. Then investigator then asked the student if it was a match, and the student answered. The investigator reinforced the student if the response was correct as related to what was actually sung. If the response was incorrect, the investigator told the student it was not a match and asked the student if it was sung above or below the example on the video tape.

If the student knew it was not a match, the investigator still asked the student if he or she thought his/her voice was above or below the example he or she heard. If the student could not give any response, the investigator told the student whether he or she were above or below the pitch. The investigator then paused the audio tape. This process was followed for the remaining two examples. Those students who matched their voices correctly with each example were thanked, invited to choose a sticker, and returned to the classroom. The investigator made note of perfect scores on the response sheet. Students who received a perfect score on the pretest were excused from the remaining part of the study and were considered accurate singers. Those students who did not sing all responses correctly remained for the teaching aspect of the study and were considered to be inaccurate singers.

Mathias made an adaptation of the Boardman (1964) scale for the posttest of the study. The judges listened to each student's pretest, posttest, and repeated posttest and then rated each response using the following scale:

7 = Accurate matching of all tones in the pattern, without hesitation.
6 = The child "slid" into one or more of the pitches in the pattern, but eventually sang all accurately.
5 = An exact transposition of the pattern.
4 = The child maintained the general contour of the pattern, but sang incorrect intervals.
3 = The child maintained the general direction of the pattern but not the exact contour.
2 = Responses which ignored the contour of pitches.
1 = Child spoke rather than sang a response or did not respond at all (47).

Electronic Measurement

Several investigators used the aid of an electronic device to record and/or measure the vocal pitch accuracy of their subjects. These devices include the use of the strobotuner,

stroboscope, Johnson Intonation Trainer, Korg Tuner, diplacusimeter, Visi Pitch, and the Digigram MIDI Interface.

Culpepper (1961) used the diplacusimeter. The diplacusimeter was built to measure a pitch deviation of an amount up to five percent, either plus or minus, between the two ears. It consisted of one fixed audio-oscillator and one variable audio-oscillator which was controlled by a tuning control knob. Measurements could be taken at the 250 cycles per second and 500 cycles per second frequencies. The two frequencies were chosen because the fourth-grade child's singing voice was noted by the investigator to normally encompass that range (18).

The diplacusimeter was designed so that the tone from the fixed oscillator was heard in one earphone and the tone from the variable oscillator was heard alternately in the other earphone. A test switch shifted the tone from one ear to the other so that the subject could compare the pitch of the two tones (18).

The volume control dial on the front side of the instrument panel permitted the subject to select the intensity level that he desired. A second set of headphones was provided so that the operator could listen as the subject matched the two tones (19).

The operating instructions for the dipacusimeter were the following:

> The operator set the switches to the 500 cycles per second tone. The subject was seated in front of the diplacusimeter and a set of headphones was used. The operator sat at the side of the instrument with another set of headphones over his ears. The subject was instructed to place his left hand on the tuning control to proceed. He was told to let the operator know when he thought the pitch in the right ear was the same as the one in the left ear. When this decision was reached, the operator recorded the dial reading, changed the switches to the 250 cycles per second tone, set the dial to the five percent mark and told the subject to match those two tones (21).

Cobes (1970) used the strobotuner to measure pitch accuracy. The instrument provided an accurate dial reading in cents (1/100 of a semitone) by comparing the frequency to be measured with internal frequency standards based upon the equally tempered musical scale. The range of the strobotuner encompassed seven octaves or eighty-four half steps in the musical scale. The procedure for measuring the sharpness or flatness of a tone involved moving the vernier control knob and pointer to the right or left until the mobile pattern on the strobotuner disc was brought to a standstill. An audio generator was used for presenting the stimulus tone, as a substitute for the human voice, thereby maintaining consistent standards in the presentation of the stimulus.

The auditory signal was delivered by a speaker since it was felt by the investigator that headphones might interfere with hearing one's voice. On top of the speaker cabinet were two yellow lights, which when lit, were used for reinforcing subjects's responses. At the bottom of the cabinet was a white light that when illuminated, signaled the subjects to emit the note sounded by the audio generator.

Pupils were initially seen in an individual session which lasted approximately five minutes. They were first asked to hum or sing a note that was comfortable for them and which could be emitted without strain. The investigator then determined the scaled pitch of this note with the aid of the audio generator, calibrated to the equal tempered chromatic scale with an A440. If the tone emitted was not a scaled tone, the closest scaled tone was chosen. The pretest task was composed of the following pitches in the given order: one step above, two steps above, one step below, and two steps below the pitch initially emitted (26).

The investigator instructed the subject to listen to a tone that would come from the speaker and as soon as the tone stopped, to emit (using the words "lu," "ah," or "hum") the same note into the microphone. Each of the four pretest pitches was then presented to the subject. The audio generator was used to present the pitch and the strobotuner was used to determine the accuracy of the subject's response. If a subject emitted a tone at least a half step sharp or flat, they were penalized with one uncertain pitch. The subject was classified as an uncertain singer if he emitted at least three uncertain pitches out of four.

The Conn Stroboscope was used by Whitman (1970). Upon completion of all stimulus recordings, tape loops were made for each stimulus pitch and the subject's response to that pitch. These loops were analyzed through the use of a Conn Stroboscope to determine the accuracy of the performance, compared to the stimulus pitch. The sharpness or flatness were recorded in cent deviation.

Richner (1976) defined the accurate singer as one who consistently sang in tune in both individual and group situations. The term inaccurate singer was used in the study to designate a person who sang out-of-tune in both individual and group situations. Inaccurate singers were initially identified as those singers who failed to sing the correct pitches when singing songs with their classmates. They were then tested individually and more specifically identified, for the purposes of the study, as singers who could accurately reproduce no more than twenty of the pitches on a *Pitch Reproduction Test.* The *Pitch Reproduction Test* was designed to utilize the strobotuner, an electronic device that eliminated the need for subjective judgment about pitch accuracy.

Porter (1977) used several electronic devices to measure pitch accuracy. A vocal pitch matching task (VPM) was constructed as follows: the student was asked to sing or hum one tone with which he or she felt comfortable. Then the experimenter determined by stroboscope, the pitch of the note of the closest scaled tone to the sound emitted. The remainder of the pretest segment was composed of the following five pitches played consecutively on the *Johnson Intonation Trainer*: the pitch just emitted by the student, a pitch one step above, one two steps above, one a step below, and one two steps below. After each pitch was played, the student was asked to match the pitch vocally. If the student emitted a tone at least one half-step sharp or flat, or 100 cents deviation, he was penalized with one uncertain pitch. Of the five pitches sung, three uncertain pitches were necessary to meet the operational definiation of an uncertain singer.

Gratton (1992) initially used the subjective judgment of the regular music teachers for his study. Each subject was determined to be an inaccurate singer by their music teachers. Subjects were then tested on a pitch-matching test. Only children who could not vocally match four different pitches out of the five tested participated in the study. This was determined by the subjective judgement of the regular music teachers (17). The equipment utilized in the pre- and posttests included an alto recorder and an electronic *Korg DT-1 Ttuner*. Tests and training stimuli were recorded with a microphone and a cassette deck. Vocal samples for the subjects working with their own voices as stimuli were created with a keyboard sampler. All sounds were presented to students through a portable tape recorder (17).

Flowers and Dunne-Souza (1990) assessed vocal pitch accuracy using a *Korg AT-12 Auto Chromatic Tuner*. Assessment of pitch pattern performance was accomplished by

listening to each tape while viewing a *Korg AT-12 Auto Chromatic Tuner* and determining within plus or minus fifty cents the name of each pitch sung by the child. Two judges independently listened to a random sample of twenty two percent of the total ninety-three tapes, and reliability was determined by comparing the proportion of agreements to total judgments. The reliability estimate was high on all measures of pitch pattern performance: r =.96 for number of intervallically correct patterns; and r =.96 for number of correct melodic contours.

Green (1990) also used a *Korg Auto Chromatic Tuner, Model No. AT-12*, to evaluate vocal pitch responses of subjects. Each subject in the study was tested individually for pitch matching accuracy on three separate occasions, each time responding to a different vocal model. The interval and pitches sung were identical for each of the three tests, the only difference being the model voice. The order of the tests was identical for all subjects: female model, male model, and child model. The three tests were given at seven-day intervals to correct for the variable of tonal memory. After entering the testing room, each subject was told to stand behind the microphone and sing the notes heard on the tape. Subjects were given only one chance to respond to the model voice; however, all subjects were familiar with echo singing and the sol-mi interval because of their Kodaly-based music instruction. Headphones were not used for playing the stimulus voice so as not to interfere with the subject's ability to hear his or her own voice.

At the conclusion of the three tests, all audiotapes containing subjects' respones to each of the three models were heard and evaluated. A *Korg Auto Chromatic Tuner, Model No. AT-12*, was used to evaluate the accuracy of each taped response. Each individual pitch of the interval was evaluated, resulting in six pitches per subject, two for each model.

Deviation, sharp, or flat, from the model pitch stimulus was measured in increments of 100 cents. A pitch response was considered accurate if the deviation from the model pitch was less than 100 cents. The two pitches for each model were then averaged into one deviation score, resulting in three data points per subject, one for each model. Each data point represented either a correct, flat, or sharp response to each model for each subject. Three trained evaluators independently analyzed forty percent of the total response with the Korg tuner. Interjudge reliability was r =.93.

The *Visi Pitch* system was used by various investigators to measure singing pitch frequency. Aaron (1991) used the mean cent deviation of thirty-three cents or greater, measured by the *Visi Pitch*, to define the inaccurate singer. The *Visi Pitch* was also used by Smale (1987) who used the criteria of the mean cent deviation of 100 cents or greater as did Goetze (1985), Goetze & Horii (1989), and Cooper (1995). The *Visi Pitch* was also used by Brown (1988) to measure pitch frequencies of singing range. Brown also used the mean cent deviation of 100 cents or greater to determine the inaccurate singer. Collins (2000) used the *Visi Pitch* to determine vocal range and pitch accuracy using a vocal pitch test. Collins did not establish criteria for the inaccurate or accurate singer. Clayton (1986) used the *Visi Pitch* to measure pitch deviance from a model. It was not the goal of the investigator to classify accurate and inaccurate singers.

Yarbrough, Bowers and Benson (1992) used the *Digigram MIDIMIC, the Opcode Studio Plus 2 MIDI Interface, Performer* sequencing software by *Mark of the Unicorn*, and the *Macintosh SE* computer system to analyze tape recordings of subjects for accuracy. Consistency of digitizing by the MIDIMIC was significant, high, and

positive ($p = .98$, N = 40, $p < .01$). Consistency of tallying correct pitches was measured by submitting twenty five percent of subjects' responses to an independent judge. Agreement between the experimenter and the independent judge was ninety seven percent. Criterion-related validity was obtained by comparing the music teacher's classifications of certain and uncertain singers to scores obtained by computer analysis. For certain singers, criterion scores were seventy percent or better; for uncertain singers, they were less than seventy percent. Percentage of agreement was seventy-seven percent.

Other Tests

Several investigators measured vocal pitch accuracy via the use of ready-made or investigator-developed test instruments. These instruments are described below.

Petzold (1966) used a self-developed *45-Item Test*. The *45-Item Test* had been constructed on the basis of an extensive analysis of more than 500 songs to identify common tonal configurations in major and minor tonalities. This basic test included patterns representing the following common types of musical contours: (1) ascending scale and/or chord; (2) descending scale and/or chord; (3) ascending-descending scale and/or chord; and (4) disjunct patterns. The test was tape recorded to insure uniformity of testing procedures and included the following: (1) a practice session of five trial items which could be repeated as often as necessary; (2) the aural presentation of each item, preceded by the number of the item; and (3) a timed interval of silence following each item that was varied according to item length, during which the child was to sing a response that duplicated the presentation. The test items were presented at a tempo of quarter note = one-hundred-twenty beats per minute. Each child was tested individually

and responses were tape-recorded during the testing situation to facilitate subsequent processing and scoring (20).

Sims, Moore, and Kuhn (1982) used a self-developed pitch-performance test. The pitch performance test consisted of a series of twenty different tonal patterns presented twice, plus one practice pattern. Pre-recorded items were presented to subjects on cassette tape. Each item was sung once for the subjects, who immediately tried to echo the pattern they had heard. The series was presented once by a female mezzo soprano voice and once by a male baritone voice, with half of the subjects responding to the female vocal stimulus first, and the other half responding to the male stimulus first. All testing took place in individual ten-minute sessions.

Items were chosen for their musicality and singability. There were four each of patterns containing one, two, three, and four pitches, and two five-note and six-note patterns. Items of differing lengths were interspersed randomly to form the twenty pattern series. Within the patterns there were twenty-two ascending and twenty descending intervals, ranging in size from minor seconds to one skip of an octave, all within the range from middle C4 through C5 one octave higher. The male stimulus sounded one octave lower. All patterns were sung with one pitch per beat, MM = 68. For each pattern, half of the patterns were sung with words set syllabically to the pitches, while the other half were sung melismatically on a neutral syllable. The Kuder-Richardson reliabity test for responses to the female stimulus was $r = .96$ and $r = .95$ for the male stimulus (105).

Kramer (1985) used an abbreviated form of the *Gould Speech and Song Response Test* to determine the inaccurate and the accurate singer. A singing score was obtained and the

students were classifed as having or not having difficulty in matching pitches vocally. Those having difficulty were classified as inaccurate singers and those not having difficulty matching pitches vocally, as singers (59). The Gould test, formulated for the purpose of determining the ability of a child to match and remember pitches contained within the sections information that helped determine:

1. The ability of the child to understand the concept of high and low pitches.
2. The ability of the child to correlate this concept with his or her singing and speaking voice.
3. The ability of the child to remember a portion of a simple melody. The ability of the child to sing a simple melody with another voice.
4. The ability of the child to match single pitches (160).

Stauffer (1986) used her self-developed tests —(1) *Melodic Echo Test* and (2) *Test of Singing Ability* to determine the vocal pitch accuracy of students. The *Melodic Echo Test* (MET) was an instrument designed by the author to measure the ability to imitate melodic patterns by singing. The test consisted of twenty prerecorded melodic patterns which a subject listened to and then imitated. Each pattern was heard only once. Patterns consisted of three, four, or five pitches within a four-beat rhythmic framework. The subject was allowed four beats in which to initiate a response. The test included patterns in both major and minor tonalities, and various combinations of ascending and descending melodic steps and skips. Administration time for each individual was approximately four minutes. The subject's responses were tape recorded and evaluated at a later time by judges working independently.

The *Test of Singing Ability* (TSA) was an instrument designed by Stauffer to diagnose and measure the singing ability of the subjects participating in the study. The TSA required each subject to sing portions of the melody of five familiar songs. The test was

administered to individual subjects in a five-minute time interval. Test responses were tape recorded and evaluated later by judges working independently.

Apfelstadt (1984) based pitch accuracy ratings on two previously designed tests. Data analysis was accomplished by the investigator, who transcribed all vocal tests onto scoring sheets. The *Boardman Test* was graded for pitch accuracy according to criteria established by Boardman (1964), and the rote songs evaluated for accuracy of melodic contour, melodic interval, and maintenance of tonal center using Ramsey's (1981) criteria.

Montgomery (1988) also used a form of the *Boardman Test of Vocal Accuracy*. The Boardman test, used melodic patterns rather than a single tone and required subjects to sing a melodic pattern after it was presented three times (27). Montgomery transposed the Boardman Test from the original keys to a lower key which encompassed the range of C4 to E5 (28). The syllable "loo" was used instead of text in singing the test patterns (29).

Boardman reported a reliability coefficient of $r = .97$ which was established using a split-half reliability coefficient randomly selected and adjusted for length through the *Spearman Brown Prophecy Formula*. Reliability of the study form of the test instrument was computed on the pretest scores using a split-halves reliability correlation with *Spearman-Brown* adjustment for test length, resulting in a coefficient of $r = .93$ (29).

To produce the test tape, ten patterns were randomly selected and each recorded twice: once by the investigator in falsetto voice, and once in normal voice. The twenty patterns were recorded in random order using a cassette recorder (29).

Teacher Ratings

Several studies have used the classroom teacher or the regular music teacher to make subjective judgments on the singing ability of the subjects. Culpepper (1961) used subjective investigator ratings of tape-recorded samples of students' singing. Accurate singers were defined as students who sang most of the pitches in tune. Inaccurate singers were singers who sang most of the pitches out of tune. Two extreme groups were selected by listening to the tape recordings. The groups included the thirty-five best singers and the thirty poorest singers. Transcriptions of recordings were used to determine improvement in singing.

Bentley (1968) sent out a survey to teachers to identify the number of inaccurate singers who "persistently sing out of tune." It was not clear if the teachers were music teachers or regular classroom teachers.

Greenberg (1970) subjectively identified ten "untuned" singers for his study. Criteria for selection was not specified but stated simply that it included "ten boys classified as untuned singers" (59).

Zwissler (1972) used the classroom teacher initially to judge if a student was an accurate or an inaccurate singer. The investigator performed the final evaluations. In the study, the accurate singers was defined as a first-grade child who was able to correctly reproduce a familiar melody at a given pitch level. The term "correctly" meant that the child "handled his voice accurately enough to maintain tonality in spite of occasional pitch inaccuracy." The inaccurate singer was defined as a first-grade child who could not correctly reproduce a given melody at a given pitch level.

Roberts and Davies (1976) used teacher evaluations to determine the accurate and inaccurate singer. They sent out a survey to prospective teachers and students. The survey discovered 745 boys and 226 girls, who were rated as monotones or droners and whose singing was described as "always completely untuneful with little variation in pitch." Among the children sampled within the age range six-plus to eight-plus, 290 boys and eighty-seven girls were rated as monotones by their teachers.

The study by Jones (1993) used the subjective judgement of the regular music teacher. This judgement was based on the student's ability to echo sing "the natural chant that was used for regular roll call."

Causes of Inaccurate Singing

Knowledge of the potential causes of inaccurate singing is necessary in determining a successful remedial treatment. A review of the extant literature on this topic has revealed several potential causes. Previous research related to inaccurate singing has indicated that there are many variables that may cause inaccurate singing. These include the following:

1. Accompaniment
2. Acquisition of singing ability
3. Age and maturation
4. Attitudes toward singing and choir participation
5. Audiation skills
6. Development of rhythmic and tonal capabilities
7. Group versus individual singing

8. Hearing impairment
9. Musical achievement and Self concept
10. Musical and personal background
11. Neutral syllable versus text
12. Piano instruction
13. Pitch Discrimination
14. Tonal music aptitude
15. Tonality and vocal range
16. Vibrato
17. Vocal modeling
18. Vocal pitch accuracy assessment
19. Vocal range and development
20. Vocal registration

A summary of the methodologies, samples and results of these studies follows.

ACCOMPANIMENT

Clayton (1986) investigated the effect of simultaneous and non-simultaneous pitch stimuli on vocal pitch accuracy of single pitches. Test pitches were selected and presented to the subjects who then responded by singing once with the stimulus tone and once afterwards, as an echo (32).

Sixty-five second grade children from four elementary schools in Bloomington, Indiana participated in the study. All classes received instruction from music specialists during

two twenty-five minute periods per week. Data for fifty-nine subjects were included in the study. Of the fifty-nine children, thirty-three were girls and twenty-six were boys (33).

Stimulus tones were taped and presented on a cassette tape recorder. The investigator chose to use her own voice for the stimulus tape portion of the experiment. All stimulus pitches were sung on an "ah" vowel and subjects were to respond singing the same vowel. For the echo portion of the test, each tone sounded for three seconds with a one-second delay before the subject's response. For the accompanied portion of the test, each tone sounded for seven seconds and the subject joined in on the fourth or fifth second (34).

To avoid recording the stimulus tone during the simultaneous pitch portion of the study, the experimenter used a contact microphone. The investigator noted that a contact microphone responds to direct contact with the vibrating object rather than to acoustical sound waves. The microphone fit around the subject's neck and rested lightly on the throat. The subjects' responses were recorded on a reel-to-reel tape recorder for subsequent analysis (35).

A *Visi-Pitch*, model 6987, was used to evaluate each sung response. The *Visi-Pitch* extracted a fundamental frequency from the complex vocal wave and generated a visual graphic representation of the sound wave. Through the use of a cursor, exact portions of the wave were measured and assigned a frequency in Hertz (35).

The investigator identified the comfortable singing range of children at that age as being approximately B3 to A4 and thus the extremes of this range, B3 and A4 were chosen as two of the four test pitches. The note D4 was chosen as a third test pitch

because the investigator identified this to be the speaking voice pitch for children of that age. The fourth test pitch, F# was chosen as a midway point between the speaking voice range and the singing voice range (35).

The test was divided into two sections, accompanied test pitches and unaccompanied test pitches. Each section began with three practice pitches to familiarize the students with the procedures. The practice pitches were either accompanied or unaccompanied according to the section they preceded. The three practice pitches were C4, G#4, and F4 respectively. Middle C4 was chosen as the first practice pitch because it was near the average range of the speaking voice. It was assumed by the investigator that most of the children would find success in matching this pitch and would therefore feel confident about the rest of the experiment. After this practice, the subjects heard and responded to the four test tones: F# = 370 Hz, B = 250 Hz, A = 440 Hz, and D = 293 Hz. Both the accompanied and unaccompanied portion of the test used the same frequencies for the practice pitches and stimulus pitches. Pitches were presented in the same sequence for each part of the test. To avoid any order effects, half of the subjects were presented the accompanied tones first and half of the subjects were presented the unaccompanied tones first (36).

Students were tested individually in a room adjacent to their classroom. Prior to the test the contact microphone was explained. The researcher asked the students to place their hand on their throat and feel the vibrations as they said "Hi, how are you today?" They were told that a special microphone would be used to pick up the same vibrations they had felt (36). After this, the subjects were told to listen to the vocal sounds on the tape and imitate whatever they heard when the light came on. Subjects were told to hold

their note until the light went out. For one part of the experiment, students were told that they would sing after they heard the voice on the tape, like an echo; for another part of the experiment, they would sing with the voice on the tape (37).

The tapes and reel-to-reel tape recorder were taken to a laboratory equipped with the *Visi-Pitch*. The investigator played the tapes into a speaker and each tone was analyzed by *Visi-Pitch*, providing a visual display of the pitches. The investigator used a cursor to dial to the exact part of the wave to be measured. For each response, the investigator measured the portion of the wave that was sustained the longest. The width of the visual wave representing the pitch allowed for some variance, depending upon whether one measured the upper edge, middle, or lower edge. Every attempt was made by the investigator to dial consistently to the center of the band. The range the width of the band covered was approximately five Hertz. The investigator said that this method was shown to be reliable for establishing a measure of the subjects' responses in Hertz by Goetze (1985) (37).

Each stimulus tone was also analyzed by *Visi-Pitch* and the frequencies were as follows: F# = 370 Hz, B = 250 Hz, A = 440 Hz, and D = 293 Hz. In order to have equal interval data, the frequencies in Hertz were transformed to their natural logarithms. The accuracy of each of the subjects' responses was measured by determining the difference between the logarithm of the model's frequencies and the logarithm of the frequencies of the subjects' responses. Plus and minus signs, which reflected the direction of the subjects' deviation from the model—either flat or sharp, were omitted by changing all logarithms to their absolute values. Overall deviation scores for each of the four test

pitches in both the accompanied and unaccompanied situation were determined in the same manner. All scores were subjected to statistical analyses (38).

Using *Visi-Pitch*, the investigator determined the frequency of the subjects' performance on each of the four test pitches in both the accompanied and unaccompanied situations. Each subject had a total of eight pitch measurements (39).

Means and standard deviations were determined for each pitch. In determining means for the raw scores, all data were left in Hertz measurements. For all other statistical procedures, the Hertz were transformed to their natural logarithms in order to obtain equal interval data. In some measurements absolute values of the logarithms were used to determine the size of the deviation from the model pitch, regardless of the sharpness or flatness. That is, the amount of deviation from the model stimulus was determined without considering the direction of the error (39).

A repeated-measures MANOVA was used for the statistical analyses. A univariate analysis of variance was used to determine the effects of the variables on individual pitches (40).

The mean and standard deviation scores of test pitches in both situations were calculated to determine the effect of the condition on pitch accuracy. Plus and minus scores were used to reveal the direction of error from the model stimulus pitch. A positive score represented an error in the direction of sharpness and a negative score represented an error in the direction of flatness. Both scores for pitch number one, F#4 were negative and when averaged (M = -.045), revealed to the investigator that this pitch was flat in both the accompanied and unaccompanied sections. The mean size of the discrepancy was approximately three-fourths of a half step flat. Both scores for pitch

number two, B4, were positive and when averaged together (M = .078) revealed that this pitch was sung sharp in both the accompanied and unaccompanied situations. The mean size of the discrepancy was equal to slightly more than a half step. Both scores for pitch number three, A4, were negative and when averaged together (M = -.100), revealed that this pitch was sung flat in both sections. The mean size of the discrepancy was almost one whole step. Both scores for pitch number four, D4, were positive and when averaged together (M = .016), revealed that this pitch was sung sharp in both accompanied and unaccompanied sections. The mean size of the discrepancy was equal to approximately one-fourth of a half step (42).

A multivariate analysis of variance was used to determine if either condition, the unaccompanied or accompanied, had a significant effect on overall pitch accuracy. The results revealed that both condition and subject gender had no main effect nor interaction effect on overall pitch accuracy (44).

To investigate additional questions of the study, a univariate analysis of variance was performed on each pitch. Initially, the interaction between condition and accuracy of certain pitches was considered. Other questions were whether certain pitches were sung significantly more in tune and whether gender had any effect or interaction on the accuracy of individual pitches (45).

Conclusions by the investigator were that the results of the study suggested that the effect of accompaniment on vocal pitch accuracy of single tones may be different for children than for adults; that is, younger children sing certain pitches presented singularly more accurately when unaccompanied rather than with a simultaneous stimulus while the reverse may be true for adults. The investigator hypothesized that one explanation for

that was that perhaps the children were attending more to the accompanying sounds than to their own voice. She cited Zimmerman (1971) who showed that young children often focus their attention or "centrate" on the dominant aspect of a perceptual field (51).

The investigator made the following recommendations for further study:

1. A replication of the study using different pitches as stimulus tones to determine if other pitches would yield the same results.
2. A similar study using both younger and older children to help confirm the differences in singing for adults and children suggested by the study findings. The same study could also help determine the age at which children begin matching pitches more accurately as accompanied tones rather than as echoes.
3. A similar study using intervals as the model stimulus as opposed to single pitches to determine if young children match intervals more accurately with a non-simultaneous stimulus than with a simultaneous stimulus (54).

ACQUISITION OF SINGING ABILITY

Guerrini (2002) conducted an investigation in order to gain insight into how children in the fourth and fifth grade acquire singing ability. Specific questions posed by the investigator were: (1) Does singing accuracy vary significantly with the nature of the singing task? (2) What tonal characteristics are most associated with difficulty in developing singing accuracy? (3) What differences are found in children's singing accuracy relative to tonal music aptitude and gender? (9)

The students in the sample comprised the entire fourth and fifth grades in a suburban middle class community in southern New Jersey. There were 174 students--seventy-six fourth grade students and ninety-eight fifth grade students. Gender was reported as being almost evenly distributed, with thirty-eight males and thirty-eight females in grade four and forty-eight males and fifty females in grade five. It was reported that although the racial make up was predominantly Caucasion, minority groups were well represented, including African Americans, Asians, and Hispanics (31).

All students were audio taped at individual, private sessions during which they were given three tasks to sing: (1) melodic patterns, (2) a long familiar song, and (3) a newly learned song. The order of songs and patterns was randomized using a modified *Latin Square* so that there was no order effect. Three judges, using the *Singing Voice Development Measure* (SVDM), by Rutkowski, as the assessment tool, evaluated the performances. Within a group setting, the same students were also administered the Tonal subtest of Gordon's *Intermediate Measures of Music Audiation* (IMMA) (31).

Vocal accuracy in echo patterns and rote songs was compared with tonal music aptitude. Vocal accuracy was examined to find possible relationships or patterns of hierarchy when singing patterns, long familiar songs, and/or recently learned songs. Vocal accuracy was also examined to find possible relationships with tonal music aptitude or gender (31).

Students were audio taped individually and privately and identified by first names only. The same three songs and patterns were sung during each session, but the order of songs and patterns was randomized so that there was no order effect (32).

During the taping session, students echo sang the patterns devised by Joanne Rutkowski in her *Singing Voice Development Measure* (SVDM) while the investigator provided the vocal stimulus, using the investigator's soprano voice. During the same session students sang *America* in the key of F major while the investigator provided piano accompaniment. The third song *Path to the Moon* in G major was chosen because it was unfamiliar to students, challenging, and required children to use their head voice throughout most of the song. Students sang only the first verse of *Path to the Moon* because it contained the same number of measures as *America*. The investigator

accompanied the students on the piano as they sang this third song. The classes rehearsed each song and pattern once a week for a total of three weeks, with audio taping taking place during the fourth week (32).

The same procedures were used in teaching each fourth and fifth grade class. During general music classes, students reviewed the song *America*, played on the piano in the key of F major. Piano accompaniment was provided because the students who scored a "five" on the assessment scale would be placed in the select choral performing group, which was accompanied by piano during the rehearsals and performances (32).

Students also echo sang without accompaniment, the patterns listed as part of the *Singing Voice Development Measure* (SVDM). The investigator's soprano voice was the vocal stimulus that the students attempted to match. The students were not accompanied on the piano while they sang the patterns because the author of the SVDM specified that those patterns be sung without accompaniment (Rutkowski, 1990). The investigator stated that because there was no definitive research of the effectiveness of piano accmpaniment versus vocal modeling, it seemed reasonable to include both (33).

All practice singing was performed as a group within the general music classes, and the investigator encouraged and expected all students to participate. The songs and patterns were reviewed in the same manner during music class the following week (33).

In general music class, students were taught a new song, *Path to the Moon*, in the key of G major. Piano accompaniment was provided and all singing was performed as a group when the students initially learned the tune. Students also reviewed this song during the next two classes (33).

All students were tested individually and audio taped in the general music classroom when no one else was in the room during the fourth week of class. The investigator noted that they did not appear anxious (33).

Students sang both the patterns and the songs during the same recording session. The patterns were echo sung with no accompaniment, while the songs, *America* and *Path to the Moon*, were sung with piano accompaniment. To accommodate for order effect, a modified *Latin Square* was used to decide the order of songs for each of the eight classes (34).

All selections were audio taped and subsequently rated by three qualified judges using the SVDM rating scale. Prior to the official judging, the judges were trained on the use of the SVDM. The investigator said that the purpose of the measure was not to discrminate a fine difference among all singers, but rather to identify singers likely to succeed as opposed to those likely to fail (36).

Judging was done on three occasions, with all three judges present. From the master tape, three separate tapes were made that had students singing only a single song or pattern. By hearing and rating one song or pattern at a time, the investigator hoped to eliminate bias that might be present if the judges heard all three selections consecutively (36).

A judging sheet was prepared for each of the three selections. The sheet was color coded for song or pattern, and the judges were identified as judge "one," "two," or "three." The judges were instructed to write the student's name on the top of each sheet as they heard each testing session, and to assign a score from "one" to "five" to each of students for each song or pattern, using SVDM (36).

Finally, the tonal music aptitude subtest of the *Intermediate Measures of Music Audiation* (IMMA) was administered to all students. The music aptitude test was given in the music room during the general music class time (37).

The investigator made the following conclusions:

1. Many students learn to sing accurately by expanding their ranges first.
2. After the range had been expanded, many students become more vocally accurate by singing short patterns, but may not be able to apply this skill to entire songs.
3. Once students are able to sing one song accurately, using notes above the lift (register break), they appear to be able to sing other songs accurately (56).

The investigator made the following recommendations for further study:

1. Determine the effect of the use of the piano in influencing vocal accuracy.
2. Do a study where the song *America* is divided into two measure segments and echo sung by students in order to discover whether it is the act of echo singing or the nature of the song itself which most influences the accuracy of beginning singers.
3. Find out why some students with high tonal music aptitude take so long to develop vocal accuracy.
4. Find out about the role that motivation plays in vocal development.
5. More longitudinal research following the development of vocal accuracy of students as they progress over time (61).

AGE AND MATURATION

Petzold (1966) conducted a longitudinal study in which he retained the same group of children for a period of several years in order to more clearly identify certain problems and characteristics relating to their musical growth and development (14).

The initial study was completed during the 1959-60 school year and a variety of data obtained for a total sample of 606 children randomly selected from the first six grades (15). Prior to the conclusion of the 1959-60 school year, additional numbers of children were randomly selected from the total population in each of the three lower

grades, and given the same tests that had been administered to the original sample. This enabled the study to keep at least 100 children per group (15).

The investigator designated three groups as follows: Group one included those children who were in grade one in 1959-60; Group two represented the grade two children of 1959-60 with five years of data; and; Group three were the grade three children of 1959-60 for whom he had four years of data for. For Group one, the 102 children with six years of data represented fifty-five per cent of the original group. The greatest loss, thirty-three children or eighteen percent of the total, occurred between first and second grade. Group two showed that ninety-six children completed the expected five years of testing, giving an overall reduction of seventy-one children or forty-two percent of the total. There also, the greatest loss occurred immediately after the initial year with twenty-six children, or almost sixteen percent of the original group, unavailable for further testing. Group three sustained a thirty-six per cent loss, represented by the fifty-six children with less than four years of data. Despite all of those factors, the investigator noted that the original goal of approximately 100 children per group was realized (18).

One of the basic measures utilized throughout the longitudinal study was the *45-Item Test* of melodic perception that had been developed for the pilot study. The reliability coefficients, by grade level, had been determined by the *Kuder-Richardson No. 20* formula. These, together with the standard error, showed that the test had a high degree of stability and internal consistency at all grade levels, suggesting that continued use of the measure could be justified (18).

The *45-Item Test* had been constructed on the basis of an extensive analysis of more than 500 songs which had been carried forward to identify common tonal configurations

in major and minor tonalities. This basic test included patterns representing the following common types of musical contours: (1) ascending scale and/or chord; (2) descending scale and/or chord; (3) ascending-descending scale and/or chord; and (4) disjunct patterns. The test was tape recorded to insure uniformity of testing procedures and included the following: (1) a practice session of five trial items which could be repeated as often as necessary; (2) the aural presentation of each item, preceded by the number of the item; and (3) a timed interval of silence following each item that was varied according to item length, during which the child was to sing a response that duplicated the presentation. The test items were presented at a tempo of quarter note = one-hundred-twenty beats per minute (20).

Each child was tested individually and his responses tape-recorded during the testing situation to facilitate subsequent processing and scoring. All of these tests were given during the first semester of each school year and, although the children remembered having been tested the previous year, none were able to recall any of the test items after an interval of one year (20).

The procedures for processing and scoring the *Pupil Response Tapes* for the *45-Item Test* were briefly summarized as follows:

1. All tests were processed aurally and the incorrect responses to each item entered as tonic sol-fa syllables on the individual data sheets for each pupil. Correct responses, those which duplicated the stimulus, were indicated on the data sheets by a check mark to facilitate scoring.
2. The tests were then played a second time, and scores assigned to each test item on the basis of the scoring system which had been developed. This served not only to verify the accuracy of the initial processing, but the combination of hearing the response while examining its syllabic notation made it possible to more accurately and consistently assign the appropriate score to the several types of possible responses.
3. The scoring system, rewarded correct and partially correct responses in terms of the number of tones that had been reproduced correctly. In

addition, proportionately smaller score values were assigned to responses which retained the general contour and/or length of the stimulus but contained no correct pitches.

4. Every fifth test was independently processed and scored at lest two months after the initial processing as a further check on the accuracy of scoring (21).

The following observations were made by the investigator:

1. All three groups followed a similar pattern in that the differences between boys and girls became larger as the children moved into the upper grades. This pattern was not as consistent when one examined the record of group two.
2. The boys and girls of group one performed with similar accuracy in the first two grades but then began to diverge for the last four years. At no time, however did these differences meet the criterion of statistical significance.
3. There was evidence that the girls in group two demonstrated a much higher level of accuracy each year of the study than did the boys. The entries for group three, grade five and six, also emphasized the fact that the girls, for some reason, were capable of better test performance (31).

The analysis of the data by the investigator for the *45-Item Test* indicated the following:

1. Grade level, at the two-year interval usually produced means that differ significantly. A three-year interval always produced differences that are significant.
2. The girls and boys of groups two and three differed significantly in their ability to perceive and respond to melodic items but this difference was not observed for group one. Although there were no significant interactions between grade level and gender, the data indicated that boys tend to reach a plateau at grade four while girls continue to show improved test performance (84).
3. When the three groups were compared on the basis of their differing grade levels at the start of the project, older children had significantly higher means than first grade children. When the groups were compared on the basis of a common grade level but with differing amounts of practice in taking the test, the younger children had significantly higher means. The group one children showed more rapid improvement taking place in the upper grades than did the children of the other two groups.
4. The competence levels established on the basis of the data showed that a substantial proportion of children remained within the same competence level throughout all years of testing. This was particularly true for the lowest and highest competence levels. The data indicated that although most children stabilized their performance by grade three, the group one

children did most of their shifting of competence levels between grades one and two and grades five and six.

5. The relative difficulty of the test items did not differ significantly between groups or between grade levels, a "very difficult" item for one group or grade level would remain "very difficult" when other groups or grades were considered. Proportionately less improvement was noted for difficult items with older children consistently demonstrating greater competence in coping with more complex items.
6. Analysis of the types of responses that were made by the children showed that they moved form a fifty four percent level of accuracy in grade one to an eighty two percent level by grade six. The probably growth pattern was to eliminate the non-melodic (Types E and F) responses by grade two; that responses showing only awareness of contour and number of tones were almost completely eliminated by grade five; and that partially correct responses became correct responses earlier for easy items than for more difficult items.
7. The difficulty of the item appeared to have a greater influence upon the type of response made by children than did grade level.
8. The data indicated that the greatest improvement took place between grades one and two and that although there was continued improvement it took place at a much slower rate (86).

The construction of the Phrase Test, was summarized as the following. The musical phrases were written with the following criteria in mind:

1. Each phrase, preferably four measures, should be reasonably musical without duplicating any familiar song.
2. Each phrase should utilize those common tonal configurations that had been identified in the analysis of songs materials. Use of familiar configurations in a new context would, it was hoped, permit the child to focus upon the learning task with a reasonable expectation of success (89).
3. Each phrase should encompass a reasonable singing range and include a simple repetitive rhythmic pattern to permit easier recognition, better organization, and greater retention (90).

Each child was tested individually and his responses were tape recorded during the testing session. The procedures for processing and scoring the *Pupil Response Tapes* for the Phrase Test were discussed in the earlier report and were similar to those described for the *45-Item Test*. They were summarized as the following:

1. All tests were processed aurally and the incorrect responses for each trial entered as tonic sol-fa syllables on the individual data sheets. Correct measures were so indicated by a check mark.
2. The test tapes were replayed later to verify the data sheets as well as to score the trials. Scoring was done on the basis of two points for each correct tone and a perfect trial of fifteen correct tones earned a score of thirty.
3. Two kinds of scores, "mean correct" and "rate of learning," were then calculated for the total test.
4. Every third test was independently processed and scored at a much later date to serve as a check on the accuracy with which such scoring was done (92).

The primary function of the task of the Phrase Test was to obtain data regarding the auditory perception of a longer and more complete melodic entity which existed as the musical phrase. The task of perceiving and responding to the larger unit, presented several times in succession, required that the child should be capable of recalling the total stimulus in order to respond accurately. In addition, the child had to be able to recall the nature of his response so that it could be compared with the model when it was presented again (94).

An analysis of the data for the Phrase Test showed the following:

1. Improvement in learning the phrase took place at a constant rate over the four-year period and only a two-year interval produced differences that were significant.
2. The hypothesis of "no difference" between the scores of the group one boys and girls continued to be tenable.
3. Approximately half of the children were able to learn the phrase at least once throughout the four-year period, usually by grade six, but only eight managed to attain this proficiency by grade four and retain that level of performance for the remaining two years.
4. Four years of practice had relatively little influence upon performance on the second task, that those sixth grade children responded with a third grade level of competence. Correlations between scores for the two phrases showed that children performed similarly on both tasks.
5. Correlations between scores on phrase one and the *45-Item Test* showed that a high relationship existed between the two tasks when "mean correct" score, which measured achievement, were considered. However, the correlations between the *45-Item Test* and "RL" scores were much

lower, suggesting that the ability to produce a learned phrase was not strongly influenced by achievement on short melodic configurations.

6. The relationship between the rate of learning and achievement remained reasonably constant from grade level to grade level.
7. Learning a phrase is a more difficult task than responding to individual melodic items and that those children were seemingly unable to recognize their efforts and make the necessary corrections in subsequent trials (114).

ATTITUDES TOWARD SINGING AND CHOIR PARTICIPATION

Mizener (1993) examined the attitudes of elementary music students toward singing and choir participation in relation to grade level and gender, classroom singing activities, previous and current out of school singing experiences, and degree of singing skill, both self-perceived and assessed. A questionnaire designed to investigate relationships between attitudes and the other variables consisted of forty-four items whose response modes varied according to the nature of the item. Items on the questionnaire were grouped into five categories: (a) singing interest, (b) choir participation, (c) classroom singing activities, (d) out of school singing experiences, and (e) self perception of singing skill. Items in the singing interest category of the survey were designed to elicit responses regarding attitude toward singing in general and singing under specific circumstances and to examine certain conditions that might influence attitude toward singing. The choir participation portion of the questionnaire was developed to determine the incidence of interest and membership in a choir and to investigate selected situations that may influence choir participation. In the classroom singing activities section of the questionnaire, items were devised to determine any relationships between singing attitude and selected common classroom music activities, such as singing from basal series books, singing with various forms of accompaniment, singing in conjunction with movement activities, and academic instruction from the teacher about songs. Items in the out of

school singing experiences section were intended to give information on students' perceptions of the extend of family involvement in music and parental participation in musical activites with children. The items in the self perception of singing skill section were developed to provide information on the student's perception of his or her own singing skills and were also used to determine possible relationships between perception of singing skill and attitude toward singing (235).

The questionnaire was read aloud in a classroom setting to the students who had answer sheets on which to respond to the items. Answer sheets were distributed and collected in a single session. A portion of the students were selected to be tape recorded for an assessment of singing skill (235).

Subjects in the study, (N = 542) included third grade through sixth grade students from seven schools in a large urban school district who returned parental permission forms. Of the total, twenty-eight percent were third graders, thirty percent fourth graders, twenty-four percent, fifth graders and eighteen percent, sixth graders. After the collection and tabulation of data from questionnaires completed by the subjects, twenty-three percent of the subjects were selected for a tape-recorded evaluation of singing skill. Each subject from the sample, (N = 123) sang two familiar songs: *Jingle Bells* and a choice of one of several familiar folk songs. For *Jingle Bells*, the starting pitch of E4 was given on a pitch pipe, and the researcher sang the first phrase of the song. The subject then sang the starting pitch on a neutral syllable, with assistance as necessary to match the pitch as accurately as possible. The subject then sang the song without further assistance. After hearing the beginning pitch a second time, the subject repeated the song. Following *Jingle Bells,* the subject chose one additional familiar song from a list of songs thought to

be familiar to most elementary school students. The list included *Twinkle, Twinkle, This Old Man, Yankee Doodle*, and *Are You Sleeping?* The subject sang the selected song beginning at any comfortable pitch (235).

A seven-point set of criteria for rating singing accuracy was designed by the investigator for analysis of singing accuracy. Following is a list of the criteria and respective singing accuracy ratings:

7 = Begins and ends in same tonality, with no loss of tonality within the song and no noticeably inaccurate intervals.
6 = Begins and ends in same tonality with no loss of tonality within the song but with some noticeably inaccurate intervals.
5 = Begins and ends in same tonality but with loss of tonality within the song and some noticeably inaccurate intervals.
4 = Ends in tonality different from the beginning tonality, with some noticeably inaccurate intervals and/or an abrupt shift in tonality within song.
3 = Begins and ends in same tonality but with little pitch variation around the tonal center.
2 = Ends in tonality different from beginning tonality, or ends in spoken tones, with little pitch variation around the tonal center.
1 = Has no clearly established tonal center and most intervals are inaccurate, or is chanted in spoken tones (235).

Analysis of twenty-five percent of the audiotapes provided reliability data. Two independent reliability observers, both experienced in working with singers of elementary school age, rated the tapes. Reliability was calculated by the method of agreements divided by the sum of agreements plus disagreements. Observers achieved reliability levels of $r = 1.00$ for beginning pitches, $r = .98$ for ending pitches, $r = .86$ for melodic accuracy, $r = 1.00$ for rhythmic accuracy, and $r = .84$ for the better trial of *Jingle Bells* for a combined reliability level of $r = .94$ (236).

The results of the study suggested that most students at all grade levels like singing under certain circumstances. Almost all of the sixth graders in the study reported liking to sing with the radio, at least sometimes, in spite of the fact that a large number of them

replied that they did not like to sing in general. Results of the study indicated that most students thought singing was an activity equally suited for boys and girls, but boys may be reluctant to say that they like to sing because in contemporary American society, males are not often encouraged to be recreational singers (241).

It was reported that the most surprising results came from analysis of data from the singing tapes. Only one significant relationship was found between singing skills and the other variables in the study: the relationship between grade level and melodic accuracy in *Jingle Bells*, with fourth and fifth graders having higher accuracy scores than third and sixth graders. It was expected that subjects with more singing skill would have a more positive attitude toward singing, but the results seemed to indicate that singing skill had little influence on attitude toward singing. It was said that the lack of significant relationships between pitch matching performance and accuracy in the performance of the self-chosen song and grade level or gender did not support results of other studies reporting that accuracy for selected singing skills increased as children matured and that girls generally sang more accurately than boys (243). Findings in the study suggested that factors other than singing accuracy may determine a child's attitude toward singing and choir participation (244).

AUDIATION SKILLS

Jones (1993) conducted a study to examine and compare the audiation skills of accurate and inaccurate singers. The study was designed to answer the following questions:

(1) How do audiation skills of accurate singers compare with those of inaccurate singers? (2) What are the implications for the development of aural and vocal skills in grades one through three? (14).

The sample for the study consisted of seventy-two accurate singers and seventy-two inaccurate singers in grades one through three (N = 144). Forty-eight subjects were selected at each grades level to include twenty-four accurate singers and twenty-four inaccurate singers, with twelve males and twelve females in each group. Classification of singers was done by the subjective judgment of the regular music teacher based on the child's ability to echo-sing accurately the natural chant that was used for regular roll call. Subjects were tested in groups of twelve to twenty-four using the tonal and rhythmic audiation tests of Gordon's (1979) *Primary Measures of Music Audiation*. Each test required twenty minutes to administer, with approximately twelve minutes of recorded examples to which the student responded. Administration of the tests was carried out in two twenty-minute sessions that were scheduled on two different days within a two-week period of time (15).

The tonal and rhythm tests each contained forty items. The child had to judge the two examples to be the same or different by drawing a circle around the box with faces that were the same or faces that were different (15).

Recorded test items were made using the *Moog Sonic Six Synthesizer* and a *Moog Rhythm Programmer*. Tonal test items were void of rhythm, and rhythm test items were void of melody. Each item presented one tonal pattern or one rhythm pattern. Tonal phrases were two to five tones in length and were within the vocal range of the child voice. Rhythm patterns were two to seven tones in length (15).

Performance on the two tests were analyzed in order to compare the scores of accurate and inaccurate singers. The *t* test was used to establish any significant differences in tonal and rhythm test scores and in the composite scores in audiation (15).

Results for the study were as follows. Mean scores were computed for the tonal and rhythm tests as well as for composite scores within each grade level. Mean scores for accurate singers in this study were higher than means reported by Gordon for these grade levels. The investigators noted that although mean scores did not reflect a large difference in raw scores, the percentile rankings associated with those scores varie with the grade levels. A mean score of seventy-one was associated with a ninety-one percentile ranking for second grade, whereas the same score at the third grade level was associated with a seventy-four percentile ranking (15).

At the first grade level, mean scores appeared slightly higher for accurate singers on the tonal and rhythm tests as well as on the composite scores. Statistical analysis confirmed the slight differences in scores on the tonal test ($t = 1.94$, $df = 46$, $p < .025$) and the rhythm test ($t = 1.6$, $df = 46$, $p < .05$); however, a marked difference in scores of accurate and inaccurate singers was found in the analysis of scores on the composite test ($t = 2.5$, $df = 46$, $p < .01$) (15).

The difference in the mean scores of accurate and inaccurate singers was reported greater at the second-grade level. There was a significant difference between the two groups on the tonal test, the rhythm test, and the composite test of audiation (16).

At the third grade level, mean scores for accurate singers were reported as being higher than the mean scores for inaccurate singers. There again, they reported a significant

difference in the scores of accurate and inaccurate singers on the tonal test and the composite test; the difference in scores on the rhythm test was not as great (16).

The investigator reported that the statistical analyses of audiation scores support the conclusion that among subjects in the first, second, and third grade, accurate singers included in the study scored higher than inaccurate singers on the *Primary Measures of Music Audiation*. They note that however, the difference in the performance of accurate and inaccurate singers among the first grade subjects included in the study was not as great as the differences found in the second and third grade subjects (16).

The investigator summarized that the main findings in the study were: (a) second and third grade accurate singers scored much higher than inaccurate singers on the tonal and rhythm tests that make up the *Primary Measures of Music Audiation* and (b) the difference in the scores of accurate and inaccurate singers in first grade was not as great. The investigator also suggested that pitch discrimination can be improved through vocal activities and it would therefore, be reasonable to expect that the audiation skills of first grade singers would improve with training in aural and vocal skills. The investigator stated that developmental studies would be helpful in providing some insight with regard to how aural and vocal skills develop from kindergarten through third grade (16).

With regard to rhythmic discrimination, the investigator concluded that it was apparent that at all three grade levels both accurate and inaccurate singers scored lower on rhythmic discrimination than they did on pitch discrimination. It was noted that differences in the two groups were more apparent on the tonal test than they were on the rhythm test. This led the investigators to a final conclusion that aural and vocal

development in the primary grades may focus more attention on pitch discrimination than on rhythmic discrimination (16).

DEVELOPMENT OF RHYTHMIC AND TONAL CAPABILITIES

De Yarman (1972) investigated the following problems; whether:

1. Children who were taught to sing songs only in usual meter performed songs in usual meter better than children who were taught to sing songs in usual, mixed, and unusual meters.
2. Children who were taught to sing only tonal songs performed tonal songs better than children who were taught to sing both tonal and nontonal songs.
3. Kindergarten children who were initially given one full year of instruction in singing songs in mixed and unusual meters would perform songs in mixed and unusual meters better than first grade children who were exposed to one year of typical kindergarten musical training.
4. Kindergarten children who were initially given one full year of instruction in singing nontonal songs would perform nontonal songs better than first grade children who were exposed to one year of typical kindergarten musical training.
5. Level of musical aptitude was a factor in determining the extent to which a child can learn to perform songs in various meters and tonalities (6).

A total of 148 kindergarten and 123 first grade children enrolled in one of twelve public elementary schools in the Iowa City Community School District, participated in the study. The population represented six separate kindergarten and five separate first grade classes. The children were said to be representative of a cross-section of those who attended the public elementary schools in Iowa City in terms of academic achievement and socio-economic status (7).

During the 1968-69 school year, each kindergarten and first grade class received music instruction during three twenty-minute periods each week. All of the music periods were taught by the investigator, an elementary music specialist. Throughout the duration of the study, each class received basic music instruction in which rhythm and tonal activities

were equally stressed. Children in all classes learned to perform songs written in duple and triple meter and they participated in rhythmic activities designed to develop kinesthetic reactions. All children learned to perform songs written in major and minor tonality, and they participated in tonal activities designed to develop aural perception (7).

Eight of the eleven classes, four kindergarten and four first-grade classes, were randomly designated as rhythmic experimental groups. Children in six of the eight groups received varying amounts of instruction in learning to perform songs written in mixed and unusual meters, in addition to usual meter. That is, two rhythmic experimental groups, one kindergarten and one first grade group, learned to perform songs written only in unusual meters. Those two groups were designated as K-A and 1-A, respectively. Children in another kindergarten and another first grade group, designated as K-B and 1-B, learned to perform songs written in usual and mixed meters. And, children in another kindergarten and another first grade group, designated as K-C and 1-C, learned to perform songs written in usual and unusual meters. Children in the remaining two rhythmic experimental groups, one kindergarten and one first grade group, K-D and 1-D, learned to perform songs written in usual, mixed, and unusual meters (7). So that children in six of the eight groups could learn to perform songs written in mixed or unusual meters, they were exposed less to songs in usual meters than were the children in the other two groups (8).

Four of the eleven classes, two kindergarten and two first-grade classes, were randomly designated as tonal experimental groups. Children in one kindergarten group and one first grade group received instruction in learning to perform only tonal songs. Those two groups were designated as K-E and 1-E, respectively. Children in the other two tonal

experimental groups, one kindergarten and one first grade group, K-F and 1-F, learned to perform tonal and nontonal songs. So that the children in two of the four groups could learn to perform nontonal songs, they were exposed less to tonal songs than were the children in the other two groups (9).

Seven songs, taken from a basic music text, were selected to serve as achievement criteria for the study. Each song was eight measures in length and contained approximately the same number of notes (9).

Each of the four songs selected to serve as rhythmic criteria included the following rhythmic patterns of at least one measure in length. They were: (1) primary beats, (2) meter beats, (3) one primary beat plus one primary beat in subdivision, and (4) one-half measure of meter beats plus one-half measure of meter beats in subdivision. Each song was in a different meter—duple, triple, mixed, or unusual (9).

The titles, along with the meters, of the songs used were: (1) *Marching Song*, duple meter, (2) *The Kangaroo*, triple meter, (3) *Swing Your Arms*, mixed meter, and, (4) *The Carpenter*, unusual meter (9).

Each of the three songs, selected to serve as tonal criteria, were written within a similar range and tessitura and included the following tonal patterns: (1) upward leaps, (2) downward leaps, (3) repeated notes, and (4) upward and downward steps. Each song was in major or minor tonality or it was nontonal. The titles of the songs used were: (1) *Friendly Town*, major tonality, (2) *Tira Lira*, minor tonality, and (3) *Two White Sea Gulls*, nontonal (10).

Each of four songs was taught successively during four two-week periods to the four kindergarten and to the four first-grade rhythmic experimental groups. First, children

only listened to the song. Then they echoed each phrase of the song immediately after it was sung by the teacher. Finally, children sang the complete song after the teacher established the tempo, meter, tonality, and beginning pitch. Children learned to chant and clap rhythmic patterns which were extracted from the song. As children were learning the song, they were given an opportunity to perform the song individually and in small groups (10).

Beginning the following month, each of three songs was taught successively during three two-week periods to the two kindergarten and to the two first-grade tonal experimental groups. The teaching procedure for the three tonal songs was the same as that used for four rhythm songs. However, instead of learning rhythmic patterns, children learned to sing characteristic tonal patterns (11).

Evaluation of children's tape-recorded performances of each criterion song was accomplished by the investigator and another elementary general music specialist through the use of a seven-point rating scale (11). Following are the rating scales used for the evaluation of children's tape-recorded performances:

Tonal Achievement:

1 = No correct response
2 = No, or very poor, sense of tonality, but general sense of direction
3 = Poor sense of tonality, general sense or direction
4 = Fair (moderately good) sense of tonality, good sense of direction
5 = Good sense of tonality, very good sense of direction
6 = Very good sense of tonality
7 = Excellent tonal performance (12)

The judges independently recorded their evaluations of the children's performances of the four rhythmic and three tonal songs on separate rating sheets for each song, for each

group. Total ratings for each child's performance were derived by computing the composite ratings of the two judges (12).

The data indicated that children who were taught just usual meter songs did not perform the criterion song in duple meter significantly better than children who were exposed to instruction in mixed and/or unusual meters. The data generally favored the children who were in the experimental groups which received varying amounts of instruction in mixed and unusual meters. Regarding children's performances of the criterion song in usual triple meter, no significant main effects were found for experimental treatments. That is, children in both grades who were exposed only to instruction in usual meters did not perform the criterion song significantly better than children in the experimental groups which received varying amounts of instruction in mixed and unusual meters. Again, as with usual duple meter, the data generally favored children who were exposed to instruction in mixed and/or unusual meters (30).

Although no significant main effects were found among experimental treatments, children who received instruction in both tonal and nontonal music performed, overall, the criterion song in major tonality better than children who received instruction in only tonal music. Regarding children's performances of the song in minor tonality, the results found generally favored children who received instruction in both tonal and nontonal music (30).

The data indicated that first-grade children who received varying amounts of instruction in mixed meter performed, overall, the criterion song in mixed meter better than kindergarten children. Unlike that found for mixed meter, the results were less conclusive for children's performances of the criterion song in unusual meter. That is,

the data indicated that kindergarten children who received varying amounts of instruction in unusual meter performed overall, the unusual meter song as well as the children in the corresponding first grade rhythmic experimental groups (31).

Considering songs in duple, triple, mixed, and unusual meters, level of musical aptitude appeared to have the least influence upon children's performances of songs in duple meter. Unlike that found for duple meter, level of musical aptitude did affect children's performances of songs in triple meter, as indicated by a significant main effect. Based upon the over-all means of children's performances of the criterion song in triple meter, high aptitude children could perform songs in triple meter better than low aptitude children. Results for children's performances of mixed and unusual meters were less conclusive than those for triple meter; however, the means did indicate that, generally, high aptitude children performed songs in mixed and unusual meters better than low aptitude children (31).

Regarding children's performances of major and minor songs and nontonal songs, level of musical aptitude appeared to be more relevant for kindergarten children than for first grade children. Although mean differences for levels of musical aptitude were not statistically significant for children's performances of any of the criterion songs, high aptitude kindergarten children performed, overall, each criterion song better than low aptitude children. Similar results were found for first grade children who received instruction in only tonal music. Opposite results, though, were found for first grade children who received instruction in tonal and nontonal music. That is, low aptitude first grade children in this tonal experimental group performed, overall, each criterion song better than high aptitude first grade children (31).

The investigator made the following conclusions on the basis of the evidence contained in the study:

1. Young children, especially those in kindergarten, would benefit from instruction in songs written in mixed and unusual meters.
2. Instruction in mixed and unusual meters did not hinder kindergarten children's performances of usual meter songs; in fact, it seemed to enhance their performances.
3. Kindergarten children's level of achievement in performing mixed and unusual meter songs was similar to the level of achievement of first grade children who were exposed to one year of typical kindergarten musical training.
4. Although, overall, high aptitude children performed better than low aptitude children, the data suggested that low aptitude children, nevertheless, could profit from instruction in mixed and unusual meters in addition to that of usual meter (32).

The investigator recommended that further evidence be obtained through additional investigations, similar in design to the study, involving a variety of student populations and different music teachers (32).

GROUP VERSUS INDIVIDUAL SINGING

Goetze (1985) investigated the following:

1. What are the relative effects of singing individually and singing in unison on the accuracy of primary grade children's singing?
   a. How does the effect of individual versus unison singing vary with grade levels?
   b. How does the effect of individual versus unison singing differ among boys and girls?
   c. What percentage of children sing accurately in individual singing but not in unison, and what percentage sing accurately in unison but not in individual singing?
2. What effect does the use of the neutral syllable "loo" have on the accuracy of primary-grade children's singing as compared to their accuracy when singing with text?
   a. How does the effect of singing with "loo" versus text vary with grade level?
   b. How does the effect of singing with "loo" versus text differ among girls and boys?

c. What percentage of children sing accurately on the syllable "loo" but not on text, and what percentage sing accurately with text but not on "loo"? (4)

The subjects for the study (N=165) were first and third graders who were from randomly selected classes in three elementary schools in Boulder and Denver, Colorado. All schools had a music specialist who provided music instruction at least once a week to the first grade and third grade classrooms. The investigator reported that the kindergarten in School Three had music with the specialist for one fifteen-minute period per week while kindergarteners in Schools One and Two had no instruction from a music specialist (63).

The two singing tests for the study consisted of two melodic phrases of equal length, range and difficulty sung on the neutral syllable "loo". One test was sung as an individual "echo" response to the investigator. The other test was sung in unison with a small group (10).

Goetze reported that the unison singing task was designed to approximate classroom singing of songs and so the task phrase was incorporated into a four-phrase song which had "aaba" form. The task phrase, "a", was repeated three times. The second "a" was used to assess the child's performance, giving the subjects one phrase to get started. If a child failed to complete the second phrase, the fourth phrase (also "a") was used instead. The third phrase ("b") was contrasting in text and melodic direction (68).

In the individual singing task each child in turn was asked to play the "echo game" with the investigator whose voice served as the model for all responses. A pitch-matching example on the descending minor third starting on A4 was used for preparation and all subsequent responses were recorded. To minimize the likelihood of one subject

influencing another's response, the order in which children were asked to respond was varied. For example, all were asked to echo-sing with "loo" in one order (1, 2, 3) then when asked to sing with text, they would singing in a different order (i.e. 3, 1, 2 or 2, 1, 3, etc.). All children were verbally reinforced through words of encouragement after each response regardless of their success on the task (71).

The procedure for the unison sing task follows:

1. After the three subjects completed the individual singing tasks, the next three subjects were admitted to the room and asked to help the group sing the song. This served as review and preparation for those who had just arrived while the first three were recorded.
2. Following this the first group of children was dismissed and those who had just arrived were fitted with microphones and the procedure continued as above.
3. The total time required for each group was eight to ten minutes.
4. Children were asked to sing under four conditions:
   a. Individual singing with text
   b. Individual singing with "loo"
   c. Unison singing with text
   d. Unison singing with "loo."
5. Because it was hypothesized that unison would be less accurate, the taping of individual singing always preceded the taping of unison singing in the procedure to allow more time to adjust to the setting and to practice singing.
6. The order in which the subjects were asked to sing with text and "loo" was alternated. Approximately half of the subjects sang all melodic examples first with text and then with "loo" while the other half sang first with "loo" then with text (72).

A *Visi-Pitch* device was used to measure each tone of the recorded responses in Hertz. The investigator reported that this was a device used by speech pathologists, audiologists, linguists, and psychologists as a teaching device and a tool to analyze voice, intonation, accent and emotion (65). It was also reported that the *Visi-Pitch* extracted cycle-to-cycle fundamental frequency from the complex vocal waveform and displayed in real time the frequency curve, the image of which could be retained indefinitely on the screen. The

investigator explained that a curser was engaged to determine the frequency of any portion of the frequency curve on the display screen (65).

The singing responses were recorded by individual contact microphones. The microphones were fitted on the neck at the level of the sternal notch in a neckband designed to be as comfortable and unobtrusive as possible (65). The investigator explained that a single track of the recording was played into the *Visi-Pitch* which then supplied a visual display in graphic form and numerical values in frequencies. A method for automatic analysis of voice fundamental frequency and intensity using *Visi-Pitch* and a computer with analog-to-digital converter was used (65). The *Visi-Pitch* produced a mean frequency in Hertz for the analyzed pitches (66).

The investigator used the following procedure to assign frequency scores to the individual subjects:

1. The tapes were taken to a laboratory equipped with the *Visi-Pitch* and a tape player.
2. The investigator played the tapes into a speaker and the sound was analyzed by *Visi-Pitch*, providing a visual display of the contour.
3. By moving the curser along the contour a frequency for each pitch unit of the singing tasks (corresponding to each syllable or not value) could be determined.
4. The width of the band representing the pitch curve allowed for some variance, depending upon whether one counted the upper edge, or center and so on.
5. Every attempt was made to consistently dial to the center of the band.
6. In some cases there was slight variation within a single pitch segment and an average frequency had to be estimated.
7. In others the beginning of the pitch would have a slight surge, then would settle at one level. In these cases, the frequency of the sustained portion was recorded (73).

Two forms of reliability were employed by the investigator. A second judge was taught the procedure for using the *Visi-Pitch* and independently evaluated thirty randomly-selected examples. After a ten-month period the investigator reevaluated the same thirty

examples. A correlation on the judgments was performed (74). The interjudge correlation coefficients for individual pitches ranged from $r = .94$ to $r = .97$ (78).

A second test of reliability was performed using twenty-seven randomly selected pitch units which were analyzed by a computer. After conversion to digital form, a computer program produced as output a mean frequency in Hertz for individual vocal periods which had been identified manually (74). A correlation of the frequencies from the investigator's original evaluation of the melody and the means determined by this method yielded coefficients ranging from $r = .96$ to $r = .99$ and mean $r = .98$ (80).

The investigator reported that in order to have equal interval data, the frequencies in Hertz were transformed to their natural logarithms. Using the computer method described in the second form of the reliability above, an average of the frequencies of five randomly selected performances was used to determine the model's frequency for each tone of melodies. Only slight variation in frequencies was noted by the investigator finding that plus or minus ten Hertz, was less than a quarter tone (25 cents). The average frequencies for the model were reported as A4 = 442 Hz, F#4 = 370 Hz, E4 = 323 Herz, and D4 = 239 Hz. Subjects' scores were based on the difference between the logarithm of those frequencies and the logarithm of the frequencies determined by *Visi-Pitch* for each subject (75).

The investigator stated that a careful consideration of the melodic direction and observation of the subjects' performance led to determining the first score which was the measure of pitch accuracy. Goetz said that subjects who did not sing accurately often droned within a narrow band of pitches near the middle to lower part of the pitch range of the melodic task (D4-F4). She added that when such a subject's deviation on the notes in

the unidirectional melodies were averaged, the score was found to be less descriptive of the student's accuracy than an average of selected notes. She elaborated that a hypothetical subject who might have sung a single pitch E4 (monotone) within the range of the song would have scored a deviation of 266 cents while using only the highest and lowest frequencies, D4 and A4, this same performance would be scored with a deviation of 400 cents. It was reported that both the melody used for unison and the one for individual singing contained two instances of the highest note, A, and one of the lowest note, D (76).

Goetze reported that plus and minus signs, which reflected the direction of the deviation, were omitted to eliminate the chance of one score canceling out another. She elaborated that the average of the deviations of the subject from the model on those three notes of the melody served as the first score, which was referred to as the "Pitch Measure" (77).

The investigator also utilized a second score called the "Contour Measure" which utilized all of the notes in the melodic patterns. In this case, the plus and minus signs were retained. A standard deviation of the subject's deviation on the notes within a pattern was calculated. That number was said to reveal how constant or erratic the subject's deviations were. For example, if all deviations equaled 100 cents, the standard deviation would have been 0.00. If the deviations were unequal, for example, from 200 to 800 cents, it would have been much larger. It was thus explained that "accuracy" on this measure was determined with a standard deviation near 0.00, indicating that the student sang the same contour as the model regardless of the pitch level at which was performed by the subject (77).

It was reported that each subject was given a Pitch Measure and a Contour Measure score for their singing performances under each of the following conditions: singing individually with text (IT); singing individually with "loo" (IL); singing in unison with text (UT); and singing in unison with "loo" (UL) (80).

The data was analyzed by a comparison of means using Analysis of Variance, Cross-tabulation and Chi Square (10). A low score on the Contour Measure revealed a response which approximated the same contour as the model, regardless of pitch level (80).

Goetze's results indicated a consistent pattern of more accurate performance in individual singing than in unison singing at all grade levels. A significant difference was indicated for the entire sample and for kindergarten and first grade. The investigator noted that even though the difference in the means at the third grade level was not significant, their performances in unison, like the younger subjects, were less accurate than in individual singing and thus contributed to the significance of the entire sample (83).

Data for the means for the four subgroups, IT, IL, UT, and UL was presented along with the difference between the means for singing with text and the means for with "loo." The comparisons revealed that the mean difference in individual and unison singing was significant for the entire sample for singing both with text and with "loo." It was added by the investigator than when analyzed by grade level, the difference was significant only within Grade One in singing with "loo." It was noted that subjects at all grade levels consistently sang with more accuracy in individual singing (85).

A comparison of the means in individual and unison singing on the Contour Measure was reported. The investigator reported that the results followed a similar pattern to the

results of the Pitch Measure. Subjects sang with greater accuracy in individual singing than in unison singing; that is, the subjects sang with a more erratic or contrasting contour in unison than in individual singing. It was reported by the investigator that for the entire sample and for kindergarten students, the differences between the means of the two conditions were shown to be significant. It was also reported that at the first and third grade level, the mean on the Contour Measure for unison was not significantly different from the mean for individual singing (86).

The investigator reported that the means of the four sub-groups revealed that subjects performed with significantly greater accuracy in individual singing with text. Goetze said that for kindergarteners, the difference between individual and unison was significant when singing with text, but not when singing with "loo." She noted that this did not mean that they were more accurate in unison singing with text than with "loo" because with "loo" the performance was more accurate (88).

The overall effects of text and the neutral syllable "loo" were determined by combining the subgroups of unison and individual singing. The investigator noted, however, that when grouped by grade levels, the largest and only significant difference was at the kindergarten level. It was also reported that when broken down into subgroups based on individual and unison singing, differences in means were not significant except for the entire sample and for first graders in individual singing. The investigator added that the subjects sang more accurately with "loo" than with text in all but two of the instances: in unison singing at the first and third grade levels (90).

Goetze created four levels of Pitch Measure scores: 1) deviations of 0 to 99 cents; 2) 100 to 199 cents; 3) 200 to 299 cents; and 4) over 300 cents (101). Cross-tabulation analyses revealed a pattern. For the entire sample, and at the first and third grade levels, the numbers of children who sang with a deviation of less than 100 was slightly higher with text than with "loo." For those who sang with a deviation of 100 to 200 cents, however, a consistently higher number sang more accurately on "loo." She noted that it appeared that for some subjects who sang inaccurately with text (beyond 300 cents), "loo" may have served as a vehicle for attaining a more accurate pitch level (115).

Goetze summarized the following findings:

1. The subjects sang more nearly at the pitch level and with more accurate contour when singing individually than when singing in unison.
2. The subjects sang more nearly at the pitch level when they sang with the neutral syllable "loo."
3. The subjects sang most accurately individually with "loo."
4. Kindergarten and first grade subjects seemed to benefit more from singing with "loo" than third grade subjects, that is, the use of "loo" rather than text improved the performance of the younger subjects.
5. In general, more third graders sang accurately than kindergartners or first graders.
6. Girls sang at a more accurate pitch level than boys; the difference was most pronounced in unison singing.
7. The difference between boys' individual and unison responses was greater than that of the girls; that is, boys were affected more by the presence of other voices than girls (123).

The investigator made the following recommendations for music teachers:

1. Children in kindergarten and first grade be given music instruction which includes singing activities (138).
2. Opportunities for individual singing responses be incorporated into classroom activities from the very beginning of instruction.
3. Caution teachers against premature classification of singers as monotones or inaccurate singers.
4. Teachers should allow children to develop unison singing skills and to refrain from discouraging slow developing singers by eliminating them from musical activities.

5. A unison singing component should be included within the audition process to determine readiness for choral singing (139).
6. Boys may require more assistance in attaining unison singing than girls.
7. Kindergarten and first grade students would seem to benefit from a regular routine of singing individually with "loo."
8. Singing without text may be an essential technique for teaching young children to sing (140).

The following were recommended by the investigator for further study:

1. A replication of this study is needed in order to determine if similar results are attainable with a different sample. In order to determine a more specific sequence of skill acquisition, such a study would ideally be conducted longitudinally over six years. Where that is not possible, a cross-sectional study which includes more grade levels is necessary to determine such aspects of singing development as the age at which most children attain success in individual singing, and the age at which most children attain unison singing skills. This additional evidence would contribute to determining whether there is an age when singing with text is equally as or more effective than singing with "loo." Also whether contour and pitch level accuracy develops simultaneously.
2. Research investigating the roles of both auditory and kinesthetic feedback as it relates to the skill of monitoring in singing with children is needed. There is also a need to determine the causes of inaccurate singing.
3. Investigate teacher evaluation of children's singing to provide insights into those characteristics of singing which influence teacher judgment and teacher perception of pitch and contour deviation within unison singing.
4. An experimental study comparing techniques which facilitate the acquisition of unison singing.
5. Research is needed which relates unison and individual singing to hemispheric specialization, hemispheric domination, cognitive styles, and gender differences (141).

Smale (1987) investigated the pitch accuracy of four-year-old and five-year-old singers. Subjects for the study consisted of 106 children. The study focused on the existing singing ability of the participants. Only pitch information was investigated to determine pitch and contour accuracy (28-29).

The song taught to the children was designed to be appropriate for young singers in interval content and range (as determined in the survey of literature), and to be similar to

songs recommended for early childhood classrooms. A phrase was constructed spanning the range of a fifth (C4 to G4). It included the interval of a third ascending and intervals of thirds and seconds descending (Sol-Mi-Sol and Mi-Re-Do). It was in duple meter and used only quarter and eighth-note values. It had only six tones (four different pitches) and took less than four seconds to sing. In order to balance the song musically, a second phrase was added, using small intervals in a descending pattern. Performance on the second phrase was not analyzed (30).

Participant and investigator voices were recorded simultaneously using individual contact microphones attached to an adjustable neckband allowing neck contact at the level of the sternal notch. This allowed a clear separation of each child's voice from that of the investigator in unison singing. Participant and investigator voices were recorded simultaneously on different tracks of a cassette using a stereo tape recorder (30).

The recordings were analyzed using a *Visi-Pitch* patched into an *Apple II* computer. This device extracted the fundamental frequencies of each child's singing voice, sampling the recorded voice every millisecond as it was played through the auxiliary input channel and, at the same time, displaying the frequency curve on its monitor. Data from each participant's performance was stored in the buffer memory, enabling the frequency curve to be retained on the screen and the frequencies to be analyzed (31-32).

The investigator taught a song to each class, then recorded the children individually performing four singing tasks: (a) singing the first phrase as an echo to the investigator; (b) singing the phrase in unison with the investigator; (c) singing the phrase as an echo, but using the syllable "loo" instead of the text; and (d) singing the phrase in unison with the investigator, using the syllable "loo." The frequency of each recorded tone was

determined and pitch accuracy scores, both pitch level and contour, were assigned to each child's responses on each of the four tasks (32).

Tape segments were collected for each child for the four vocal tasks. Incomplete recordings and recordings of insufficient gain to permit use of the *Visi-Pitch* instrumentation (i.e., poor signal to noise ratio) were rejected from analysis (34).

Each recorded segment was analyzed to determine the frequencies of each of its six tones. The image of the curve of one singing task was retained on the screen while two user-controlled cursors were moved to define the portion of the curve that represented one tone the child had sung. The *Visi-Pitch* automatically calculated the average frequency of that cursor-delineated portion of the curve. The cursors were then moved to define the portion of the curve that represented another tone, until all six tones had been evaluated to establish their average frequencies. In this way, frequencies were assigned to each of the six tones of the melody for each of the participants (34-35).

Two measures of the pitch accuracy of each child's response were calculated. A "pitch score" represented the average deviation of the child's pitches from the investigator's model. It reflected how closely the child approximated each individual pitch of the model. A "contour score" represented the average deviation of the deviations of the child's pitches from the investigator's model. It reflected how closely the child approximated the shape of the melody, regardless of pitch level. The pitch score was calculated by determining the mean of the absolute values of the logarithms which represented the intervals between each child's performance and the investigator's model. The contour score was calculated by determining the standard deviation of the absolute values of those logarithms (36).

Eight scores were calculated for each child: a pitch measure and a contour measure for each of the four conditions. The scores were expressed in cents, a measure which divides the semitone into 100 equal parts. These scores were subjected to a statistical analysis (36).

Another musician in the program was trained in the use of the *Visi-Pitch* and independently determined frequencies for the same forty-eight recorded segments. The frequencies determined by the original evaluation, the investigator's re-evaluation, and the second judge's evaluation were compared using the *Pearson Product Moment* correlation. The correlations obtained were $r = .99$, $r = .99$ and $r = .99$. Mean correlation was $r = .99$ (36).

The most significant conclusion determined in this study was that children sang more accurately alone than in unison with a vocal model. The mean pitch deviation scores were sixty cents higher on the unison tasks than on the individual; the mean contour scores were nearly forty cents higher (72).

The direction of this increased deviation was consistent according to the frequency means. They indicated that the children actually sang at a lower pitch level when singing in unison compared to singing individually. Some children sang more than 100 Hertz lower on their unison tasks than they did on their individual tasks (72).

The investigator made the following conclusions:

1. Children ages four and five sing with more accurate pitch level and contour when singing individually than when singing in unison with another singer.
2. The use of the neutral syllable "loo" rather than text does not significantly effect the accuracy of pitch level or contour of four-year-old and five-year-old singers.
3. Boys and girls ages four and five are similar in vocal pitch accuracy (72).

The investigator made the following recommendations for further study:

1. That a replication of the study be done to determine if similar results are attainable with a different sample.
2. That the participant pools in further studies be broadened to include children of different socio-economic backgrounds and children with different levels of musical experience so that influences of environment and experience can be investigated.
3. That the participant population in further studies be broadened to include children older than five years old and young than four years old so that developmental influences can be investigated.
4. That longitudinal studies be undertaken to provide further insight into developmental changes in children's response to the factors that affect their pitch accuracy through the preschool and elementary years.
5. That relationships be investigated between the use of neutral syllables or text, the frequencies at which children sing, and the pitch and contour accuracy children display.
6. That further investigations be made that compare children's reproduction of pitch level and contour, preferably longitudinal studies or cross-sectional studies that include a wide age range (73).

Cooper (1993) investigated selected factors related to children's singing accuracy. The purpose of the study was to investigate children's singing accuracy as a function of the presence or absence of a unison accompanying voice, pitch discrimination ability, gender and age (222).

A sample (N=169) of first-grade through fifth-grade subjects for the main study was from a large urban elementary school in Akron, Ohio. Subjects were tested individually, and taped responses were analyzed with *Visi-Pitch* technology. The vocal model was a tape-recorded child's voice singing a four-beat melodic pattern on the neutral syllable "loo." For individual singing, subjects echoed the model; for unison singing, subjects echoed the pattern simultaneously with the model (223).

A *Visi-Pitch* with computer interface was used to determine the frequencies of each recorded stimulus and response pitch. The *Visi-Pitch* sampled the recorded voice and

displayed the frequency curve of the criterion pattern on the computer monitor. The user then moved cursors to outline the segments of the curve representing the pitch to be analyzed, and the *Visi-Pitch* automatically calculated the frequency in Hertz of the pitch area between the cursors (225).

Because frequencies in Hertz are not equal-interval data, logarithms were used to calculate the interval or deviation in cents (100 equal cents = one semitone) between each response pitch and its corresponding stimulus (Campbell & Greated, 1987). For each subject, five such cent deviations were calculated for the individual condition and five for the unison condition (225).

The mean of each subject's five cent deviations for the individual condition was his or her individual *Vocal Pitch Accuracy* (VPA) score; similarly, the mean of the five cent deviations for the unison condition was each subject's unison VPA score. Absolute values were used in these calculations to avoid the possibility of positive and negative cent deviations for example, sharp and flat responses, respectively, on different pitches within the pattern canceling each other out. Therefore, although VPA scores represented overall deviation from the model, they did not provide an indication of the direction of deviation or contour of the response. Because VPA scores represented divergence from the model, lower scores indicated more accurate performance (225-226).

To determine reliability of the *Visi-Pitch* frequency evaluations, the investigator re-evaluated the responses of thirty-five randomly selected subjects six weeks after the first evaluations were made. The same thirty-five responses were independently evaluated by a second judge who had been trained in the use of the *Visi-Pitch.* Pearson correlations

resulted in coefficients of $r = .97$ and $r = .99$ for intrajudge and interjudge reliability, respectively (226).

For the entire sample, the mean for unison singing was slightly lower than that of individual singing. In both conditions, the lowest means, representing the most accurate singing, were reported for fourth-grade. For both fourth-grade and fifth-grade students, the mean cent deviation in both conditions was within a semitone, or 100 cents, and the standard deviations were relatively small, indicating that those grades were more accurate and more consistent than others in VPA performance. The highest mean scores, representing the least accurate singing, were reported for grade three in both conditions, followed by grades one and two (226).

A repeated-measures *MANOVA* was performed to examine individual and unison VPA means in relations to the between-subjects factors of grade and gender, and the within-subjects factor of singing condition. The within-subjects factor of singing condition was not significant, nor was its interactions with any other factors; however, grade emerged as a significant between-subjects effect and accounted for eleven percent of the variance (227).

No significant difference between individual and unison vocal pitch accuracy was evident in the present study. The investigator noted that among the factors that could have contributed to this discrepancy was the choice of vocal model, since the study was the first to use the child's voice recorded on tape as the stimulus (228).

The investigator hypothesized that short-term memory requirements dictated by pattern length and complexity may have been another contributing factor. The study used a four-beat melodic pattern for the vocal task (228).

Differences in the vocal register were also noted to be contributing factors to the discrepancy of results. According to the investigator, the criterion pattern in the study was pitched entirely below the register shift that generally occurs around G4 or A4, whereas other studies used melodic content requiring a shift into head voice (229).

The investigator defined the accurate and inaccurate singer as: Subjects with a VPA score within a semitone, or 100 cents, in both conditions were classified as accurate, while subjects with a VPA score of 100 cents or greater were classified as inaccurate (229).

The investigator made the following conclusions based on the study:

1. Singing accuracy varies widely within and across grade levels. Teachers may expect, therefore, to encounter a wide range of singing abilities, from highly accurate to highly inaccurate, in any given class.
2. While older children, for example, fourth-grade and fifth-grade students in the study, may generally be expected to sing more accurately than younger children, improvement may not occur progressively at successive grade levels in a predictable pattern.
3. Differences teachers may observe in their classrooms between boys' and girls' singing accuracy should not be attributed to gender alone. Other factors, such as motivation, peer pressure, or control of the vocal mechanism may be contributing to the perceived differences (230).

The investigator made the following recommendations for future study based on these results:

1. A replication of the study extending the methodology to accommodate comparison of several melodic patterns.
2. A study including two pitch levels of the criterion pattern, for example, both head and chest registers, would facilitate comparisons of singing accuracy by pitch level, condition, and the interaction of pitch level and condition.
3. Compare different levels of measurement on which accuracy judgments are based (230).

Green (1994) investigated the effects of unison singing versus individual singing on the vocal pitch accuracy of children in grades one, two, three, and five, as well as to observe the effects of gender and maturation on children's singing ability (105).

Subjects for the study were 241 students enrolled in grades one, two, three, and five at one inner-city elementary school during the 1991-1992 school year. The number included the entire population of the grade levels with the exception of fourteen students who were eliminated from participation in the study because of absences or difficulty in data-collection procedures. The particular school population reportedly represented a wide variety of socioeconomic levels, racial and ethnic groups. No attempt was made to determine musical aptitude or ability prior to the time of the study. All subjects received regular music instruction from a music specialist (108).

Before beginning the data collection phase of the study, the researcher visited the subjects' regularly scheduled music classes to become acquainted with the subjects and to acquaint the children with song material and the general procedure to be used in the data collection process. The children's pentatonic game song *Bow Wow Wow* was taught to each class using rote song teaching procedures, and was practiced until each class could sing the song independently of the teacher. A starting pitch of D4 above middle C4 was used, which resulted in the range of the song being from D4 above middle C4 to B4. Procedures to be used in data collection were also explained and demonstrated to the children at the time of the researcher's initial visit (108).

At the beginning of each data-collection session, the song was reviewed and practiced for the students participating in the study during that particular session. After reviewing the song, subjects were randomly selected to come in groups of four to a separate room

adjacent to the music classroom for recording of student responses. Subjects were chosen in groups of four due to the limitations of the recording equipment. For the initial group of subjects, eight students were selected so that there would be eight children present for the group-singing task, even though only four would be recorded initially. Subsequently, subjects came to the testing room in groups of four, and the four subjects who just completed the recording remained in the room to participate in the group-singing task for the next four subjects. Subjects were seated in chairs placed in two rows of four chairs each. Each of the students in the first row was given a miniature cardioid fixed-charge condenser microphone and was reminded where and how to hold the microphone. The responses of those four subjects were recorded individually on separate channels of a tape recorder (108).

Each subject was asked to sing the song both individually and simultaneously with the entire group of eight subjects. The order of the task presentation—singing alone and singing with the group—was alternated with each group of four subjects. For each singing task—individual and group, the researcher sang the entire song, with a starting pitch of D4, and asked the subjects to echo the song back into the microphone. The researcher did not sing with the subjects during the recording of responses for either group or individual singing (109).

At the conclusion of the data collection process for all 241 subjects, the audiotapes containing subjects' responses were heard and evaluated for accuracy by the researcher. The song used in the study, *Bow Wow Wow*, contained seventeen pitches and sixteen melodic intervals. Subjects were given credit for singing correct melodic intervals even if the actual pitches sung were incorrect. Therefore, a possible perfect score of thirty-

three could be achieved if all pitches and intervals were sung correctly. If a subject sang the correct melody but at an incorrect pitch level, a score of sixteen was earned. Each subject was given two scores, one for the individual singing task and one for the group singing task, with each score reflecting the number of pitches and intervals sung correctly out of a possible score of thirty-three. A trained evaluator independently scored twenty-five percent of the total responses. Reliability, determined by agreements divided by the sum of agreements plus disagreements, was reported at $r = .94$ (109).

Data was analyzed using a three-way analysis of variance (ANOVA) with repeated measures (grade x gender x performance condition). Results indicated the following:

1. A significant difference due to gender was reported; out of a possible total of thirty-three correct pitches and intervals, the mean score for females was M = 16.89, which was significantly better than the mean score for males, which was M = 13.22 (109).
2. A significant difference due to grade was reported; Fisher's test of least significant difference was computed on grade-level means. Fifth-graders scored significantly highest while gradual though insignificant improvement occurred between grades one and two and between grades two and three. Students in grade three scored significantly higher than did those in grade one (109).
3. A significant difference due to performance condition was reported; Out of a possible total of thirty-three correct pitches and intervals, the mean score for the group singing task was M = 16.42, which was significantly better than the mean score for the individual singing task, which was M = 13.83 (110).
4. A significant interaction between performance condition and grade was reported. As indicated, scores for both group and individual singing improved with each grade level. Improvement was gradual for both individual and group singing for grades one, two, and three. However, there was a more noticeable improvement in group singing scores between grade three and grade five than in individual singing scores for the same grade levels (110).

The investigator made the following conclusions:

1. Singing in unison and singing individually did have an effect on the vocal accuracy of children in grades one, two, three, and five.
2. Children at each of those grade levels sang more accurately when singing in unison with their peers than when singing individually.
3. Both group and individual scores improved with each successive grade level.
4. The difference in scores between groups and individual singing was much greater for children in grade five than for those in the other three grades.
5. In each of the grade levels and for each of the singing conditions, girls sang more accurately than boys.
6. A child would have difficulty singing simultaneously with an adult voice, which has a completely different timbre than does the child's own voice.
7. Conversely, a child may find it easier to hear himself or herself and to sing accurately in unison with voices exhibiting a timbre similar to that of his/her own voice, which could be a possible explanation for the findings in the study (111).

The researcher recommended that future investigations should be conducted to study placement of a weak singer within a group of strong singers, as well as placement of younger singers with older singers to ascertain those effects on children's singing accuracy (113).

HEARING IMPAIRMENT

Culpepper (1961) conducted a study of the hearing impairments in defective singers. The purpose of the study was to investigate defective singing among 113 fourth-grade students in a public school of Macon, Georgia. The four elements of hearing that were examined included diplacusis, hearing acuity, pitch discrimination, and tonal memory. The term diplacusis was defined by the investigator as follows. "Diplacusis is an anomalous characteristic of hearing in which a fixed tone is heard at two different pitches in the separate ears" (121). Deficiencies in any one of those hearing factors that could be associated with defective singing were considered as a hearing impairment for music.

Comparisons were made of the scores of the students who sang a majority of notes on pitch with those who did not in order to determine the relationship between the various hearing impairments and defective singing (2).

The investigator also analyzed the characteristics of defective singing. This phase of the experiment was accomplished by subjectively transcribing into musical notation the tape-recorded singing of the entire group of 113 students used in the experiment. A study was made of those transcriptions as a means of determining any factors causing the singing to be defective. The investigation also included daily observations of a randomly selected group of fifteen extremely defective singers in order to discover whether or not there was evidence of any behavior pattern being followed as a student learned to sing in tune (3).

Audiograms were made, diplacusis readings were taken, and the individual voices of all 113 students were recorded on tape as they sang two of the songs they had been learning in class. The songs used were *Tadpoles* and *My Partner*. Both songs were taken from the *New Music Horizons, Book Three* series (15).

The audiometer, diplacusimeter, record player and the tape recorder were set up in a sound-proof room. Students reported individually to the investigator for this phase of the testing. The hearing test was administered first. The operation of the diplacusimeter was explained next and following this, the student attempted to match the tones at both the 500 cycles per second and 250 cycle frequencies. The recording procedure was explained next and the student was given an opportunity to sing both songs on a trial

basis before the tape recording was made. Phonograph records were used as an accompaniment to the subject's singing both during the practice and recording sessions (17).

Two extreme groups were selected by subjectively listening to the tape recordings. The groups included the thirty-five best singers and the thirty poorest singers according to the subjective judgement of the investigator. By random sampling, fifteen students were selected from the group of extremely defective singers as the experimental group. The remaining fifteen subjects comprised the control group. The experimental group reported daily to the investigator for thirty minutes. During this time, students listened to and sang numerous recorded songs with the phonograph. The control group remained in their classrooms and received their regular music instruction. This instruction consisted of a twenty-minute period each day in which the classroom teacher taught the music (18).

The diplacusimeter was built to measure a deviation of an amount up to five percent, either plus or minus, between the two ears. It consisted of one fixed audio-oscillator and one variable audio-oscillator which was controlled by a tuning control knob. Measurements could be taken at the 250 cycles per second and 500 cycles per second frequencies (18).

The operator set the switches to the 500 cycles per second tone. The subject was seated in front of the diplacusimeter and a set of headphones on. The operator sat at the side of the instrument with another set of headphones over his ears. The subject was instructed to place his left hand on the tuning control to proceed. He was told to let the operator know when he thought the pitch in the right ear was the same as the one in the left ear. When this decision was reached, the operator recorded the dial reading, changed the

switches to the 250 cycles per second tone, set the dial to the five percent mark and told the subject to match those two tones. The instrument used for testing the hearing acuity of the student was an *A.D.C. Pure-Tone Audiometer, Model* (21).

Nine boys and six girls comprised the randomly selected group. Of the total number, five students were from one section of the fourth-grade, six were from another section, and four students were from the third section. The mean age of the subjects was nine years (23).

Students in the experimental group reported to the investigator in a sound-proof room. Students sang familiar songs and were taught a new song. Students sang along with a record (24).

This procedure was followed for three weeks before the first recording session was begun. Thereafter, the individual voices of all fifteen students were recorded weekly. Thirteen such recordings were made. The songs that were used for this purpose were selected from the group of listening songs that the students had been singing during the week. The tape recordings were subjectively transcribed into musical notation by the investigator and from the study of those transcriptions an attempt was made to establish patterns in the process through which a child learns to sing in tune (25).

Diplacusis scores were obtained for the students in the experimental group over a period of nine weeks. Nine diplacusis scores were obtained for each student in the experimental group over a two and one-half month period. Students were given four weekly practice trials prior to recording their diplacusis scores. Readings were taken from the diplacusimeter which was built specifically for this experiment. Information

gained from those readings was used in determining the nature and characteristics of diplacusis among that group of defective singers (25).

The following results were reported by the investigator:

1. Results of all comparisons indicated that at no time were there significant differences in the means of the various groups on hearing acuity. It was reported that one student with a seventy-two decibel loss of hearing in one ear was one of the thirty-five best singers. It was therefore concluded by the investigator that if the severity of a person's hearing impairment does not prevent his hearing music, it has little or no effect on his singing.
2. All comparisons indicated that there were significant differences in the means of the singer group and the defective groups on pitch discrimination, tonal memory, and diplacusis (117).
3. Since there were significant differences in the means of singers and defective singers on pitch discrimination, tonal memory, and diplacusis, product moment correlations were used to measure the relationships between each of the three measures and the number of notes that were sung correctly.
4. The results of the tests indicated that the correlation of each of the three measures with the number of notes sung correctly was significantly different. This indicated to the investigator that each of the three measures had some bearing on the performance of the students (118).

In order to discover if the pitch discrimination, tonal memory, and diplacusis tests were measuring the same capacities, correlations were made of the interrelationships among the three measures. Using the scores from all 113 students, it was found that the relationship between pitch discrimination and tonal memory was significantly different. As a result of those tests it was concluded that the three measures were different distinct characteristics, and were present in various amounts among students (118).

At the end of the experiment, gain scores were obtained on the four hearing measures for the singer, experimental, and control groups. Comparisons were made of the mean gains of the various groups on all measures. The gains made by the singer group on pitch discrimination was significantly different from that of the experimental and control groups. The experimental group made improvements on pitch discrimination whereas the

control group became worse on that measure. This inferred to the investigator that those students whose singing was extremely defective could, with additional help, make improvements on that factor while others in the same category without additional help were likely to become worse (119).

There were no significant differences in the amount of gain shown by the singers and both of the defective groups on tonal memory, diplacusis, and hearing acuity. It was noted that the improvements made by the experimental group were almost directly proportional to those of the singer group on both tonal memory and diplacusis. The amount of improvement shown by the control group on both measures was less than that of the singer group and the experimental group. Since there were no significant differences between the mean of any of the groups on hearing acuity at the beginning, nor in the mean gains on hearing acuity at the end of the study, this factor was disregarded (120).

A comparison of the gains of the experimental group and the control group revealed that although the experimental group showed more improvements on pitch discrimination, tonal memory and diplacusis, the amount of improvement was not statistically significant (120).

Diplacusis was found to be subject to minor changes from day to day both as to the amount and the direction of deviation above or below the fixed tone. Over a two and one-half month period of time, the fourteen students in the experimental group made nine judgments of both the 250 and 500 cycle tones. The amount of change in diplacusis among those students on the nine trials indicated to the investigator that the extent of the changes varied only slightly. Judgments of those students on the 250 and 500 cycle

frequencies indicated also that the pitches of a tone were heard consistently sharp or flat of the fixed tone, a majority of the time (120). Two of the fourteen subjects were 100 percent invariable as to the direction of deviation of their pitch judgment from the fixed tone (121).

A comparison of the subjects' first and ninth judgments on the diplacusis testing indicated that the students' scores were worse on their final trial than they were at the beginning. Since practice over a two and one-half month period did not result in any improvement on this measure, it was concluded that diplacusis was not a learned response. Instead, it appeared to be a relatively fixed and stable mental trait (121).

The investigator observed certain characteristics about defective singing from an analysis and comparison of the transcribed singing of all 113 students:

1. Defective singers usually employed a smaller singing range than that required by the ambitus of the song.
2. In some instances, students sang with as wide a range as that required by the ambitus of the song but the position of their range was lower than that of the ambitus.
3. The mean singing range used by all fourth-grade students on the recorded songs extended from C4 to B4 on one song and from Bb3 to Bb4 on the other song.
4. Students deviated from the true pitch of the notes in the musical score and, with few exceptions, the pitch error was below that of the note in the song.
5. In regard to melodic intervals appearing in the music, students sometimes missed intervals of a major third and smaller.
6. The percentage of students who missed notes became larger as the size of the intervals increased.
7. Students more often missed the second tone of a melodic interval of a given size when the second note moved upward from the first tone than when it moved downward.
8. In many cases, the highest note required in the song was out of the singing range of the students.
9. A relationship was noted in the amount of defectiveness in a student's singing and the interval between the highest note in the score and the student's response (122).

10. A relationship was observed between the percentage of notes a student missed and the percentage of tonics that he missed in singing the song.
11. Note values had little effect on whether or not the pitches were sung correctly (123).

Product moment correlations were employed to measure the relationships between the number of notes correctly sung by the students and the several characteristics found in their singing. From those correlations, it was concluded that there was a significant relationship between:

1. The number of notes sung correctly and the number of tonics sung correctly.
2. The number of notes sung correctly and the extent of range used by students in their singing.
3. The number of notes sung correctly and the position of range used by the students.
4. The number of notes sung correctly and the size of the interval between the highest note and the student's response.
5. The number of notes sung correctly and the size of the interval between the first tonic and the tone sung for it.
6. The number of notes sung correctly and the number of semitones used by students in singing the widest interval in the song (124).

During the experiment, students in the randomly selected group recorded one song each week, except during school holidays. Thirteen individual tape recordings were made of each subject's singing of nine different songs. Four of the songs were re-recorded on different dates during the experiment. From the transcriptions of the thirteen weekly recordings, the following additional information was gained about defective singing (125):

1. A majority of students made improvements on all six of the factors related to singing.
2. At no time did a student retrograde on all of those factors.
3. More improvements were evident in the repeated songs when the lapse of time between the two recordings was of longer duration.
4. The singing of extremely defective students can be improved with additional help (125).

Another factor that influenced the students' week-by-week improvements was the difficulty of the song that was being sung. Some of the elements that affected the difficulty of songs used in the experiment were:

1. The position and extent of the ambitus of the song.
2. The position of the tonic note.
3. The highest note required in the score.
4. The size of melodic intervals in the song.
5. The accompaniment used on the phonograph recording and especially the introduction played on the record (126).

The investigator reported that a study of the students' singing of the tonic notes revealed the following pattern of voice movement as the singing became less defective when singing *Tadpoles,* the easier of the two songs:

1. A majority of students whose singing was ninety-five to 100 percent defective sang, for the tonics, notes that were an interval of a sixth below the tonic.
2. Among students whose singing was eighty percent to ninety percent defective, a majority sang, for the tonic notes, tones that were an interval of a fourth below the tonic.
3. In the sixty percent to seventy percent defective category, the majority of students sang, for the tonics in the score, notes that were a second below the tonic.
4. Those students whose singing was less than fifty percent defective, sang a majority of the tonics correctly.
5. The intervals of a sixth, fourth, second, and prime represented the position of the students' voices on the tonic notes, for varying stages of defectiveness (127).

A comparison was made of the 113 students singing *My Partner*, the song of greater difficulty. The following results were noted:

1. Over fifty percent of the students who were in the six to sixteen percent category of defectiveness on the easier song dropped to the twenty to thirty percent category on the more difficult song.
2. Approximately the same percentage of students in each of the other categories dropped to the next level of defectiveness on their singing of the more difficult song.
3. Many students who sang the tonic notes correctly on the easier song, sang an interval of a second below the tonic on the more difficult song.

4. Many students who sang an interval of a second below the tonic on the easier song often dropped to an interval of a fourth below the tonic on the more difficult song.
5. Students who sang notes an interval of a fourth below the tonics in the score on the easier song, sang notes an interval of a sixth below the tonics in the more difficult song (127).

It was observed by the investigator that a majority of students in every category of defectiveness sang the intervals of an octave, sixth, fourth, and prime below the tonic. It was also found that students progressed upward through the intervals of a sixth, fourth, second and the prime as their singing became less defective (128).

The following recommendations were made as a result of the study:

1. Music specialists should determine the mean singing range of each of the various groups of singers they work with and transpose songs whose ambitus is beyond their ability to a key that is better suited to their singing range.
2. Music specialists should utilize part of their teaching time on practice sessions involving tone matching, singing of melodic intervals of various sizes, singing arpeggios, and helping the students sense the tonality of the songs.
3. It was recommended that classroom teachers be encouraged to utilize the phonograph in their teaching of songs. The phonograph was found to be an effective and valid device for teaching songs to children.
4. Defective singers should be given additional help since the majority of them can make improvements.
5. A classification of the difficulty levels of songs appearing in the music textbooks for each grade should be made by the music specialist.
6. The classroom teachers should be provided with an outline containing songs of graduating difficulty to be used throughout the school year.
7. Music textbook publishers should, in future revisions, attempt to classify songs according to their levels of difficulty for all grades.
8. A sequence should be followed in teaching where the easier songs are placed toward the front of the textbook and the more difficult songs toward the end.
9. Publishers of music textbooks should consider printing a non-graded remedial music text with worthwhile songs of limited range and interval requirements, similar to remedial textbooks used in the academic areas (121).

MUSICAL ACHIEVEMENT AND SELF CONCEPT

Greenberg (1970) investigated the relationship between musical achievement and self-concept in students grades four through six. It was the hypothesis of the investigator that continued inaccurate singing and underachievement in music were primarily results of a low self-concept of the student's ability to succeed in musical endeavors (57).

Greenberg defined self-concept as the perception the individual has of himself. The definition referred specifically to "the ways in which an individual characteristically sees himself and feels about himself" (57).

Greenberg stated that self-concept is learned and that people learn about who they are from the ways in which they have been treated by those who have contact with them. Greenberg elaborated that people develop their self-concepts in music from the kinds of experiences they have had in dealing with music. He said is it logical to infer that to produce a positive self-concept in music, it is necessary to provide experiences that show individuals they are doing well in music (58).

Greenberg explained that there is an ongoing impact between the self and the continuum of experiences involved in the process of learning music at school. Greenberg postulated that the learner perceives, interprets, receives, resists, or rejects what he is taught in music in light of his self-system. He said in the normal course of the development of the self-concept in music, the student is involved in a continuing process of assimilation and integration of new musical experiences and discoveries concerning his own resources and limitations (59).

Greenberg hypothesized that underachievement in music, including the lack of ability to sing in tune, is in part a function of the individual's self-concept. He hypothesized that

an individual who conceives himself as inadequate—who has negative self-concepts—also tends to achieve less than his potential capacity and aptitude would indicate (59).

The subjects in the study were ten inaccurate singers in grades four through six from the University of Hawai'i Elementary Laboratory School. The ten inaccurate singers were subjectively selected by the investigator. The ten subjects were placed in the school select chorus group. Greenberg hypothesized that changing self-perceptions of untuned singers would result in improved singing (59).

The school chorus consisted of eighty-four students from grades four, five, and six, including the ten boys classified as inaccurate singers. Admission to the chorus was primarily based upon the student's ability to sing accurately, with good tone quality, and an interest in being in chorus. All students in grades four through six were first asked to write down whether they wished to be in chorus. Of the 152 students in grades four through six, 112 answered "yes," including ten of the fifteen inaccurate singers—all boys—within the three grades (59).

The ten inaccurate singers who wished to join the chorus were allowed to participate, as part of the experiment. In order to determine how the inaccurate singers felt about their own abilities to sing, a questionnaire was devised that attempted to assess each student's self-concept of his various abilities, including music. Scores on self-assessment in music averaged lower than all academic areas except mathematics. As a result of the questionnaire, the investigator concluded that: (1) those boys had a lower self-concept of their own singing abilities, yet felt that they liked to sing; and (2) the main reason those

boys wanted to join chorus was not that they felt they could contribute to the success of the chorus, but the fact that it would enhance their self-esteem and prestige (60).

In rehearsals, the inaccurate singers were placed on the outer edge of the chorus seating. They were never placed next to each other. Although the investigator often heard those students sing inaccurately, nothing was said to them by the instructor (60).

Chorus rehearsals were held twice a week for one half-hour session over a period of eleven weeks, beginning in October and ending before the Christmas holidays. The ten inaccurate singers received no additional instruction during the experimental period except for their regular musical instruction in the classroom, which amounted to approximately 120 minutes a week. Approximately thirty minutes per week were devoted to singing activities. The investigator stated that the regular classroom instruction would not substantially affect the inaccurate singers' abilities to sing in tune, since regular musical instruction since the first grade did not improve their ability to sing. The investigator hypothesized that any singing improvement in the ten inaccurate singers would be attributed to chorus participation, additional experience, or changing self-perceptions (60).

At the conclusion of the experimental treatment sessions, the inaccurate singers were again tested by the investigator. Classroom teachers of the inaccurate singers were asked to write down comments about the students' aptitude, academic achievement, socio-emotional growth, and concepts of self (60).

The investigator summarized that the ten subjects exhibited the following:

1. Poor peer relationships
2. Academic underachievement
3. Lack of a strong feeling about self, or unrealistic self-concepts. Only two cases seemed to have well-integrated personalities (63).

The investigator concluded that the evidence pointed to emotional or psychological factors, and not musical factors, as the main cause of inaccurate singing (63).

Results of the study were reported by the investigator as follows:

1. Four inaccurate singers made little progress in singing as a result of being in chorus.
2. One boy made some progress and five boys made significant improvement (63).

The investigator noted that though the rigid controls needed in a research study were lacking, evidence of the study indicated that growth in achieving a positive self-concept in music led to improved singing accuracy (63).

The investigator asked the following research-related questions as a result of the experiment:

1. Do untuned singers know that they are singing inaccurately?
2. At what age or grade level should students be told that they are inaccurate?
3. What is the correlation between academic achievement, musical achievement, and personality factors?
4. What are some remedial techniques useful with inaccurate singers who are not suffering from inability to discriminate pitch?
5. What is the effect on students who are excluded from special groups? (64)

MUSICAL AND PERSONAL BACKGROUND

Eikum (1963) investigated factors affecting singing ability. It was the purpose of the study to:

1. Identify students who were not able to sing in tune.
2. Find out as much as possible about their musical and personal backgrounds.
3. Apply a terminology which indicated the extent of the inability.
4. Identify the causes of the inability.
5. Offer a means by which they could possibly overcome their singing deficiency (2).

A questionnaire was prepared and distributed (24). Following were the types of information about the child included on the questionnaire:

1. Musical heritage.
2. Home musical environment.
3. Attitude toward music and singing.
4. Interest in music and singing.
5. Intelligence as determined by the *Lorge-Thorndike Intelligence Tests*.
6. Occupational and family status of parent.
7. Health record.
8. Attendance record for current year.
9. General health rating by the classroom teacher.
10. Emotional stability rating by the classroom teacher.
11. Character traits rating by the classroom teacher.
12. Oral reading ability as rated by the classroom teacher.
13. Causes of oral reading weakness as determined by the classroom teacher.
14. Physical defects, handicaps, severe emotional disturbances, or severe nervous disorders as rated by the classroom teacher.
15. Singing performance score (TR score).
16. Prognostic music test score of selected sample students as determined by the *McCreery Elementary Rhythm and Pitch Test* (25).

All students were instructed to complete the preliminary information about age, school, grade, gender, etc., with the help of the classroom teacher or parent. The classroom teacher and music teacher completed the remainder of the questionnaire. Classroom teachers were instructed to observe the descriptions for the ratings of "four," "three," "two," "one," and "zero" as "excellent," "good," "average," "fair," and "poor," respectively, in all cases where such a number rating was employed (26).

All students had from six to eight weeks familiarity with the songs or tone patterns to be used in the recording procedure (27). Familiar and unfamiliar songs and tone patterns were used. Students were recorded individually. They responded after the first tone of the song or pattern to be sung was sounded. If the student experienced difficulty, the music teacher helped by singing the pattern in the same voice range as the student. Students were allowed two attempts to complete a song or pattern. In some cases, the

relative range of the song or tone pattern was lowered to fit the range of the student being tested (28).

The tape recordings were evaluated by two public school music teachers and one elementary classroom teacher. The judges were instructed to rate the students on three levels of ability with scores of "two," "one," and "zero"--meaning "excellent," "average," and "poor," respectively. They were instructed to consider standards of performance on the basis of accuracy of note and pattern reproduction alone. A rating system was devised for the TR, or tape-recorded scores. The scores "four," "three," "two," "one," and "zero" conformed, respectively, to the terms "excellent," "good," "average," "fair," and "poor." Tests of significance were applied to the differences in TR mean scores for the various subgroups of the test sample. The significance of a difference was determined by calculating a *t* ratio score and then determining the probability, by reference to a *Fisher's Table* of *t* values, that such a value of *t* could be obtained through chance (29).

The using of a music prognosis test as part of the research data was included. Norm or average scores were established for each existing grade level for grades two through nine by administering the *McCreery Test* to whole classrooms at each grade level included, and then averaging the scores for each grade level. Random selection of five or six students at each of three TR performance score levels ("zero," "two," and "four" from grades four, six, and eight) provided a control group including the same students who had three years previously been in grades one, three, and five. *McCreery Tests* were given to the control group and results compared with the average scores established for the whole classroom test situations for the corresponding grade levels. Standard deviations of the

McCreery Test scores were calculated for grades two, four, six, and eight in establishing norms for the control group. Significance was noted by comparing the *McCreery Test* scores, for the TR score categories listed, "zero," "two," and "four" with the averages for the grade levels included in the control group (31).

A total of 1,196 students were enrolled in three different elementary schools (grades one through six) for the 1957-58 school term. Six-hundred-thirty-one students were enrolled in grades one through three; 565 were enrolled in grades four through six. Of the 1040 students included in the study, 595 were in grades one through three and 445 were in grades four through six. Nearly eighty-seven percent of the total school enrollment was included in the research study (33).

The researcher reported that there were only four fewer boys (N = 518) than girls (N = 522) in the entire test sample. Grades one through three included 304 boys and 291 girls; grades four through six included 214 boys and 231 girls (33).

Students who made TR scores of "three" or "four," were classified as singers. Students not classified as singers conformed with the following list (39):

1. Sings in the speaking range only and in a very limited range.
2. Sings consistently higher than the correct melody but seldom, or not at all, lower than the melody.
3. Follows the general direction of the melodic line but does not sing the correct note at the proper time.
4. Has trouble matching even one tone correctly.
5. Imitates only two or three, or a very small number of tones or part of a phrase.
6. Sings with the group but not alone.
7. Sings alone but gets confused when singing with the group.
8. Does not differentiate between the speaking and singing voice but says the words in rhythm with no tune or noticeable variation in pitch.
9. Possesses similar characteristics to number eight but pronounces words infrequently and in a sporadic nature with relation to the rhythm of the song (18).

Data on responses concerning musical heritage as reported by the investigator follows:

1. Singing ability of off-spring is affected by the singing ability of parents and grandparents (hereditary factors).
2. For all "yes" respondents, the average TR score was above the mean TR (an individual tape recording performance score based on the average score evaluation of the tape recorded performance of the student by three qualified judges) score for the sample to a highly significant or very highly significant degree.
3. For all "no" respondents, the average TR scores were below the mean TR score for the sample to a highly significant or very highly significant degree.
4. The greatest significance occurred for average TR scores above the mean TR score for students giving *yes* responses to questions about whether or not parents or grandparents sing (40).

Data on responses concerning the home musical environment was reported as follows:

1. Singing ability is affected by the home musical environment.
2. For nearly all "yes" respondents, the average TR scores were above the mean TR score for the sample to a highly significant or very highly significant degree.
3. The average TR score for all students giving "no" responses were significant, highly significant, or very highly significant (40).

The data for attitude toward music items on the questionnaire were reported as follows:

1. In six of nine questionnaire items for students giving "no" responses, average TR scores were below the whole sample average TR score to a significant, highly significant, or very highly significant degree.
2. Questionnaire items nine, thirteen, and seventeen showed only slight or no significant difference in average TR scores for "no" or "yes" respondents.
3. The number of students giving "no" responses for any one item was relatively small, thus, effecting a highly significant average TR score difference without affecting the TR average of "yes" respondents to a significant degree.
4. The interest inventory data showed that a preference for musical activities through the fifth rank affected the average TR scores to a highly significant or very highly significant degree.
5. The average TR score for those selecting three or more music activities in the first five ranks of the interest inventory, was above the whole sample average TR score to a very highly significant degree.
6. For those with one or no music activities in the first five ranks, the average TR score was below to a highly significant degree (41).

Data reported in regards to attitude toward singing follows:

1. A relationship did exist between attitude toward singing and singing ability.
2. Whether or not a student would like to be a musician had no particular significance for the ability to sing.
3. A relationship between attitude toward singing and singing ability was reported. Many more girls than boys placed the singing activity in higher order ranks of the interest inventory (42).

Data on interest in music was reported as the following:

1. There was a relationship between interest in music and the ability to sing.
2. Students who played an instrument had average TR scores above the whole sample average TR score to a highly significant or very highly significant degree.
3. Students with two or more years of instrumental training showed the highest TR scores.
4. Students who did not play an instrument showed average TR scores below the whole sample average to a highly significant degree.
5. Interest in music, as expressed through attitudes in questionnaire items eight, ten, twelve, thirteen, fourteen, fifteen and seventeen showed significant or very highly significant differences in average TR scores for "no" respondents in five of the seven items.
6. Average TR scores for boys were above the whole sample TR average score through seven ranks for the singing activity on the interest inventory.
7. The same was true for the girls through the first six ranks of the interest inventory.
8. Both boys and girls who showed better than average ability in singing also had a fairly wide range of interests (43).

Data in regards to interest in singing was reported by the investigator as follows:

1. A relationship did exist between interest in singing and the ability to sing.
2. Erratic fluctuations in the average TR scores of the singing activity for progressively lower ranks on the interest inventory were reported. Those fluctuations were smoothed out by combining rank groups for average TR scores (43).
3. Statistical differences were apparent between the boys and girls, both in respect to the relative numbers making high order rank choices for singing, and the average TR scores obtained.
4. Boys showed higher average TR scores than girls in both the high order ranks and the low order ranks.
5. Both boys and girls in middle order ranks showed the same average TR scores (50).

Intelligence scores ranged from a low of fifty to a high of 144, with an average of M = 107.58 for all students. It was the conclusion of the investigator that intelligence did affect the ability to sing (51).

Some general characteristics of above-average singers as determined the survey included students:

1. Whose father, mother, grandparents, brother, and sisters played an instrument or sang.
2. Who had one or more years of instrumental training (may have included one year of pre-instrument program training).
3. Who had a piano at home.
4. Who expressed a desire to perform in a choir or chorus.
5. Who ranked singing no lower than sixth, if a girl, or seventh, if a boy, on the interest inventory.
6. Who had two or more musical activities listed among the top five choices on the interest inventory.
7. Who had average or above average intelligence.
8. Whose family-occupational status was not in the laborer, skilled worker, or broken-home classification.
9. Who had not had a severe bronchial infection or asthma.
10. Who had fewer than thirty days absence from school for the current year.
11. Who had average, or better, ratings in general health.
12. Who had average, or better, ratings in emotional stability.
13. Who had average, or better, character traits ratings.
14. Who had average, or better, ratings in oral reading ability.
15. Who had listed no physical defects, handicaps, severe emotional disturbances, or severe nervous disorders.
16. Who had average, or better, scores on the *McCreery Music Prognostic Test* (60).

Specific causes for an inability to sing, as determined by the survey study results included:

1. Poor musical heritage.
2. Poor home musical environment.
3. Poor attitude toward music and singing.
4. Little interest in music and singing.
5. Low intelligence as determined by the *Lorge-Thorndike Intelligence Tests*.
6. Inadequate ability due to occupational and family status.
7. Poor health record—in a few instances.
8. Chronic absence from school.

9. Low emotional stability.
10. Poor character traits.
11. Some physical defects, handicaps, severe emotional disturbances, or severe nervous disorders (61).

The investigator concluded that it appeared that singing ability improved with an increase in grade level or age. The findings of the study also indicated that intelligence, heredity, and home environment affect singing ability (61).

Recommendations were made as follows by the investigator:

1. Consideration must be given to the relative abilities or stages of vocal development of the students in the classroom.
2. Methods and materials should be in keeping with the students' abilities and interest according to the developmental learning processes involved.
3. Materials with high student appeal should constantly be used or devised in order to hold student interest.
4. A simplified and more abbreviated form of the questionnaire might well be devised and employed periodically by the music teacher to check the progress of the singing program.
5. Coordination between classroom teacher and music teacher should be maintained in order to note student progress in all areas of behavior and experience at school.
6. Similar or identical testing should be done in other elementary school systems throughout a state or in many states to establish more objective methods in evaluating the singing ability of elementary students.
7. More systematic research should be done concerning the causes for the inability to sing.
8. Sociological and psychological aspects of the causes for a student's inability to sing, as related to the occupational and family status, may prove to be an interesting area for investigation as well as beneficial to the field of music education (62).

NEUTRAL SYLLABLE VERSUS TEXT

Levinowitz (1987) conducted an experimental study of the comparative effects of singing songs with words and without words on children in kindergarten and first grade.

The specific problems of the study were the following:

1. To determine the comparative effects of song instruction with and without words on levels of developmental music aptitudes of children in kindergarten and first grade.
2. To determine the comparative effects of song instruction with and without words on the singing achievement of children in kindergarten and first grade (5).

Three classes of children in kindergarten and three classes of children in first grade from a parochial school in Trenton, New Jersey, participated in the study. Prior to the experimental music instruction, the students received no music instruction. The students were said to be representative of an urban population of diverse ethnic and socio-economic backgrounds (20).

Before instruction began, the Tonal subtest of the *Primary Measures of Music Audiation* was administered to each of the three classes of children in kindergarten and to each of the three classes of children in first grade. The rhythm subtest of the *Primary Measures of Music Audiation* was administered the following week (20).

The *Primary Measures of Music Audiation* was designed to measure the developmental music aptitudes of children in kindergarten, first, second, and third grades. The tonal subtest was a recorded series of forty paired tonal patterns and the rhythm subtest was a recorded series of forty paired rhythm patterns. For each subtest, the children were asked to listen to each of the paired patterns and to determine whether the paired patterns sounded the same or different (21).

The three classes of children in kindergarten were assigned randomly to Experimental Group T1, T2, or T3; the three classes of children in first grade were assigned randomly to Experimental Group T1, T2, or T3. For one academic year, all children in each of the three experimental groups met with the investigator for one thirty-minute music class

each week. During each of those music classes, the children participated in rhythm activities and sang at least six rote songs (21).

The children in kindergarten and first grade who were assigned to Experimental Group T1 sang songs primarily with words; no more than two songs without words were sung during any one thirty-minute music class. The children who were assigned to Experimental Group T2 sang songs primarily without words; no more than two songs with words were sung during any one thirty-minute music class. The children who were assigned to Experimental Group T3 sang only songs with words; all six rote songs sung during the thirty-minute music class were performed with words. The repertoire of songs, with and without words, for all experimental groups comprised major, harmonic minor, dorian, phrygian, lydian mixolydian, and aeolian tonalities. The songs sung by the children in all experimental groups were in duple, triple, and unusual meters (22).

The technique used to teach the rote songs to all children in all experimental groups was as follows:

1. The teacher established the tonality of the song by playing the appropriate harmonic progression or by singing the appropriate melodic sequence.
2. The teacher sang the song several times.
3. The children were encouraged to sing along whenever they wished, but they sang only if they so desired. Songs without words were sung on a neutral syllable such as "bum" or an onomatopoetic sound such as "moo," "swish," or "drip" (22).

During the last month of instruction, two criterion songs with words were taught to all children in all of the experimental groups. One of the criterion songs was in major tonality and triple meter, the other criterion song was in minor tonality and duple meter. Both songs were sung approximately the same number of times by all of the children in each of the three experimental groups (23).

At the end of instruction, both subtests of the *Primary Measures of Music Audiation* were readministered to all children. Each child was also tape recorded singing the two criterion songs (23).

Two, five-point rating scales were constructed by the investigator to assess the children's tonal achievement and rhythm achievement. The tonal rating scale and the rhythm rating scale were adapted versions of the tonal and rhythm rating scales used in a previous study by the investigator to assess children's tonal and rhythm achievement (23).

Two judges independently evaluated the children's recorded tonal performances and rhythm performances of each criterion song. The combined tonal ratings of the two judges for both criterion songs combined served as the children's tonal achievement scores. The combined rhythm ratings of the two judges for both criterion songs combined served as the children's rhythm achievement scores (24).

Results of the study follow:

1. For the tonal analysis for children in kindergarten, and the tonal and rhythm analyses for children in first grade, the two-way interaction effects between level of aptitude and time of PMMA administration were found to be statistically significant.
2. For the rhythm analysis for children in kindergarten, the three-way interaction effect was statistically significant.
3. No significant interaction or treatment effects were found for any of the four singing achievement analyses.
4. Main effects for level were found for the rhythm achievement analysis for children in kindergarten and for both the tonal and rhythm achievement analyses for children in first grade.
5. No main effect for level was found, however, for the tonal achievement analysis for children in kindergarten (50).

The investigator made the following conclusions based on the evidence in the study:

1. Tonal and rhythm audiation of young children with low tonal and rhythm developmental music aptitudes were enhanced by song instruction that comprised songs both with words and without words.
2. There was no evidence to suggest that one type of song instruction was more beneficial than the other for developing tonal and rhythm audiation of children with high tonal and rhythm developmental music aptitudes.
3. In particular, audiation of first grade children with high rhythm developmental music aptitude improved with song instruction that comprised only songs with words.
4. There was no evidence to suggest that any one of the three types of song instruction was superior to the other for developing young children's singing achievement (51).

PIANO INSTRUCTION

Whitman (1970) conducted a study to determine if there was a significant difference in the accuracy of pitch reproduction in the solo voice performance between performers who had piano instruction and those who did not have piano instruction. Whitman (1970) also sought to determine if a consistent pattern of intonational differences existed, depending on whether the performer had no piano instruction, one year of piano instruction, two years of piano instruction, or three years of piano instruction (3).

Seventy-two students were selected from the fourth-grade, fifth-grade, and sixth-grade of the public schools of Levy and Marion counties in Florida to serve as subjects. Only those students who fell into one of the following groups were used for the study:

1. Group 4-0: Fourth grade performers without previous musical instruction.
2. Group 4-GM: Fourth grade performers who had had general music class instruction.
3. Group 4-GMI: Fourth grade performers who had had general music class and one year of instrumental instruction.
4. Group-4GMP: Fourth grade performers who had had general music class and one year of piano instruction.
5. Group 5-0: Fifth grade performers who had had no previous musical instruction.

6. Group 5-GM: Fifth grade performers who had had general music class instruction.
7. Group 5-GMI: Fifth grade performers who had had general music class instruction and two years of instrumental instruction.
8. Group 5-GMP: Fifth grade performers who had had general music class instruction and two years of piano instruction.
9. Group 6-0: Sixth grade performers who had had no previous musical instruction.
10. Group 6-GMI: Sixth grade performers who had had general music class instruction and three years of instrumental instruction.
11. Group 6-GMP: Sixth grade performers who had had general music class instruction and three years of piano instruction (10).

Each student in grades four, five, and six in the public schools of Levy and Marion counties in Florida completed a questionnaire under the supervision of a classroom teacher. The questionnaire provided the researcher with biographical and musical experience information about the student. Approximately seven-thousand students completed the questionnaire (11).

The subjects were then categorized into thirteen populations according to the information provided. Twelve of the populations corresponded to those with which the research was concerned and the thirteenth contained all other subjects. To avoid sampling bias, six subjects were randomly selected for the study from each of the twelve populations corresponding to the twelve groups. Three tape recorders were used for the testing procedure--one to produce the instructions, one the stimulus tones, and one to record the response. Students (N = 60) were tested individually (10).

Subjects were randomly assigned to performance conditions to eliminate the order of the presentation affecting the pattern of the outcome. The subjects within each group

performed the three pitches in a different sequence. The sequence was randomly assigned each performer. The six tonal sequences were as follows:

1. G#-C-E
2. G#-E-C
3. C-E-G#
4. C-G#-E
5. E-G#-C
6. E-C-G# (12)

Instructions were recorded identically for each order of presentation. Each subject was instructed to use the syllable "mah" in matching the pitch. Thirty seconds of music were presented on the tape between the presentation of each stimulus tone to prevent the response of one tone affecting the response of the other tones. Music was selected for the interlude between performances in the keys of Gb and D major in order to minimize the occurrence of the stimulus tones in those interludes (12).

Each stimulus pitch and the corresponding vocal response to that pitch were recorded on a single magnetic tape in order to prevent variances which might be caused by variations in the recording equipment. Upon completion of all recordings, tape loops were made for each stimulus pitch and the subject's response to that pitch. Those loops were analyzed through the use of a *Conn Stroboscope* to determine the accuracy of the performance, compared to the stimulus pitch. The sharpness or flatness was recorded in cent deviation (13).

The data was subjected to an *F* test to determine the significance of any difference between groups, and a profile analysis was performed to determine if parallelism was present in the variances (13).

The *F* test was used to determine the statistical significance in analyzing the difference between the means of the groups in the experiment. This test consisted of dividing the

square of the larger standard deviation by the square of the smaller standard deviation and then referring to the *F* table for an interpretation of the significance of the difference between the variances. A profile analysis was computed to determine if parallelism existed between the variances of the groups under consideration. The statistics were subjected to the *F* test at the $p = .05$ level of significance (14).

Fifteen separate comparisons of variance were tested for equality and none was found to be significant at the $p = .05$ level. The same comparisons were subjected to the *F* test at the $p = .05$ level to determine if parallelism that existed between the sample size may have been insufficient to allow the computations to reject parallelism in the experiment (15).

The hypothesis that subjects who had had piano instruction perform with greater pitch accuracy than performers who had not had piano instruction was rejected according to the statistics from the experiment. The investigator noted that since there was no significant difference between any of the groups, the second hypothesis that subjects who had had piano instruction perform with greater pitch accuracy than subjects without piano instruction at each grade level was also rejected (15).

The investigator mentioned that even though the difference between the groups was not statistically significant, there did appear to be several patterns of apparent differences in the performance of the test subjects. The observations were as follows:

1. Those groups without previous music instruction were consistently more inaccurate when performing each of the three stimulus pitches.
2. The G# was performed most inaccurately of the three pitches for eight of the twelve groups involved in the experiment.
3. The largest individual deviations and the largest group standard deviation occurred when responding to the G# stimulus pitch.
4. The stimulus pitch C was the most accurately performed pitch for six of the twelve groups.

5. The E was not the most inaccurately performed pitch for any of the twelve groups.
6. Sixty-four percent of the pitches were performed flat, ranging from five hundred and fifty-five cents to one cent below the stimulus pitch.
7. The range of the thirty-six percent of the pitches which were performed sharp was from one cent to three hundred and sixty-two cents above the stimulus pitch.
8. Twenty-six percent of the subjects responded sharp to the stimulus pitch G#, thirty-two percent were sharp when performing the E and fifty-one percent performed sharp when responding to the C.
9. The order in which the pitches were performed was not a factor in the vocal accuracy of the subjects.
10. Those groups who had had general music and instrumental instruction performed the C most accurately, the E next most accurately, and the G# least accurately.
11. Those groups who had had instrumental instruction and those who had had piano instruction exhibited stronger tendencies toward performing flat than did the other groups at each grade level (16).

The investigator made several conclusions in regards to the study:

1. The skill required to match vocally an isolated pitch in the experiment was not affected to a statistically significant degree by the musical experience of those subjects within the test population.
2. Those test subjects who had not had any prior musical training performed consistently more inaccurately than did those with prior musical training.
3. Musical training did affect the particular skill of pitch matching as done in the experiment to some degree, but the specific type of musical training was of little significance (17).

The investigator summarized that the data from the experiment indicated that the G# stimulus pitch may have introduced a variable regarding the vocal range of the performers. It was reported that seventy-three percent of the subjects performed flat when attempting to perform the pitch G#, compared to only forty-eight percent performing flat on the lowest stimulus pitch C. This considered with the fact that the G# was the most inaccurately performed pitch for eight of the twelve groups caused the researcher to conclude that the placement of the pitch was a factor in the performer's accuracy (17).

The investigator reported that there was evidence in the research that tonal memory was a factor in the subject's performances. It was reported that some of the subjects performed within twenty-five cents of the stimulus pitch on two of their three performances while the third vocal response was more than two hundred cents variation from the stimulus pitch. The researcher concluded that the possibility of the inaccuracy resulted from a temporary lapse of concentration on the part of the subject which caused the loss of the stimulus pitch concept (18).

The investigator made several recommendations for future research in the area of testing the vocal accuracy in matching stimulus pitches:

1. Consider stimulus pitches which will eliminate the variable induced by limitations of the vocal range of the performer.
2. Use a recording of the voice for the stimulus pitch rather than an instrument with unfamiliar timbre.
3. Isolate the vocal music experience of the subjects as one of the variables to be considered in addition to those variables used in the experiment.
4. Eliminate the dependency upon memory in the subject's attempt to duplicate the stimulus pitch.
5. Record the stimulus pitch as the subject hears it and record the subject's response to that pitch on the same equipment. This should eliminate any deviations in the stimulus pitch caused by electrical or climatic conditions on the equipment in use.
6. Consider an experimental design which would provide more accurate control of the experimental treatment received by the subjects.
7. Eliminate the apprehension felt by the subjects or measure its effect on their performance.
8. Duplicate the experiment, utilizing as many of the above recommendations as possible.
9. Examine the correlation between the subject's ability to match an isolated stimulus pitch and his ability to match the same pitch in the context of a simple melody.
10. Examine the effect of instruction on other types of instruments on vocal accuracy.
11. Examine the effect of specific elementary music curricula on vocal accuracy. This should include such areas as specific learning activities, in addition to the effect of the music teacher's versus the classroom teacher's responsibility for those activities.

12. Compare the effects on vocal accuracy of a newly initiated (less than three years) elementary music curriculum with an elementary music curriculum which has been well established in the school over a long (more than three years) period of time.
13. Examine the incidence of vocal inaccuracy in the elementary school grades.
14. Eliminate known nonsingers from the test population.
15. Examine the incidence of nonsingers in the elementary school grades (19).

PITCH DISCRIMINATION

Pitch discrimination has been investigated in relation to vocal pitch accuracy to determine if it is a possible cause of inaccurate singing—that singers are unable to hear differences between pitches and thus sing inaccurately. If poor pitch discrimination is determined to be a cause of singing inaccuracy, then a remedial strategy may be developed to aid inaccurate singers. Following are some investigations related to pitch discrimination in relation to the inaccurate singer.

Bentley (1968) conducted two studies to investigate: (1) the incidence of monotonism; (2) the memory and pitch discrimination abilities of normal singers compared with monotones. Subsidiary aims were to discover if there was any evidence of a gender difference in memory and pitch discrimination; and the degree of relationship, if any, between memory and IQ and pitch discrimination and IQ (27).

The first investigation (1954) was undertaken in schools in and near Reading, England. This was followed up two and a half years later by a further investigation, again in the Reading schools using a different population in each age range, but also bringing in schools from a wider area. The results of the second investigation confirmed those of the first: that there was a general pattern of a high proportion of monotones at seven

years of age, this proportion gradually decreasing as the children became older (Bentley, 1957; 28).

The first survey was sent to several schools in the Reading area requesting a return of the numbers of children who "persistently sing out of tune," the figures were to be given separately for boys and girls, and in year age ranges from seven to thirteen, with "Nil" returns where appropriate. Pupils in the upper age ranges who showed signs of changing voice were not to be included. In addition to the monotones the numbers of normal singers were to be given in each group (28).

The investigator gave the following characteristics of the monotone singer:

1. Children who persistently sing out of tune.
2. Are easily identified in the classroom because they are frequently enthusiastic vocalists, at least in their early years.
3. They become, to some teachers, such a source of irritation that they cannot fail to be noticed (28).

The investigator defined the normal singer as "merely the rest of that particular class or age range" from which the monotone singers were derived (28).

The investigator addressed the point that all the teachers involved might not have the same concept of monotonism by stating that "most of the individual returns showed the same general tendency as was shown by the total number, thus suggesting that there were few, if any, misunderstandings." He said that this was confirmed by personal conversation with several of the teachers involved (28).

Twenty-one primary and six secondary public schools completed the return, along with three private preparatory schools, covering a total of 7,361 children from seven to thirteen years of age. The total numbers of boys and girls sampled were roughly equal, according to the investigator who stated that the raw figures gave an approximate idea of

the comparative numbers of monotones between the genders. Percentage figures were calculated in order to present a more reliable picture for purposes of comparison (29).

The following results were reported by the investigator based on the survey:

1. Both boys and girls showed a rapid improvement between the first and second age ranges.
2. As the proportions became smaller, the curves tended to level out, with four to five times as many boys as girls continuing as monotones.
3. More boys than girls were monotones.
4. The monotonism tended to work itself out with increase in age.
5. The proportion of boys still showing inaccurate singing at the top of the primary school was surprisingly high (fifteen percent).
6. When the next age range was reached (eleven to twelve) the proportion was reduced to almost one-third of the previous (five and nine-tenths percent), with an average of seven percent for the first two years of the secondary school stage (29).

The investigator attributed the latter figures to less careful ascertainment, or to the sampling; only a small number of secondary schools made the return, and the total number of boys reviewed was only 571, as compared with 3,147 boys in primary schools (29).

The first investigation was followed up two-and-a-half years later, and the range of schools was extended beyond the Reading area. Teachers overseas, as well as those in the United Kingdom, were invited to send information (29).

In addition to the local schools, fourteen others sent a return of their comparative figures: five primary schools, four secondary, two preparatory schools for boys, and two independent schools for girls, covering the entire age range of seven to thirteen years; in addition, one school overseas made the return, a preparatory schools for boys in Tasmania. Those fourteen schools covered a total sample of 3,279 pupils, and it was reported by the investigator that their trends and proportions confirmed those of the first Reading and district survey (29).

In the second Reading survey an entirely new set of assessments was given on different children, and in several cases by different teachers, from those involved in the former survey. It was reported by the investigator that the results confirmed those of the first survey. The investigator noted that it was therefore possible to combine all the returns to give the overall picture for a total of 16,699 children, a sample large enough to permit fairly valid generalizations about the proportions of normal singers and monotones (30).

A comparison between the first and second surveys revealed the following:

1. Similar proportions and tendencies.
2. The decrease in the proportions of monotones was steady from year to year until the secondary stage was reached, when there appeared to be left what the investigator considered to be the "hard core" of those who were still monotones (between seven percent and eight percent of boys and about two percent of girls).
3. At the age of seven to eight years there were about twice as many boys as girls reported who were monotones.
4. From the age of eight years upwards there were between three and four times as many boys monotones as girls.
5. There was again, a reportedly a return of a high proportion of boys (fourteen percent) showing inaptitude at the top of the primary school, with a big drop at the beginning of the secondary school stage (30).

The figures concerning the children of eleven to thirteen years of age were compared by the investigator and the following results were obtained:

1. In the first survey, out of 1331 boys and girls of eleven to thirteen years, fifty were assessed as monotones, or three and seventy-five-hundredths percent.
2. In the second survey, out of 4,005 boys and girls in the same age range, 176 were assessed as monotones (4.38%) (30).

Pilot tests of memory and pitch discrimination were conducted by Bentley (1968). Two standardized tests which were available at the time were: (1) *Measures of Musical Talent* (Seashore; 1939); and (2) *Standardized Tests of Musical Intelligence* (Wing; 1947). Each

contained a test of memory. Seashore's contained a test of pitch discrimination; Wing's *A Test of Pitch Change* (31).

It was noted by the investigator that the tests were not entirely suitable for use with children of the age range under review. The investigator stated, therefore, that only certain aspects of them were used as a model for the tests of the investigation developed by the investigator. These are described below (31).

It was decided by the investigator to model the new memory test on Wing's, since this approached the more nearly to the actual musical situation: the wing memory test consisted of recognizably musical material, i.e. actual melodies or successions of pitch sounds within a rhythmic framework, involving both the tonal and rhythmic elements simultaneously, whereas the content of the Seashore tonal memory and rhythm tests were described as the material of music. However, the element of counting was considered to be a distraction for young children, so this was eliminated and they were instead, asked to state merely whether the second tune of each pair was the same as, or different from, the first. It was apparent, then, that some of the items had to be the "same," whereas all the Wing items were different (33).

Thirty pairs of tunes were played on the piano, ranging in length from three notes to ten. The subjects had to write "S" for "same" if the second tune of a pair was the same as the first (eight items), and "D" for "different" if the second contained an alteration in pitch (twenty-two items). In the latter case not more than one note was altered (33).

For the devised *Pitch Discrimination Test* both the smaller-than-semitone differences of the Seashore pitch discrimination test and the partial masking of the change of the Wing pitch change test were thought to be too difficult or too confusing for young children by

the investigator. The investigator explained that at that stage the aim was neither to discover a pitch discrimination threshold nor merely to solicit judgments of direction of pitch movement, but to find out how the size of the interval affected pitch discrimination judgments (34).

The investigator said that therefore, in the new pitch discrimination test, both the unison and intervals from the semitone to the octave were included, employing all the semitonal steps within the range: A6 to E6 (34).

The memory and pitch discrimination tests were applied to eighty-eight monotones within the same age range as the normal singers, nine to eleven years. Those children took the tests at the same time as, and along with, the accurate singers, and they were not aware that they were in separate groups (40). The investigator reported that as a group, the monotones performed the tests less ably than the accurate singers, and that there was significant difference between the mean scores in both memory and pitch discrimination (42).

In regards to the *Memory Test*, it was reported the lowest score in both groups was seven, but where no monotone scored higher than twenty-three, nearly ten percent of the normal singers scored over twenty-four (43). The distribution of the scores of the accurate singers was reported to approximate the normal curve; that of the scores of the monotones while less regular (due in part, according to the investigator, to the smaller size of the sample) still showed a tendency to normal distribution. The difference between the means of the two groups was reported to be significant by the investigator ($M = 2.70$, $SE = 0.44$) (43).

In regards to the devised *Pitch Discrimination Test*, it was reported by the investigator that both groups showed scores up to the maximum of thirty, but the nature of the distribution differed considerably; whereas the scores of the accurate singers were entirely skewed, those of the monotones were more evenly spread. The difference between the means was four and six-tenths, which was reported as being highly significant; but, the investigator noted that caution should be taken in interpreting the results since: (1) the test proved too easy for accurate singers, and (2) their score distributions were so far from normal. The investigator concluded that the evidence appeared to indicate that a group characteristic of monotones was relative deficiency in both memory and pitch discrimination (44).

At a later date a further comparison was made between accurate singers and monotones, and between boys and girls. In the meantime a new test battery had been developed. Three of the tests from that battery were used in the comparison: pitch discrimination--slightly different from the published test in that the pitch discrimination test used contained fifteen items gradually decreasing in frequency differences from twenty-six cps, or the semitone, through differences of eighteen, twelve, six down to three cps. In the published test there were twenty items, still from twenty-six cps to three cps, but from the twelve cps differences the steps were ten, eight, six, five, four, down to three cps (46).

Sample A consisted of seventy-nine accurate singers and seventeen monotones. The age range was eleven and eight-tenths to twelve and five-tenths with a mean of twelve years. The IQ range for was seventy-one to 119 (M = 97.90) for the accurate singers. The IQ range for monotones was seventy-one to 115 (M = 91.10) (47).

Sample B consisted of 270 accurate singers and thirty-five monotones. The ages of the subjects in this sample were approximately eleven years. The IQ range was 140 – 100 (M = 111.80) for the normal singers. The IQ range was 139 – 100 (M = 113.90) for the monotones (48).

Examination of the scores of numerous subjects who had previously done the tests indicated that the scores of children of lower general ability, as measured in terms of IQ tended to be less reliable, and to show more influence of guessing, than those of children of higher general ability. This further comparison was therefore based on the most reliable evidence that could be obtained. Only the scores of subjects of IQ 100 or more were now used; i.e. children of average and above average general ability. The investigator justified that those children would be less prone to loss of concentration, and less likely to guess an answer as the line of least resistance (49).

The last year of the primary school, mean age eleven years was chosen by the investigator on the basis that those children who had reached this age and still remained monotones might be regarded as chronic. It was noted that most of those who had started school as monotones were now singing in tune with their fellow students (49).

Eleven different schools were utilized in the study. The investigator reported that there were considerable differences in both IQs and test scores, between the different schools. In view of this, and of the non-normal distributions of individual IQs and test scores, school mean scores were used as the basis for the calculation of mean differences between normal singers and monotones and in testing the significance of such mean differences (50).

It was furthermore reported that since there were far more boy monotones than girl monotones, the means for boys and girls were treated separately. It was justified by the investigator that if both genders had been combined the comparison of monotones and normal singers would also have been a comparison of a sample consisting largely of boys with a sample having a higher proportion of girls than boys (50).

Differences between the means for both boys and girls, in each part of the test and IQ were calculated for each school, added together, and the total divided by the number of schools involved (N = 11) to give an average difference (50). The difference between the IQs of the boys and the girls was not significant and was ignored (M = 0.35, SE = 1.79, DF = 2.23) (52).

The difference was then calculated for boys and girls together of IQs of accurate singers and monotone groups. This was also reported as being not significant. The investigator thus assumed that in the total sample of 305 boys and girls there was not much evidence that the difference between normal singers and monotones could be explained by differences in IQ (53).

The results of the *Pitch Discrimination* test follows. The mean score of the boys was higher than that of the girls (M = 0.66). This difference was reported by the investigator as not being significant. In view of this, and of the fact that there were only two girl monotones, the calculation of the difference between the accurate and monotone groups was made on boys only. At each of the eleven schools the accurate means were higher than the monotone means (M = 2.37). This difference between normal singers and monotones was significant (p =0.1, t = 6.07) (53).

The *Tonal Memory Test* results were reported as follows. The difference between accurate boys and girls was not significant; they were therefore taken together. The difference between accurate singers and monotones was significant ($p = 0.1$) (53).

In regards to the *Rhythmic Memory Test* the results indicated that there was no significant difference between normal boys and girls, so both genders were combined. The difference between normal and monotone groups was not significant ($p = .05$) (53).

In regards to the *Chord Analysis Test*, there was no significant difference between normal boys and girls. In regards to the Full Battery there was no significant difference between accurate boys and girls (53).

The investigator summarized the studies as follows:

1. Monotones, as a group, revealed significant deficiencies in both pitch discrimination and tonal memory, but not in rhythmic memory.
2. Approximately twenty-five percent of boys and twelve percent of girls were monotones between the ages of seven and eight years, the proportions falling to about seven percent boys and one to two percent girls at twelve to thirteen years.
3. After the age of eight years between three and four times as many boys as girls were monotones.
4. Monotonism, as a condition, appeared to be a form of retarded development in the case of most children, but in a small proportion (about four percent of the population) the condition persists throughout childhood (63).

Pedersen and Pedersen (1970) investigated the relationship between pitch recognition and vocal pitch production in sixth-grade students. The investigators sought to determine the level of formal musical training by administering a musical symbols test (266).

The following relationships were investigated: (1) pitch discrimination with vocal pitch production; (2) both pitch discrimination and vocal pitch production with gender;

and (3) both pitch discrimination and vocal pitch production with knowledge of musical symbols (267).

Three measuring instruments were administered to each of two sixth-grade classes. The first class, Group I, (N = 26) consisted of thirteen males and thirteen females. The second class, Group II, (N = 29) was comprised of twelve males and seventeen females (267).

The *Music Symbols Test* (MST) was used to measure knowledge of musical symbols. It was a test consisting of thirteen matching and short-answer items that required the subjects to identify common musical symbols and terms. No time limit was given and each subject had ample time to attempt every item. A split-half reliability estimate (r = .87) was obtained by applying the Spearman-Brown formula to the odd-even correlation (267).

The *Pitch Discrimination Test (PDT)* was given to measure the ability to discriminate whether two tones or sets of tones had the same or different pitches. The test administration was standardized by recording the instructions and test stimuli on magnetic tape (267).

The test consisted of ten trials in each of three series. In all trials a first stimulus was presented, and then three seconds later a comparison stimulus was presented. The three series differed in the number of pitches comprising both the first and comparison stimuli. Series A used single pitches, Series B used an interval of two pitches, and Series C used a sequence of three pitches. The subjects were asked to identify whether the comparison stimulus was the same as or different from the standard stimulus. Responses were written on an answer sheet during the intertribal interval, which was approximately

fifteen seconds. In each series the order of the trials was randomly determined. There were five matched and five unmatched trials for both Series A and C, and four matched and six unmatched trials for Series B (267).

Four scores were obtained for each subject--three subscores representing the number of correct responses on each series, and an overall score representing the total number of correct responses on the test. Split-half reliability estimates were obtained for each of the four scores by applying the *Spearman-Brown* formula to odd-even correlations. The reliability coefficients for Series A, B, C, and the Total test were $r = .65$, $r = .84$, $r = .82$ and $r = .86$, respectively (268).

The *Vocal Pitch Production Test* (VPPT) was given to measure the ability to produce vocally a pitch that had been sounded on a musical instrument. The subjects were to reproduce vocally the standard stimulus presented in each trial during the fifteen-second inter-trial interval. If the first attempt was incorrect (and the subject recognized his error), and second attempt was permitted provided that it was confined to give a complete response. The responses were rated on the following scale:

> 4 = Perfect vocal production of the standard pitch or pitches.
> 3 = Vocal production of the standard pitch or pitches one octave higher or lower.
> 2 = Vocal production of the standard interval pitches or standard sequence pitches in another key but with proper pitch relationships (this did not apply to Series A).
> 1=Correct vocal production of one out of two pitches in Series B or two out of three pitches in Series C.
> 0=Unsuitable vocal response (268).

The *Vocal Pitch Production Test* (VPPT) yielded four scores—three subscores representing the sums of the numerical ratings for each series, and an overall score obtained by summing the three series scores. Both of the test stimuli and vocal pitch productions made by the subjects were recorded for Group II on a second tape recorder.

Subsequently, a qualified independent judge used the same rating scale to rate the performance of each subject from the tape recording. The inter-rater reliabilities for Series A, B, C, and the Total test were $r = .98$, $r = .97$, $r = .97$, and $r = .98$, respectively (268).

The experimental procedure used was identical for both groups. *The Music Symbols Test* (MST) was administered to the entire class. Approximately two weeks later the *Pitch Discrimination Test* (PDT) and the *Vocal Pitch Production Test* (VPPT) were administered individually in a single testing session. The PDT was followed immediately by the VPPT. The three tests' yielded nine scores for each subject: a total test score for the MST, four scores for the PDT, and four scores for the VPPT. The gender of each subject was included as an additional variable. A numerical value of "one" for males and "zero" for females was arbitrarily assigned (268).

Four additional scores were obtained on the *Vocal Pitch Production Test* (VPPT) for the subjects in Group II in order to assess the inter-rater reliability of the VPPT ratings. This required a slight procedural variation for the group as previously described (268).

All forty-five intercorrelations among the ten variables were computed for the two groups separately. The means and standard deviations for each variable were also obtained. Eight different types of relationships were represented by the correlations: (1) correlations of the *Pitch Discrimination Test* scores with the *Vocal Pitch Production Test* scores; (2) correlations of the *Pitch Discrimination Test* scores with the *Music Symbols Test* score; (3) correlations of the *Vocal Pitch Production Test* scores with the *Music Symbols Test Score*; (4) correlations of the various *Pitch Discrimination Test* scores with gender; (5) correlations of the *Vocal Pitch Production Test* scores with

gender; (6) correlations among the *Pitch Discrimination Test* scores; (7) correlations among the *Vocal Pitch Production Test* scores; and (8) correlations of the *Music Symbols Test* score with gender (269).

Results of the study indicated that the subjects in Group I performed better on the average than the subjects in Group II on all tests. An analysis of the means for the *Pitch Discrimination Test* series scores indicated that Series A was the easiest task, followed by Series C and then Series B. The reason for the obtained difficulty was unknown to the investigators. The investigators hypothesized that Series C may have been inherently simpler than Series B because more cues were provided to the subject regarding the similarity of the standard and comparison pitches. The investigators also added that the selected stimuli for Series C may have been inadvertently easier than those for Series B (269).

The investigators concluded that it was not possible to analyze the relative difficulty of all three series in the *Vocal Pitch Production Test* (VPPT). It was noted that the rating scale for Series A involved fewer values than for Series B and C, but note that Series B was easier than Series C. The investigators reported that the finding may have been the result of interval pitch production being inherently easier than sequence vocal pitch production, or that it may have resulted from the selection of simpler pitches for the subjects to produce in Series B. It was concluded that since all the pitches in the VPPT were within the singing range of the subjects, the former explanation seemed more likely (270).

An examination of the standard deviations by the investigators indicated that the lower means for Group II probably resulted from the presence in that group of some subjects

with extremely low scores on the variables. They added that this would have produced both the lower means and the higher standard deviations obtained for Group II (270).

The investigators mentioned that the correlations of principal interest were those between the *Pitch Discrimination Test* (PDT) and *Vocal Pitch Production Test* (VPPT) scores. They reported that those correlations tended to be somewhat larger for *Group II* than for *Group I* and that all but one of those sixteen intercorrelations were significant beyond the $p = .01$ level for Group II. The remaining one was reported to be significant at the $p = .05$ level. For Group I, nine out of sixteen were significant beyond the $p = .05$ level and three of those were significant at the $p = .01$ level. All correlations among corresponding scores on the PDT and VPPT were significant beyond the $p = .05$ level except two for Group I—PDT Series A with VPPT Series A, and the PDT Series B with VPPT Series B. There was a tendency for the correlations among corresponding PDT and VPPT series scores to increase with the complexity of the task. The investigators concluded that there was a fairly strong relationship between pitch discrimination and vocal pitch production. The correlation of $r = .69$ between PDT Total and VPPT Total for Group II indicated that total pitch discrimination accounted for forty-seven percent of the total vocal pitch production variance. The investigators concluded that a correlation of that magnitude provided a basis for fairly good predication, especially in combination with other predictor variables. It was mentioned that the fact that the correlations were somewhat different for the two groups confirmed the desirability of cross-validating results in correlational studies (271).

The investigators reported that the correlations of the *Pitch Discrimination Test* (PDT) scores with the *Music Symbols Test* (MST) score for Groups I and II were not encouraging. They reported that the only significant correlations were between PDT Series B and MST, and between PDT Total and MST for Group I. Also reported was that the bulk of the evidence indicated that although some relationship may exist between pitch discrimination and knowledge of musical symbols, it was not sufficiently large or consistent to have much predictive value (271).

All correlations between the *Vocal Pitch Production Test* (VPPT) scores and the *Music Symbols Test* (MST) score for the two groups were significant beyond the $p = .01$ level except the correlation between VVPT Series C and the MST for Group II (just short of significance), and the correlation between VPPT Total and MST for Group II (significant at the $p = .05$ level). The investigators found that as was true for the correlations between the *Pitch Discrimination Test* (PDT) and the MST, the correlations between the VPPT and the MST were larger for Group I than for Group II. The correlations were not only found to be significant, but also moderately larger. The investigator concluded that the findings indicated that the kind of understanding of music that was measured by the MST was more highly related to vocal pitch production than to pitch discrimination (271).

The examination of the correlations between gender and both the *Pitch Discrimination Test* (PDT) and the *Vocal Pitch Production Test* (VPPT) revealed that only one of the sixteen correlations reached significance. The results indicatwd that sixth-grade males did not differ in their ability either to discriminate or to vocally produce single, interval, or sequence pitches (271).

In regards to the results for the two groups, all except two intercorrelations among the *Pitch Discrimination Test* (PDT) scores were significant beyond either the $p = .05$ or $p = .01$ level. The evidence indicated that the various types of pitch discrimination—single, interval, and sequence—were moderately interrelated. A person who did well on one type of pitch discrimination was likely to do well on the others. It was reported that the correlations between the series scores and the total score were largest, partly because they were part-whole correlations. The investigators reported that whenever overlapping items contributed to both a part score and a whole score, the correlation between the part and the whole was reportedly high. It was reported that the descending order of magnitude of the intercorrelations among the PDT series scores was as follows: rBC, rAB, and rAc. For those subjects Series B and C were most similar, and Series A and C were most dissimilar (272).

The intercorrelations among the *Vocal Pitch Production Test* (VPPT) scores were reported to be extremely large. All were reported to be significant beyond the $p = .01$ level. The investigator concluded that those consistently large correlations indicated that the ability to vocally produce pitches were general and not specific to the type of task. He added that on the other hand, the *Pitch Discrimination Test* (PDT) intercorrelations indicated that pitch discrimination ability was more dependent upon the type of task. It was concluded that those consistent and large correlations probably accounted for the fact that the descending order of magnitude of the intercorrelations among the VPPT series scores was not consistent for Groups I and II as was true for the PDT (272).

The final relationship reported was between gender and the *Music Symbols Test* (*MST*). The correlation was significant for Group II. It was reported negative for Group I but not

significant. The investigators reported that girls tended to perform better on the MST than boys (272).

Zwissler (1972) investigated the pitch discrimination skills of first-grade children identified as accurate singers and those identified as inaccurate singers (1). The study was designed to determine the relationship between the ability of first-grade children to sing and their pitch perception. The study sought to answer the following questions:

1. Is there a significant difference in the pitch discrimination skills, measured over a range of three octaves, of first-grade children identified as accurate singers and those identified as inaccurate singers?
2. Is there a significant difference in the pitch discrimination skills of those two groups as measured in the normal singing range of first-grade children?
3. Is there a significant difference in the pitch discrimination skills of those two groups as measured one octave above the normal singing range of first-grade children?
4. Is there a significant difference in the pitch discrimination skills of those two groups as measured one octave below the normal singing range of first-grade children?
5. Are the pitch differences of larger tone intervals more easily discriminated by first-grade children than those of smaller intervals? (39)

The sample population consisted of 100 first-grade children—fifty classified as accurate singers and fifty as inaccurate singers. The subjects for the study were randomly selected from the twenty-four first-grade classrooms in the Whittier City School District in California (39).

The accurate singer was defined as a child who was "able to correctly reproduce a familiar melody at a given pitch level." The term meant that the child "handled his voice accurately enough to maintain tonality in spite of occasional pitch inaccuracy." Correct rhythm was not a consideration in the study (7).

In October, all first-grade teachers were asked to teach the song *Twinkle, Twinkle, Little Star* to their students (40). The classroom teacher's subjective evaluation of the

children's singing ability was used as the criterion for identifying the two best and the two poorest singers in each classroom. Some teachers were asked to select more than two "best" and two "poorest" singers to ensure sufficient subject numbers for the study. There was mention of how those teachers were selected. An audiometer test was administered by the school nurse to each child selected by the teachers and those with abnormal hearing acuity were eliminated from the study (42).

The following suggestions were made to the first-grade teachers for evaluating the singers:

1. Make a game of children taking turns being the solo singer on the first phrase of *Twinkle, Twinkle, Little Star*.
   a. To choose best singers: listen for students who sing exactly on pitch, i.e., they start on the correct pitch, as sounded on the tone bell C, and they also sing the melody correctly.
   b. To choose "poorest" singers: listen for students who cannot match the beginning pitch or who cannot sing the tune so it is recognizable.
2. Make a point of walking around the class during the music period, listening to individual students. Poor singers will probably be singing at a very low pitch or you can hear that they are not singing accurately. A child who sings at a pitch level either below or above the correct pitch should be classified as a poor singer even if he sometimes appears to be singing the melody correctly at his pitch level.
3. A child who refuses to sing at all cannot be used in this study.
4. Non-English-speaking children should not be chosen for the study (41).

In November, a test of vocal accuracy, Test A, was administered to each student by an examiner who was both a private piano teacher and a certificated public school teacher. The student, identified only by number, was asked to tape record the song *Twinkle, Twinkle, Little Star* and two melodic patterns as described in *Test of Vocal Accuracy* (42).

The investigator listened to the tapes and judged each subject to be an accurate or inaccurate singer, unless he could not be clearly classified as either. In the latter case, the subject was eliminated from the study. Since almost perfect correlation was obtained

between the four music educators who judged the performances of subjects on the test of vocal accuracy in the pilot study, a panel of judges was not deemed necessary for the study (42).

Of the 131 students recommended for the study by the classroom teachers, eleven were eliminated because they failed the audiometer test, three moved before the tests were administered, several were eliminated because they performed only two-thirds of the test correctly, and ten were eliminated because they failed to complete the test of vocal accuracy due to shyness or forgetting the words of the song. The sample population for the study consisted of the remaining 100 students, fifty identified as accurate singers and fifty as inaccurate singers (43).

A Test B, pitch discrimination test, was administered by the investigator to the 100 subjects of the sample. It was not known to the investigator which students were accurate and inaccurate singers. A total pitch discrimination score and a score for each of the three octaves tested were obtained for each subject (43).

A pilot study was conducted in the spring and summer of 1970. Trials of both the test of vocal accuracy, Test A, and the pitch discrimination test, Test B, were performed and appropriate adjustments were made (44).

Pilot study data was used to determine the reliability of Test A, the pitch discrimination test, by the split-half technique. Using the *Spearman Brown* prophecy formula, the estimate of reliability obtained was $r = .78$ (47).

The pitch discrimination instrument was constructed to determine the pitch discrimination skills of the accurate and inaccurate singers in three different tonal ranges: (1) The normal singing range of first-grade children; (2) One octave above the normal

singing range of first-grade children; and (3) One octave below the normal singing range of first-grade children (49).

For the *Test of Vocal Accuracy*, the subject was asked to record the song *Twinkle, Twinkle, Little Star* and two short melodic patterns (C C E E G, eighth notes and quarter note on G) and (Bb Ab G F Eb, all eighth notes and quarter note on Eb) (48). The test of *Vocal Accuracy* was individually administered and required from three to five minutes (49).

The tonal stimuli of the first section were within the frequency range of 261.63 cps to 523.25 cps, those of the second section within the frequency range of 523.25 cps to 1046.50 cps, and those of the third section within the frequency range of 130.81 cps to 261.63 cps. All frequencies were based on the equally tempered scale with A4 = 440 cps (52). The stimuli, pairs of tones, were produced on an electric organ and recorded on magnetic tape. A reel-to-reel tape recorder was used for both the recording and playback. The first tone of each pair was held for one-and-a-half seconds, followed by half a second of silence, then by the second tone held for one-and-a-half seconds. Four seconds of silence occurred between each pair (54).

The reliability of the pitch discrimination test was determined by the split-half technique. Using the correlation coefficients between the odd and even scores, the estimated reliability coefficient for the entire test was obtained by the *Spearman-Brown* prophecy formula. The coefficient was $r = .91$ (54).

The correlation coefficients between chance-halves scores for the accurate singers and for the inaccurate singers were $r = .91$ and $r = .57$, respectively. The difference between

those two coefficients was significant beyond the $p = .01$ level. This indicated to the investigator that the test was less reliable for the inaccurate singers (55).

The test for the study was individually administered by the investigator. Testing time, including the training session and practice items, varied from fifteen to twenty minutes. When the subject entered the room he was told that he would be asked to listen to some musical sounds and answer some questions about them. Attention was called to a set of one-octave diatonic step bells beginning on middle C and numbered from one to eight.

The training sequence for each subject involved having: (1) The subject play up and down the "musical stairs" on bells; (2) The subject sing the sound of the bells as they were played up and down; (3) The subject identify and sing melodic contours that go "up" and "down"; (4) The subject listen to melodic examples that go "up" and "down" and then write down answers on a score sheet (58).

When the investigator was sure that the subject understood the testing procedure, the test was administered. The tape was stopped between the first and second section to explain that the sounds the subject would hear next were sometimes higher than he could sing, but he could "feel" that his voice would want to go up or down. A similar explanation was given between the second and third sections. During the test, comments or instructions were only made if necessary to maintain the subject's interest and attention (58).

Four scores were recorded for each subject:

1. Number of correct discriminations in the normal singing range of first-grade children.
2. Number of correct discriminations in the octave above the normal singing range of first-grade children.

3. Number of correct discriminations in the octave below the normal singing range of first-grade children.
4. Total number of correct discriminations (59)

The investigator reported the following findings from the study data:

1. The major hypothesis of the study was that there is no significant difference between first-grade children identified as accurate singers and those identified as inaccurate singers in their pitch discrimination skills, measured over a range of three octaves.
2. A statistical analysis of the data indicated that there was a highly significant difference in the pitch discrimination skills of these two groups of singers, and the hypothesis was rejected at the $p > .001$ level.
3. Since no significant interaction was found between accuracy of singing and range, the three secondary hypotheses were considered to be subsumed under the major hypothesis and were also rejected. Those hypotheses were that there is no significant difference between first-grade children identified as accurate singers and those identified as inaccurate singers in their pitch discrimination skills, measured: (a) in the normal singing range of first-grade children; (b) one octave above the normal singing range of first-grade children; and (c) one octave below the normal singing range of first-grade children.
4. There was a significant effect for octaves at the $p = .05$ level. All subjects, without regard to singing ability, could identify pitch differences more easily within the normal singing range of first-grade children than within the octave above or below (70).
5. The accurate singers clearly performed better than the inaccurate singers in all three ranges. This did not appear to the investigator to be consistent with the theory that the larynx assists pitch discrimination, since the accurate singers maintained their advantage outside their normal singing range.
6. There was no evidence in the data of the study to indicate that first-grade children hear the pitch differences of the larger tone intervals of an octave (sixths, sevenths, the octave) more easily than those of the smaller intervals (seconds, thirds) (71).

The investigator made the following conclusions within the limitations of the study:

1. The fact that first-grade children who sang accurately in their normal singing range also possess, in general, a more accurate sense of pitch than inaccurate singers indicated a strong relationship between the inability to sing and poor pitch perception. Therefore, the case for teaching pitch discrimination at the kindergarten and primary levels was well established.
2. Since both groups of subjects could identify pitch differences more easily within the octave designated as the normal singing range of first-grade

children than within the octave above or below, early training in pitch discrimination skills should begin within this range. The fact that the accurate singers performed at a higher level than the inaccurate singers on all octaves indicates however, that pitch discrimination training should quickly be expanded to include recognition of pitch differences in ranges above and below the normal singing range of first-grade children.

3. When dealing with the intervals of an octave, the size of the interval did not appear to be significant in the pitch discrimination skills of first-grade children. Therefore, it was not necessary to start training in recognition of pitch direction with the wider intervals of the octave.
4. The fact that five children in the inaccurate singing group scored between seventy and ninety-five percent suggests that factors other than pitch discrimination skills must have been the cause of their inability to sing (72).

The investigator concluded that the study supported the viewpoint of research which suggested that poor pitch perception is an important cause of the singing problems of elementary school children. It was added that it confirmed that pitch discrimination and the ability to sing are closely related (72).

The following recommendations were made by the investigator:

1. Musical experiences, through which children will develop those concepts, should be provided at the kindergarten and first-grade level.
2. An explanation needs to be found for why the inaccurate singers, especially those who sang at a pitch level below the defined normal singing range, also performed better in the normal singing range.
3. Conduct further studies which could define the exact singing range of each child and then test the pitch discrimination skills of each subject in his defined range (78).

The investigator reported that the fact that the pitch discrimination test was less reliable for the inaccurate singers indicated the need for refinement of the test used. Some of the factors to be considered according to the investigator were: (1) how the student will label the response he is to make; (2) the time intervals between stimuli and also between the pairs of stimuli; (3) the problem of inattention of the subject; (4) the use of a standard tone; and (5) the pitch range of the test (78).

The investigator concluded that the study had clearly established that the pitch discrimination skills of first-grade children classified as accurate singers were superior to those of the first-grade children classified as inaccurate singers. It was recommended by the investigator that an attempt should be made to ascertain the importance of the factors of tonal memory and the physiological aspect of vocal production in the ability of very young children to find and effectively use their singing voices (78).

Geringer (1983) examined the relationship between pitch-discrimination and vocal pitch-matching abilities of preschool and fourth-grade children. Specifically, would preschool and fourth-grade subjects grouped according to ability on a pitch-discrimination task show significant differences performing a vocal pitch-matching task? (94)

Subjects were four-year-old and five-year-old children (N =72), randomly selected from a large, ethnically mixed preschool with a total enrollment of 450, and fourth graders, (N =72), randomly selected from five public schools (95).

All experimental procedures had been pilot tested with other samples of four-year-old children (N =18), and fourth graders (N =12). All subjects were tested individually for both the discrimination and pitch-matching tasks. The order of task testing was counter balanced to prevent possible test-order bias (95).

The prerecorded pitch discrimination test (PD) consisted of twelve trials of tonal pairs. The intervallic structures of test items were a descending tritone, and ascending minor third, a descending quarter tone, one ascending and one descending eight tone, and four unison pairs. All tones were produced on a *Johnson Intonation Trainer*, with timbre setting number "two," and were said to be within the vocal ranges of the subjects. The

investigator did not elaborate on the specifics of this device. Three tapes were recorded with different test item orders. All tapes were monitored with a *Conn Chromatic Stroboscope* to insure pitch accuracy. Each tone had a duration of one second with a tenth of a second silence between tones of a pair. The inter-trial duration was ten seconds to allow subjects adequate time to respond. Subjects verbally stated whether the second tone of a pair was the same or not the same as the first tone, or whether they were uncertain. Criterion for aural testing was three consecutive correct discriminations and verbalizations regarding the visual stimuli (95).

The recording used for the vocal pitch matching test (VPM) consisted of a simple, unaccompanied three-measure song with simple words, sung on scale degrees "do," "re," "mi," "re," "do." The master recording was made in three keys, C, E, and F# major by a female vocalist viewing a strobotuner. A panel of graduate students in music education selected the most accurate version in each of the three keys, while viewing the strobe, so that all pitches were within plus or minus five cents relative to equal temperament. The song was presented to individual subjects in each of the three keys, in counterbalanced orders. The experimenter instructed the subjects to sing back the sustained, final tones of the examples, exactly as on the tape recorder. The investigator said that the pitches to be matched were well within the suggested singing ranges for both age groups. Subject responses were tape recorded for subsequent analysis (95).

An analysis of variance indicated significant differences in pitch-matching abilities between age groups of subjects ($p < .01$), but showed no differences among the discrimination-based ability groups on pitch-matching scores ($p > .10$). A Spearman

coefficient revealed a moderate correlation, $r = .61$, between the two measures within the high ability fourth graders (97).

The investigator cautioned the interpretation of the results. He said although extreme care was taken to verify the ability of the preschool children to respond in accordance with their best judgment, it was possible that response biases may have been operating. It was reported that different dependent variables—verbal estimation, production, or reproduction methods—possibly accounted for much of the inconsistency in the research literature. He said because young children lack verbal sophistication, it is especially critical to develop reliable and valid methods of measurement for this population (98).

The investigator said that the relative lack of correlation between pitch discrimination and vocal pitch matching ability raised some questions of relevance to music educators. It was hypothesized that the limitations of pitch discrimination ability do not operate until the vocal pitch matching deviations are small enough to be affected by the limit. It was also hypothesized that pitch discrimination and pitch matching are simply two independent abilities, or that maturation and training are necessary to develop an interrelationship (98).

TONAL MUSIC APTITUDE

Jaffurs (2000) investigated the relationship between singing achievement and tonal music aptitude. Fourth-grade students (N = 56) from three intact classes of a public elementary school in Michigan participated in the study. They were from a large, affluent, suburban area of Detroit. The students received music instruction twice a week

for thirty minutes each session as part of their regular curriculum. The researcher was also the students' music teacher (36).

The study examined the relationship between a student's singing ability and tonal aptitude. *Intermediate Measures of Music Audiation* (IMMA) a standardized aptitude test, was used to measure the developmental, tonal aptitude of the students (36).

Rutkowski's (1986) *Singing Voice Development Measure* (SVDM) was used to rate singing achievement. A short, minor song used in Rutkowski's 1996 study was used as the criterion song. It was decided that the subjects would sing the criterion song both with and without text (39).

Two weeks before the testing, students were taught the criterion song. A five to ten minute lesson was included in each of the fourth-grade music classes to teach the song. The lessons focused on encouraging solo singing of the song so that students became comfortable singing alone (39).

In November 1999, the researcher administered the tonal portion of IMMA, and tape-recorded individual students singing the criterion song. It took the researcher two weeks to administer the singing achievement test to all students. Two weeks prior to administering SVDM, the teacher sent letters to the parents inviting students to participate in the study. The letter included the nature of the testing and assurance of confidentiality of test results. At the time of the study, there were sixty-eight students enrolled in the fourth-grade. Fifty-six students participated in the study (40).

The criterion song for the singing achievement test was taught to the class for two weeks before the testing began. The fourth-grade classes met twice a week, so there were four instructional periods before the actual testing began. Students were taught the song

in a rote, unison, and echo manner. The researcher used five to ten minutes at the beginning of each instructional period to echo sing the song with the class. In each of the four classes, students were randomly chosen and eventually all of the students who participated in the study were given the opportunity to sing individually the criterion song before testing (41).

The testing of the singing achievement began the same week that the students took IMMA. Testing of students was agreed upon and arranged between the researcher and the classroom teachers. The students from the three fourth-grade classrooms were mixed randomly so that individual classes were not intact. Students were sent two at a time to the music room where they were tape-recorded. The researcher played the first three notes of the song and then said "ready sing." The students sang the song individually, once with, and once without the text and the researcher tape recorded the performances (43).

Two judges were asked to listen to the tape and independently rate the singing performances using SVDM. The judges were chosen based on their vocal study and familiarity with the elementary student voice. Both judges were vocal music education teachers in public school districts in Michigan (43).

Results showed that even though scores for singing without text and singing with text both were found to be significantly correlated with tonal aptitude, there was a difference between the strength of the relationship when singing without text and singing with text. It was hypothesized that this was possibly due to the method in which students sang the criterion song. All of the subjects were asked to sing the criterion song without words

first and then with words. It was said that those two should have been alternated to eliminate an order effect (48).

A significant relationship between the singing accuracy scores and the tonal subtest of IMMA was found. The results indicated that students who had high tonal aptitude also tended to be able to use their singing voices as measured by SVDM (50).

The following recommendations were made by the investigator:

1. The SVDM may be inadequate for use with fourth grade students.
2. A standardized measurement for singing accuracy with older children is not available. One should be developed.
3. Replicate the study fixing the design flaw related to text and no text.
4. Train judges in person rather than over the phone.
5. Replicate the study with a larger and more diverse population (52).

TONALITY AND VOCAL RANGE

Flowers and Dunne-Sousa (1990) assessed young children's abilities to echo short pitch patterns in relation to maintenance of a tonal center in self-chosen and taught songs. Additional considerations were (1) age differences in ability to maintain a tonality and echo pitch patterns; (2) accuracy of vocal reproduction in echoing pitch patterns; (3) age differences in use of vocal range; and (4) size of vocal range used for different singing tasks (102).

The subjects for the study were the total enrollment of three-year-old, four-year-old, and five-year-old children at two preschools (N = 93). Prior to testing, the experimenter spent at least four music sessions with the children in order to build rapport and to teach them a new song that they would be asked to sing individually during testing. Following the small group music sessions, each child was individually recorded singing a self-chosen song, singing a taught song, and echoing twenty short pitch patterns (104).

Twenty pitch patterns were developed, consisting of equal numbers of one-note, two-note, three-note, and four-note patterns, including ascending, descending, combination ascending/descending, and single pitch contours. All items were four beats in length. Patterns spanned the range of one octave across the twenty items and included instances of all semitones within the octave. The pitch patterns were sung by a soprano on the neutral syllable *loo* while she viewed a *Korg AT-12 Auto Chromatic Tuner*. They were recorded on cassette tape at six pitch levels: As (the A below middle C) A3, Bb3, Bb4, B3-B4, C3-C4, C#3-C#4, and D3-D4. The different pitch levels were intended to accommodate the most comfortable or preferred singing ranges of the individual children. Furthermore, patterns were sung at two tempos: (1) beat = approximately 60 MM, with eight beats response time (slow), and (2) beat = approximately 80 MM, with six beats response time (fast). The six pitch levels and two tempos resulted in a total of twelve different tapes (104).

Children's vocal performance in echoing the pitch patterns was assessed according to four measures: (1) number of pitches, out of fifty total, sung within plus or minus fifty cents of the recorded vocal model; (2) number of patterns sung entirely correctly, each pitch within plus or minus fifty cents of the vocal model; (3) number of patterns sung with correct intervallic relationships but at an incorrect pitch level—for example, G4, E4 sung as C4, A3; and (4) number of patterns with correct melodic contour but incorrect pitches or intervallic relationships. To be counted as correct on melodic contour, the overall dimensions of the contour had to be maintained. It was intended that comparisons among those four criteria would provide information about children's early attempts to model melodies, in particular, whether they more closely approximated specific pitches

or overall contours. Assessment of pitch pattern performance was accomplished by listening to each tape while viewing a *Korg AT-12 Auto Chromatic Tuner* and determining within plus or minus fifty cents the name of each pitch sung by the child. Two judges independently listened to a random sample of twenty-two percent of the total ninety-three tapes, and reliability was determined by comparing the proportion of agreements to total judgments. The reliability estimate was reportedly high on all four measures of pitch pattern performance: $r = .96$ for number of intervallically correct patterns; and $r = .96$ for number of correct melodic contours (105).

Two measures of the child's ability to maintain a tonal center were established. First, each child's self-chosen song performance was placed into one of the three categories: (1) "modulating"—more than three modulations; (2) "somewhat modulating"—one to three modulations; or (3) "not modulating"—none or one modulation with the song beginning and ending in the same key. Three performances were dropped from the study because they were chanted or otherwise unclassifiable. If a song was too short to apply the given modulation criteria, the experimenter made a judgment concerning the category into which that performance was placed. Interjudge reliability based on a random sample of twenty-two percent of the self-chosen song performances was $r = 1.0$ (105).

A second and more stringent measure of the child's ability to maintain a tonal center was a comparison of the phrases in *The Little White Duck*, which was taught during the preliminary music sessions. This song was chosen because it was unfamiliar and could be taught consistently to children in the study; it encompassed the range of an octave as did the pitch patterns, allowing a comparison of vocal ranges between the two singing tasks; the tessitura of the song fell mid-range; and it consisted mostly of motion in steps

of thirds with a few larger intervals. In assessing the child's performance of *The Little White Duck*, one pitch from each of the six phrases was designated for comparison with the designated pitch of the previous phrase; the first phrase was compared with the starting pitch given by the experimenter. This was intended to provide an overall measure of the child's ability to maintain a tonal center throughout the six phrases of the song. The designated pitch from each phrase was counted as correct if it was sung within plus or minus fifty cents of the proper intervallic relationship to the designated pitch from the phrase directly preceding it. For example, if the child modulated once and subsequently maintained that tonality, a score of five would result. Scores ranged from "zero" to "four"—the best performances had at least two modulations, that is, four phrases that maintained a correct tonal relationship to the preceding phrase. The interjudge reliability estimate on this measure was $r = .80$ based on a random sample of twenty-two percent of the performances. Because this measure and the categorization of self-chosen songs according to number of modulations were intended to provide similar information, a comparison of scores on the two tasks was made. A three ("modulating," "somewhat modulating," or "not modulating" on self chosen song) by four ("zero," "one," "two," or "three" to "four" tonal phrases in *The Little White Duck*) contingency table was produced to make this comparison. A significant relationship was found between the scores on those two measures. That is, the modulation scores assigned to children on the two measures were similar, further indicating consistency in use of the modulation criteria (106).

In each testing session, the child was introduced to the room and made familiar with the recording equipment. Children who were uncomfortable with leaving their classmates or

entering the testing environment were not required to participate in individual testing. Five three-year-olds at Preschool One and four three-year-olds at Preschool Two did not complete the individual singing tasks. All four-year-olds and five-year-olds participated fully (106).

The first task was to sing a self-chosen song. As the child did so, the experimenter recorded the performance and viewed a *Korg Auto Chromatic Tuner* in order to make a general assessment of the child's singing range. Based on this assessment, one of the six pitch pattern tapes was assigned according to the lowest pitch sung comfortably by the child. The first four pitches of *The Little White Duck* were given to the child at the same pitch level that was assigned on the pitch patterns. Following the self-chosen song performance, the child sang the pitch patterns and *The Little White Duck*, the order of which was randomly assigned. Children at Preschool One (N = 42) sang the pitch patterns at the slower tempo. During the testing of those children, it was observed that a faster tempo was desirable to keep their attention during and between items. The faster pitch patterns were sung by the children at Preschool Two (N = 51). Testing time was approximately ten to fifteen minutes and varied considerably depending on the amount of conversation with each child (106).

The study compared aspects of young children's singing on several tasks that were supposedly common in early music classes. The investigators said singing songs were a more functional music skill than echoing pitch patterns, yet echoing melodic fragments was a component of many preschool and elementary music curricula. The investigator noted that it was often hoped that such activities would lead to successful singing; however, that idea was not entirely supported by the results of the study. Ability to

maintain a key center was another factor that was presumed to affect the quality of a child's vocal performance. The study found that forty-seven percent of the children were classified as modulating singers on their self-chosen song. It was found that the ability to maintain a consistent tonality did not seem to be closely related to echoing pitch patterns, although it was found that youngsters who modulated most frequently also had difficulty in reproducing pitches or contours. For three-year-old to five-year-old children maintaining tonality of a song and echo singing pitch patterns appeared to be largely separate skills (111).

The three-year-olds had the most difficulty in correctly approximating contours, yet their responses did not seem totally unrelated to the model performance, even if incorrect, according to the investigator. A typical response on *The Little White Duck* and also many of the self-chosen songs was to modulate whenever a pitch moved beyond the child's physical or self imposed vocal production limits. On the low pitches, it was said that this seemed to be because of physical limitation; on the higher pitches, the child simply did not extend the voice, opting to modulate instead. Following a modulation, the melodic contour of the phrases or fragments was usually intact, but the overall effect was reportedly one of considerable wandering. It was reported that children took more risks in echoing the pitch patterns. It was reported that even though many sang below the given pitch, their overall range was nearly always greater than in singing songs. It was once again reported that those tasks seemed to represent different skills. It was noted that the immediacy of hearing and responding to brief patterns seemed less demanding than the complex self-monitoring over time that is required in singing a song. It was stated that learning to use the higher pitches through echoing exercises might help the child to

develop a singing range that would allow melodic contours to be more accurately produced, perhaps necessitating fewer modulations (113).

The investigators recommended that more research in this area of education is needed. It was also recommended that future research should continue to individualize procedures whenever possible and address the issue of variability as well as mean performance, particularly with very young children (113).

VIBRATO

Yarbrough, Bowers, and Benson (1992) investigated the effects of vibrato on the pitch-matching accuracy of certain and uncertain singers. All children in kindergarten through grade three of a university laboratory school, (N = 200) were selected to participate in the study. Students in grade one, (N = 50) grade two, (N = 51) and grade three, (N = 54) had completed differing levels of the Kodaly-based curriculum used by the school music teacher. Kindergarten students, (N = 45) were beginners in the school music curriculum (32).

Each student was classified as a certain, (N = 91) or uncertain singer, (N = 109) on the basis of their pitch-matching ability as shown in a preceding pitch-matching study. There were ninety-five girls and 105 boys (32).

Stimulus tapes of the models were made to insure maximum accuracy. Each model sang a descending minor third (G to E) using the syllable "la." The female model sang both the nonvibrato and vibrato models. Taped models were analyzed for accuracy by the *Digigram MIDIMIC*, the *Opcode Studio Plus 2 MIDI* interface, *Performer* sequencing software by *Mark of the Unicorn*, and the *Macintosh SE* computer system. Accuracy rates were 100 percent (child), 100 percent (female nonvibrato), and 79.53 percent

(female vibrato). Subjects responded to all three models, with diversionary music in contrasting keys played between each model to control for tonal memory. Order effect was controlled by the rotation of six possible model presentation orders: (1) child, vibrato, nonvibrato; (2) child, nonvibrato, vibrato; (3) vibrato, nonvibrato, child; (4) vibrato, child nonvibrato; (5) nonvibrato, vibrato, child; and (6) nonvibrato, child, vibrato (32).

Consistency of digitizing by the *MIDIMIC* was significant, high, and positive (N = 40, $p < .01$). Consistency of tallying correct pitches was measured by submitting twenty-five percent of subjects' responses to an independent judge. Agreement between the experimenter and the independent judge was $r = .97$. Criterion-related validity was obtained by comparing the music teacher's classifications of certain and uncertain singers to scores obtained by computer analysis. For certain singers, criterion scores were $r = .70$ or better; for uncertain singers, they were less than $r < .70$. Percentage of agreement was $r = .77$ (32).

The results of the study demonstrated that vibrato did affect pitch matching of uncertain singers. The vibrato model containing correct pitches for 79.53 percent of the duration of the model elicited significantly less correct responses from kindergarten through grade three uncertain singers than the nonvibrato model with 100 percent correct pitches. In addition, the vibrato model elicited fewer, though not significantly fewer, correct responses than did the child model that contained 100 percent correct pitches (36).

According to the investigators the presence of the vibrato did not create an immediately perceptible out of tune model. The model sang the descending minor third with a natural, light vibrato rather than an exaggerated one. The more common protocol for determining

pitch accuracy, for example, the number of correct pitches sung or cent deviations from target pitches, according to the investigators, would have judged those pitches as in tune or plus or minus twenty to twenty-five cents deviation. It was said that although this relatively small vocal fluctuation did not affect the pitch matching accuracy of certain singers, it did so significantly for uncertain singers. Thus the small percentage of time, about twenty percent, during which the vibrato model fluctuated from the target pitches seemed to impact the ability of uncertain singers to match the vocal model (36).

The results of the study suggested that when the accuracy of the vocal model decreased by twenty percent or more, that model became detrimental to children who were experiencing pitch matching problems. The study supported the notion that not only is timbre of the voice important but also the ultimate pitch accuracy one can only achieve without vibrato (37).

The following recommendations were made by the investigators for future research:

1. Future research be conducted with combinations of more disparate timbres and pitch inaccuracies to further isolate the parameters within which music educators need to operate in order to effectively remediate the problems of uncertain singers.
2. Encourage those responsible for teacher training in music education programs to use new technology to help prospective teachers perfect vocal modeling skills.
3. For younger children, and especially for uncertain singers, the presence of vibrato in the voice of the teacher should be reserved for solo work and should be kept out of the elementary classroom (37).

VOCAL MODELING

Sims, Moore, and Kuhn (1982) investigated the effects of female and male vocal stimuli, tonal pattern length, and age on vocal pitch-matching abilities of young children from England and the United States. Subjects were drawn from two populations. Thirty

five-year-old and thirty six-year-old children were selected at random from an infant school in a middle class suburban city in England. The children had begun group music training at the age of five, receiving approximately twenty minutes of instruction five times per week from a male instructor (105).

The second group was also selected in a similar manner from a middle class suburban public elementary school in the United States. The sixty children had received school music training beginning at age five, meeting with a female music specialist once a week for thirty minutes (105).

The pitch-making test consisted of a series of twenty different tonal patterns presented twice, plus one practice pattern. Pre-recorded items were presented to subjects on cassette tape. Each item was sung once for the subjects, who immediately tried to echo the pattern they had heard. The series was presented once by a female mezzo soprano voice and once by a male baritone voice, with half of the subjects responding to the female vocal stimulus first, and the other half responding to the male stimulus first. All testing took place in individual ten-minute sessions (105).

Items were chosen for their musicality and singability. There were four each of patterns containing one, two, three, and four pitches, and two five-note and six-note patterns. Items of differing lengths were interspersed randomly to form the twenty pattern series. Within the patterns there were twenty-two ascending and twenty descending intervals, ranging in size from minor seconds to one skip of an octave, all within the range from middle C4 through C5 one octave higher. The male stimulus sounded one octave lower. All patterns were sung with one pitch per beat, MM = 68. For each pattern, half of the patterns were sung with words set syllabically to the pitches,

while the other half were sung melismatically on a neutral syllable. *The Kuder-Richardson 20* estimate of test reliability for responses to the female stimulus was $r = .96$ and for the male stimulus, $r = .95$ (105).

Data generated on the pitch-matching test was subjected to a five-factor analysis of variance with repeated measures on two factors. A data point was the percentage of pitches sung correctly, within a quarter step of the actual pitch, per each pattern length for each stimulus. For example, the number of pitches sung correctly in all four, three-note patterns sung under the female stimulus conditions were totaled, and then converted into a percentage to become one data point, as were all the pitches in three-note patterns to which a subject responded correctly under the male stimulus. This was done for every pattern length, resulting in two data points for each of the six pattern lengths, for a total of twelve data points per subject, (N = 120) (105).

The results of the ANOVA indicated that a significant difference existed between countries. It was reported that the English children scored significantly higher on the pitch-matching test than did the American children (105).

It was also reported that a significant difference was present between responses to the vocal stimuli. A significantly higher percentage of pitches was sung correctly under the female vocal stimulus than under the male stimulus condition (105).

Another significant main effect was pattern length. Analysis revealed that while there was no difference between the percentage of pitches sung correctly in five-note and six-note patterns, all other pattern length comparisons were significantly different, with each successive length from one through six eliciting a lower percentage of correct responses (105).

No significant main effects were found between age groups or stimulus orders. The only significant interactions were country by stimulus, stimulus order by stimulus, country by pattern length, stimulus by pattern length, and country by stimulus by pattern length (106).

The researchers recommended that further research is needed to explore other variables which may affect pitch matching ability, as well as teaching techniques which would help to increase the number or correct pitch matching responses. It was also recommended that future studies should present stimulus items which are centered in the child's comfortable singing range (107).

Small and McCachern (1983) investigated whether children in first grade could, prior to practice, match pitch more accurately with a female than with a male vocal model. The investigators also sought to determine whether there would be any significant difference in pitch-matching accuracy between groups of first grade children after a period of practice with a female versus a male vocal model (227).

All first grade students in an urban public school were invited to participate in the study. Subjects receiving parental permission to take part, (N = 55), were given two pretests for pitch-matching accuracy. In an attempt to approximate classroom singing activities as well as provide for first graders' ability to relate to the performance task, the investigators used originally composed song fragments for testing stimuli. The music was sung with words (229).

All testing and treatment material along with instructions were tape recorded on a reel-to-reel recorder by a female and a male vocal model. Animal puppets, one with a typically masculine name and one with a typically feminen name, were manipulated by

the experimenters to enact respective male and female recorded voices so that first-graders might relate to instructions and experimental stimuli (229).

Stimulus material for one pretest, Model F, consisted of three melodic fragments tape recorded by a female vocal model. The other pretest, Model M, consisted of three different melodic fragments tape recorded by a male vocal model. All of the test song fragments consisted of two measures of 2/4 meter—two quarter notes and a half note moving at MM: = 56. Model F Test contained three different orders of the stimulus tones "do," "re," and "mi" while Model M Test contained three other different orders of those scale degrees. The six different arrangements resulted from all possible combinations of "do," "re," and "mi." On each test the fragments were written in three different keys: Db, D, and Eb major (229).

Subjects were pretested and posttested individually. Intensity on the tape recorder playing stimulus material was maintained at a comfortable, easily heard level (229).

After the subject entered the testing room, spoken and sung "interactions" between puppet and subject were encouraged through question-answer and repetition-type trials. The descending minor third was used in the vocal stimulus at that point. It was reported that there was ample silent spacing on the stimulus tape for subjects to respond verbally or vocally. Following the warm up period, the subject was asked to repeat the test song fragments in response to the puppet's "singing." Vocal test responses were recorded on a cassette tape recorder. After having received two trials on each of the three song fragments, with eighteen pitch trials, the subject was dismissed (230).

All subjects with perfect pitch matching scores on both pretests, (N = 8) were eliminated from further experimental participation. Remaining subjects were randomly

assigned to one of three groups: (1) practice with female model; (2) practice with male model; and (3) no contact control. Each day for five days subjects in the practice groups went collectively to two different rooms for a thirty-minute practice session with the respective male or female vocal model. Four different and previously unused song fragments in varying keys with varying texts were presented during each practice session. The song fragments were taken from children's existing song literature. In the practice sessions, the same male and female model puppets used in the pretests were manipulated to pantomime the instructions and singing that were recorded on tape. Since each song fragment contained a previously unused text, part of the practice session was devoted to verbal repetition of the new words to be used in that session (230).

Song fragments presented in the practice sessions were similar to the ones used in the pretests. The four fragments were presented in sequence, one fragment per child, until all subjects had rehearsed individually all four song fragments. Two presentations of each fragment, along with appropriate silent spacing between presentations, provided each subject with two rehearsals of each of the four fragments (230).

On the first treatment day, subjects were allowed to choose their own seating. During the remaining days, subjects drew numbered pennies indicating respective chairs and prescribing a chance seating arrangment for each day. Subjects held the pennies as long as they maintained appropriate behaviors and pennies were taken away when subjects failed to cooperate. Later in the week, additional pennies with stars on them were distributed for good behavior. Pennies were not given as reinforcers of correct singing, nor was feedback given specifically regarding correct or incorrect pitch responses. Positive verbal and physical expressions followed appropriate behaviors and

participation. It was the intention of the investigators to study vocal modeling, apart from specific feedback, as a possible influence on first graders' pitch accuracy (230).

At the end of the five-day rehearsal period, forty-four subjects were posttested individually with the Model F and Model M tests used at the beginning of the experiment. Three subjects were not available for posttesting (230).

The results of the study were reported as the following:

1. There were first graders who had little difficulty matching pitch whether the vocal model was male or female.
2. There were first graders who experienced pitch-matching difficulty with both models.
3. The data indicated that subjects had somewhat more difficulty matching pitch with the male vocal model than with the female model, however, it was observed that the grand means were not widely disparate.
4. It was found that practice with vocal models apart from feedback produced no significant differences in pitch-matching accuracy means (232).

The investigators noted that a larger number of subjects could have produced a greater difference in accuracy means and more clear-cut musical conclusions. This was recommended for future studies (232).

Montgomery (1988) investigated the effect of two means of vocal modeling by a male music teacher on third-grade children's vocal accuracy in singing pitch patterns. The two vocal modeling techniques were the use of the falsetto voice in the child's range and the use of the normal voice of the adult male, one octave below the child's range (5).

Students in two intact classes of an elementary school in Danville, Virginia, served as the subjects for the study. The school was described as middle-class by the director of instruction and consisted of sixty percent black students and forty percent white students (26).

Both classes were taught by the investigator, formerly an elementary music specialist himself. Each class comprised of twenty-one students, was randomly assigned to one of two experimental treatments: E1, in which the falsetto voice was used for vocal modeling, and E2, in which the normal male voice was used (26).

The design of the study consisted of four phases: (a) a two-week orientation period, (b) a vocal pretest, (c) a twelve-week treatment period, and (d) a posttest identical to the pretest. During the orientation period, the investigator taught the classes twice weekly. The instruction was comprised of music listening, rhythmic activities, and group singing aided by the use of recordings. To avoid contamination, no male vocal models were included in those recordings; all models were either children or adult female. At no time during the introductory sessions did the investigator use his singing voice, thus controlling this factor as a contaminating influence on the treatment (27).

The test instrument was a form of the *Boardman Test of Vocal Accuracy*. The Boardman test, used melodic patterns rather than a single tone and required subjects to sing a melodic pattern after it was presented three times (27). The patterns used in the Boardman test were said by the investigator to be similar to those found in children's vocal literature of the time (28).

The investigator's analysis of the Boardman test revealed that the pitch levels of six of the patterns were higher than those recommended in research by Wilson (1970), Young (1976), Vaughn (1980), and Goetze (1985). Those patterns were therefore transposed from the original keys to a lower key which encompassed the range of C4 to E5 (28). The syllable "loo" was used instead of text in singing the test patterns (29).

Boardman reported a reliability coefficient of $r = .97$ which was established using a split-half reliability coefficient randomly selected and adjusted for length through the *Spearman Brown Prophecy Formula.* Reliability of the study form of the test instrument was computed on the pretest scores using a split-halves reliability correlation with *Spearman-Brown* adjustment for test length, resulting in a coefficient of $r = .93$ (29).

To produce the test tape, ten patterns were randomly selected and each recorded twice: once by the investigator in falsetto voice, and once in normal voice. The twenty patterns were recorded in random order using a cassette recorder (29).

A pretest was administered for purposes of establishing the level of vocal accuracy at the outset of the study. The subjects were tested in a quiet room equipped with a low table, two chairs, two cassette recorders, a microphone, and two pairs of headphones. The time required for testing each subject was nine minutes (30).

To maintain consistency all instructions were written and read to each subject. To prepare the subjects for the test procedure, the investigator recorded at the beginning of the test tape one sample pattern in falsetto, and one in normal vocal range. The subject was instructed to respond by echoing. When it was apparent to the investigator that the subject understood the procedure, the pretest was administered (30).

The twelve-week treatment period followed the pretesting procedure. The investigator met each class twice weekly for thirty-minute instruction periods. At least half of each thirty-minute class period was devoted to singing instruction comprised of echo-singing of pitch patterns and learning of rote songs. All songs were taught using the syllable "loo." The investigator first sang the entire song using "loo," then divided the song into phrases to be echoed by the students. This procedure was designed to facilitate tonal

accuracy before the addition of text. Only the concepts of tempo, dynamics, rhythm, timbre, and form were studied to avoid contamination from emphasis on melodic content. Each class received equal amounts of instruction in each concept area. Class periods included singing, performing on classroom percussion instruments, listening to recorded instrumental selections, and movement activities (31).

To evaluate the consistency in instruction, fifty percent of the classes were observed by the school's music specialist. The observer completed an observation form which indicated that the appropriate means of vocal modeling was used consistently with each group (31).

At the end of the twelve-week treatment period the pretest was readministered as a posttest. The investigator replicated the testing conditions and procedures used in the pretesting (31).

In order to assess subjects' vocal accuracy, their recorded responses were transcribed by the investigator into music notation. The vocal tests were scored using the following procedure established by Boardman (1964) for each of the twenty patterns:

7 = Accurate matching of all tones in the pattern, without hesitation.
6 = The child "slid" into one or more of the pitches in the pattern, but eventually sang all accurately.
5 = An exact transposition of the pattern.
4 = The child maintained the general contour of the pattern, but sang incorrect intervals.
3 = The child maintained the general direction of the pattern but not the exact contour.
2 = Responses which ignored the contour of the pattern.
1 = The child spoke rather than sang a response or did not respond at all.(36).

Data was collected on forty-two subjects, two of whom were not included in the statistical analysis due to their being older than eight. Data on both pretest and posttest

were numerical scores based on an evaluation of the transcription of an individual subject's recorded responses. Each subject had four scores, each between ten and seventy: pretest falsetto (F1), posttest normal (N1), posttest falsetto (F2), and posttest normal (N2) (32).

To determine the reliability of the scoring, an independent rater, who was an experienced musician, scored a random sample of twenty percent of the taped responses. Correlations between rater and researcher were computed and resulted in a coefficient of $r = .99$ (33).

The dependent variable was vocal accuracy as measured by the *Boardman Test*. The independent variables were: (a) means of instruction (E1/E2), and (b) mode of testing (falsetto voice/normal voice) (33).

The data obtained from the pretest (total score: F1 + N1, with a possible maximum of 140 points) were analyzed by a t-test procedure to determine if there was a significant difference between groups at the outset of the study. Posttest scores of vocal accuracy (each student having and *F* and an *N* subscore maximum of seventy points each) were analyzed by a two-way analysis of variance (ANOVA) to determine if there were significant differences on the dependent variable due to: (a) means of instruction—comparing E1 with E2 on total score (F2 + N2); (b) mode of testing-comparing the F2 scores for combined groups with N2 scores for combined groups; and (c) interaction between means of instruction and mode of testing. An alpha level was established at $p = .05$ (33).

Pretest data analysis revealed that the *Boardman Test of Vocal Accuracy*, used as a pretest had a mean score for E1 (instruction using falsetto) of M = 89.3 and a standard deviation of SD = 17.59. The mean score for E2 (instruction using normal range) was M = 84.6 with a standard deviation of SE = 16.89 (35).

The investigator determined that because the obtained t value was $t < 2.021$ (the critical value for thirty-eight degrees of freedom at the $p = .05$ alpha level), it was concluded that the group means for E1 and E2 were not significantly different. This indicated that the groups were not significantly different on the *Boardman Test* at the outset of the study (35).

Having determined that the groups were comparable, the investigator used an analysis of variance (ANOVA) procedure to analyze the data (36). Group means scores on the two types of pitch patterns used in the posttest indicated that for E1, the mean score on falsetto patterns was M = 55.56, and the mean score for normal range patterns was M = 44.90. For E2, the mean score for falsetto patterns was M = 51.60, and the mean score for normal voice patterns was M = 46.50. For the combined groups the mean score for falsetto patterns was M = 53.63, and for normal voice patterns the mean score was M = 45.70. It was observed by the investigator that both groups scored higher in response to patterns modeled in falsetto than to those modeled in normal voice (36).

ANOVA results for the posttest were reported as the following:

1. The first main effect, means of instruction, was not significant ($F = .29$, $df = 1$, $p = .59$).
2. The second main effect, mode of testing, was significant ($F = 12.24$, $df = 1$, $p = .01$).

3. A significant effect within combined groups due to the mode of testing was reported. That is, all subjects regardless of means of instruction responded with greater accuracy to patterns modeled in falsetto than to patterns modeled in normal voice (37).
4. Interaction between the independent variables in the study (means of instruction and mode of testing) existed when the effect of the method of instruction upon the vocal accuracy scores was not the same in the two different modes of testing (38).

A graphic representation of mean posttest scores was used to illustrate a plotting of cell means in order to explain the nature of interaction between the independent variables. ANOVA was nonsignificant at the $p = .05$ level but it was found that minimal interaction did occur (39).

The investigator concluded that the results of the study indicated that at the outset of the experiment the group mean for E1 was M = 89.3 and for E2 was M = 84.6. It was reported that both groups showed gains after instruction, increasing their mean scores to M = 100.5 (E1) and M = 98.10 (E2). It was also reported that the differences were not significant at $p = .05$ level, indicating to the investigator that means of instruction did not affect students' demonstration of vocal accuracy in matching pitch patterns as measured by the Boardman test. The investigator claimed that given the fact that all subjects demonstrated greater vocal accuracy in response to patterns modeled in falsetto, one might expect that the posttest scores of the group instructed in falsetto (E1) would significantly surpass those of subjects instructed in normal voice. It was reported that results of the study indicated otherwise. It was hypothesized by the investigator that in the study the effect of the octave difference between the subjects' and the investigator's vocal ranges was minimized by the investigator's consistent use of a light vocal quality within the normal tessitura of the adult male voice. The investigator hypothesized that had a heavier vocal production model been used (i.e., sung in chest register, resulting in

louder and more forceful singing), the subjects might have attempted to imitate the timbre of the model by singing in chest voice. It was reported that by doing so, subjects would have sacrificed vocal accuracy since it was noted that the chest voice range was relatively restricted (44).

The investigator reported that the gain in mean scores could have been attributed to the following factors which individually or collectively may have affected a change in scores:

1. Subjects' familiarity with the testing procedure or test-wiseness, may have been responsible for improved posttest scores.
2. Subjects experienced less anxiety on the posttest because they were then more familiar with the investigator who administered the tests and thus improved their scores (44).
3. Singing on a neutral syllable was an unfamiliar task to students on the pretest and adversely affected the scores.
4. The consistent use of "loo" during the treatment period may have accounted for the increase in vocal accuracy as well as for the lack of hesitation or use of incorrect syllables on the posttest (45).

Analysis of differences between individuals' pretest and posttest scores also revealed that most students, (N = 98) demonstrated greater pitch accuracy following vocal and melodic instruction by a male teacher. It was hypothesized that the consistent and concentrated attention given not only to modeling and echo singing, but also to teaching of melodies in phrase units was as beneficial as the means of modeling employed (45).

The investigator reported that during testing, children sang with significantly greater accuracy in response to patterns modeled in falsetto rather than in normal range. The investigator concluded that the group means indicated that it would appear that the use of falsetto by a male teacher is an effective means of vocal modeling (45).

Several recommendations were made by the investigator:

1. Replication of the study with a different and larger sample.
2. Replication of the study should include a longer treatment period.
3. Replication of the study including several grade levels.
4. Replication of the study to determine the effect of vocal modeling in terms of pitch-giving skills for initiating singing. The question was "is it necessary for the male to use falsetto voice only for establishing pitch, and then just as effective to revert to normal range?"
5. Replication of the study with the addition of a third group instructed by a male using a heavy vocal quality within the normal tessitura of his voice.
6. Further study is needed to investigate the association of group instruction and individual testing.
7. Research is needed concerning the stereotyping of males in elementary music education.
8. Male elementary music teachers should be encouraged not only to use their normal singing voice in vocal modeling, but also to use falsetto voice as frequently as possible to encourage accurate pitch matching by students (45).

Gratton (1992) investigated the effect of three vocal models on inaccurate singers' ability to match and discriminate pitches. Fifty-six elementary school children between six and eight years of age participated in the study. Each of them was determined to be an inaccurate singer by their perspective music teachers. The subjects attended one of three Montreal schools. Subjects were tested on a pitch-matching test. The inaccurate singers used in the study were identified as the children who could not vocally match four different pitches out of the five tested (17).

The equipment utilized in the pretests and posttests included an alto recorder, and an electronic tuner, *Korg DT-1*. Tests and training stimuli were recorded with a microphone and a cassette deck. Vocal samples for the subjects working with their own voices as stimuli were created with a *Prophet 2000* keyboard sampler. All sounds were presented to students through a portable tape recorder (17).

The pretest and the posttest were identical. The timbre used during both tests was that of an alto recorder. The pretest consisted of two subtests. The first subtest dealt with pitch discrimination. It consisted of ten prerecorded intervals of different sizes. Those intervals were all within the range of an octave (A3 to A4). Intervals were presented in an increasing order of difficulty from the octave to the minor second. The students' task was to verbally state whether the second tone of the pair was higher or lower than the first tone (18).

The second subtest required children to match vocally pitches presented to them. Students attempted to match five different pitches: the personal note, a half tone and a whole tone higher than the personal note, and a half tone and a whole tone lower than the personal note. Each singer's predominant emitted pitch had been referred to as the singer's "personal note" (19).

Vocal stimuli were recorded on seven different tapes. Three tapes were prepared for pitch discrimination training, and four were prepared for pitch matching training (19). The three pitch discrimination tapes were identical, except for the timbre used. They contained sung intervals consisting of all intervals an octave or smaller in size other than the major seventh and the tritone. The pitch range of those intervals was G3 to A4 (20).

The three pitch matching tapes were also identical, except for their timbres. All chromatic tones between F#3 and C5 were recorded on those tapes. The range was said to be large enough to include the personal notes and pitches surrounding them for all subjects in the study (20).

Pitch-matching exercises on tape seven were custom made for each student in the "own voices" group. In order to produce the sounds required for each of the exercises, personal

notes from all subjects in that group were recorded on tape and transferred to the keyboard sampler. Sampled sounds were then created for the five pitches each subject was trained on. Tape seven thus contained fourteen different pitch-matching exercises, each of which consisted of the recording of a subject singing his or her personal note, and recordings of the subject's voice singing four different pitches: a half tone and a whole tone higher than the personal note, and a half tone and a whole tone lower than the personal note (20).

A four-group pretest-posttest design was employed. Fourteen children were randomly assigned to one of the following groups:

1. Own Voices group: Training in pitch matching was attempted using subjects' own voices. Training in aural discrimination was attempted using sounds from a model child voice of the same gender as that of the subject. The two model children sung with accurate intonation and good tone quality.
2. Model Child group: Training in pitch matching and aural discrimination was attempted using sounds from a model child's voice of the same gender as that of the subject. Both this group and the own voices group worked with the same pitch discrimination tapes.
3. Female Adult group: Training in pitch matching and aural discrimination was attempted using sounds from an adult female voice.
4. Control group: Subjects in this group received no training (21).

The students and the investigator met individually for ten minutes a day on ten consecutive school days. Each child took the pretest and posttest on the first and the tenth day, respectively. Approximately eighty minutes of training were given to every student in the three treatment groups (21).

A training program to be used with every student in the treatment groups was designed by the investigator. However, the investigator said that in order to meet individual needs, this program was used flexibly throughout the training period. Approximately seventy-five percent of treatment time was spent on developing pitch-matching and

discrimination skills with drill exercises. The rest of the time was used for testing, games, and concept labeling (22).

In regards to the "personal note," it had been found by the investigator that almost all uncertain singers were capable of producing at least one pitch. During initial meetings between each student and the investigator, the personal note of every subject was obtained. The personal notes from subjects in the "own voices" group were recorded in order to prepare their pitch matching exercises (22).

Before children in treatment groups attempted to vocally match pitches, they first produced a variety of inflections using their speaking voice. The objective of those exercises was to increase awareness of the fluctuation of the voice, and to attempt a transition from the speaking to the singing voice. Series of expressive word formulations permitting the exploration of a variety of vocal inflections were prepared. Key formulas such as "Oh!," "Eureka!," "Ah, oui?" or "Ding, dong!" were first expressively spoken by subjects. Children gradually attempted to sustain the words' syllables so that the inflections became pitches (23).

Before children in the treatment groups attempted to discriminate pitches, the terms "high" and "low," in reference to pitch, were described. During this time, children first played high and low notes on an alto metallophone placed vertically. After subjects showed the correct location of high and low notes on the metallophone, the following high/low discrimination games were played using the instrument (23).

The investigator and the child sat facing the vertically placed metallophone. During the first game, the investigator played two notes on the instrument and the child played back either the highest or the lowest of the two notes. After this game was mastered, a second

game was played. The investigator asked the child to play two notes in the high/low order. The child then closed his eyes and the investigator played back the two notes in the same or in a different order. The subject's task was to say whether the interval had been played in the high/low or in the low/high order. The objective of the games was to distinguish high and low from other musical concepts. When a child's answers were "high" or "low," terminology for other concepts was not introduced. When answers were "soft," "loud," "short," or "long," an attempt was made to distinguish those concepts from "high" and "low" (23).

For pitch discrimination training, subjects in the "own voices" group were trained using the same tapes as those used in the model child group. This was necessary because the creation of a discrimination tape required accurate reproduction of fourteen different pitches, whereas only one vocal sample from each subject in the "own voices" group was available. The subjects' task was to say if the second tone of a pair was higher or lower than the first tone. Three successive correct responses were required before training was begun on a new interval. Children were first trained on the octave. Subsequent intervals became successively narrower. The final interval tested was the minor second. Children who were unsure about the concepts of "high" and "low" received additional exercises on the metallophone. If they became discouraged or fatigued, pitch discrimination training was interrupted and children received pitch-matching exercises instead (24).

Pitch-matching exercises employed successive approximations. Subjects were instructed to wait for the sound to end before trying to vocally reproduce it (24).

Reinforcement consisted of verbal approval. It was given in the following manner: the first answer given was rewarded. It also determined an accuracy criterion to be attained.

This criterion was usually a response that was a half step more accurate than the initial response. Subsequent responses were rewarded only if they were at least as accurate as the new criterion response. After three successive responses, a new criterion was selected. This was the next closest pitch to the correct response. A maximum of eight minutes per session was spent on pitch matching exercises. When subjects became fatigued or discouraged, training was interrupted. Subjects then attempted pitch discrimination exercises (25).

The results of the study were reported by the investigator as follows:

1. An analysis of variance for pitch matching errors during training revealed a significant relationship between the type of timbral treatment and the number of errors.
2. Analysis showed that subjects made fewer errors when matching pitches derived from their own voices than did subjects who matched pitches of a model child's voice.
3. There were no significant differences in errors made between either of those groups and the female adult group (41).

The investigator discussed the following factors observed during the pitch-matching training of the "own voices" group to try to explain why sounds from one's own voice were matched with fewer errors than those derived from a child's voice:

1. An emotional component was observed to be present which may have caused a higher resistance to fatigue. Subjects who heard their own voices appeared to be more excited and interested than did subjects in the other two groups (41).
2. Two factors may have been a function of a familiarity effect:
    a. The recognition of one's own voice seemed to have helped children from the "own voices" group concentrate their efforts on pitches instead of on other irrelevant elements of the sounds.
    b. Unfamiliarity with the timbre of the stimuli in the two other experimental groups seemed to have caused periodical loss of attention to the task's relevant elements (42).
    c. The rapid recognition of the personal notes may have been another factor in the function of familiarity. Most subjects in the own voices group were said to match their personal note after very few

trials, and subjects in the other groups required significantly more trials to match to the same criterion.

3. The procedure of recording a child's personal note was said to have had some treatment value. Children from the "own voices" group were the only subjects from whom personal notes were recorded, and all showed remarkable enthusiasm, according to the investigator, during his procedure (42).
4. The investigator reported that the difficulties encountered by subjects in the model child group seemed to have been due to the relatively poor vocal quality of the model children as compared to the female vocal model. The investigator recommended that voices of nine-year-old or ten-year-old models might have provided a stronger vocal model (42).

The investigator made the following observations and interpretation of results:

1. Matching to a whole tone below the personal note was reportedly more difficult than matching to any of the other four tones for subjects in the female adult and the model child groups.
2. Subjects in the model child group made significantly fewer errors than subjects in the female adult group on the pitch discrimination task.
3. Errors made by subjects in the "own voices" group, who were working with the same tape as that of subjects in the model child group, did not significantly differ from errors made by subjects in the model child group.
4. The lack of a difference between those two groups suggested that if it was impossible to develop an extended pitch discrimination tape with a synthesized vocal sample of one's own voice, a model child's voice might be a good alternative (43).
5. Subjects in the "own voices" group made fewer errors than subjects in the female adult group when discriminating pitches, however, there was not a statistically significant difference between the two groups. The investigator hypothesized that that may have been a consequence of the two following factors: (1) Subjects from the "own voices" group were working with a novel timbre; (2) Subjects from the own voices group may have been influenced by an interesting timbral effect as follows: On the tapes of the two model children, singing large intervals in the range of A3 to A4 required a change of vocal register. Such a change involved a variation in vocal quality. Children from the model child group seemed to have heard this effect earlier in their training than did children from the own voices group. The difference in vocal quality between high and low notes when performed by the female adult voice appeared to be much less evident (44).

Results from the study lead the investigator to make the following summary:

1. Similar timbres affected uncertain singers' abilities to match and discriminate pitches.
2. The child's own voice seemed to have been more accurately matched than a model child's voice.
3. A female adult voice seemed to have been matched as accurately as the child's own voice but the "own voice" timbre was the only one to have created a familiarity effect.
4. A six-year-old model child's voice seemed to have been more accurately discriminated than an adult voice because of the variations of the vocal qualities of a child's voice.
5. It was not possible to use a pitch discrimination exercise with a child's own voice as a stimulus in the study (44).
6. The investigator reported that the significant increase in correct responses from pretest to posttest for all three experimental groups on the pitch matching subtest, as compared to no increase in correct responses for subjects in the control group, may have reflected the efficiency of the successive approximation technique.
7. The use of approval may also have prevented discouragement and diminution of interest.
8. The use of a vertically placed metallophone may have been a key element in increasing comprehension of the two concepts of "high" and "low."
9. The use of vocal inflections appeared to be helpful in placing some singing voices and in finding personal notes (45).
10. All experimental groups improved from the pretest to the posttest on pitch matching and pitch discrimination.
11. However, it was also reported that there were no significant differences between them.
12. The difference between the experimental groups and the control group on the posttest for pitch discrimination approached significance.
13. In general, the larger intervals on the test were well discriminated.
14. Unequal duration of training could have also influenced the results. Children had ten items to discriminate and five to match but their training time was equally divided between the two tasks. Children progressed from the larger to smaller intervals. Some subjects reportedly did not have the time to complete the last items of the pitch discrimination training.
15. Results for pitch discrimination showed that the average gain scores from the pretest to the posttest were significantly greater for the three experimental groups than for the control group.
16. The overall absence of pretest-posttest differences indicated that subjects needed further training in pitch discrimination.
17. It was reported by the investigator that the interaction between tests and treatments approached significance.
18. Experimental groups appeared to improve, and the control group did not (46).

Recommendations for future study were made as follows by the investigator:

1. In subsequent research using the same age population, older children's voices might be used as stimuli for comparison with a female adult voice as well as with one's own voice.
2. An older child's voice is more mature than the voice of a six year old and might represent a better model for inaccurate singers.
3. The creation of a pitch discrimination tape for subjects in the own voices group would be desirable in order to compare such a vocal timbre with others.
4. Replication of the study might be attempted without pairing a subject with a model child of the same gender as that of the subject. Instead, matching every subject of one group to one model child voice might simplify procedures (48).

Green (1990) investigated the effects of adult female, adult male, and child vocal modeling on the pitch-matching accuracy of children in grades one through six. Subjects for the study were 282 elementary students enrolled at a major southern university laboratory school during the 1986-1987 school year. This number included the entire population of grades one through six with the exception of two students who were not present at the beginning of the study and were eliminated from further participation. Audiotapes of three different model voices—an adult female soprano, and adult male tenor, and a nine-year-old child's unchanged voice were prepared for use in the study (227).

The descending minor third was chosen as the stimulus interval. Pitches constituting the interval were G4 and E4 above middle C4, because it was determined by the investigator to be in the comfortable singing range of preschool and first-grade children. Model stimuli were sung on the neutral syllable "la" to decrease changes in vocal timbre. The descending minor third was sung and recorded by each of the three model voices at a tempo of quarter note = 60 beats per minute, with each pitch receiving one beat (227).

Each subject in the study was tested individually for pitch matching accuracy on three separate occasions, each time responding to a different vocal model. The interval and pitches sung were identical for each of the three tests, the only difference being the model voice. The order of the tests was identical for all subjects: female model, male model, and child model. The three tests were given at seven-day intervals to correct for the confounding variable of tonal memory. When each subject entered the testing room, she or he was told to stand behind the microphone and sing the notes heard on the tape. Each subject was given only one chance to respond to the model voice; however, the investigator mentioned that all subjects were familiar with echo singing and the sol-mi interval because of their Kodaly-based music instruction. Headphones were not used for playing the stimulus voice so as not to interfere with the subject's ability to hear his or her own voice (227).

At the conclusion of the three tests, all audiotapes containing subjects' responses to each of the three models were heard and evaluated. A *Korg Auto Chromatic Tuner, Model No. AT-12*, was used to evaluate the accuracy of each taped response. Each individual pitch of the interval was evaluated, resulting in six pitches per subject, two for each model. Deviation, sharp, or flat, from the model pitch stimulus was measured in increments of 100 cents. A pitch response was considered accurate if the deviation from the model pitch was less than 100 cents. The two pitches for each model were then averaged into one deviation score, resulting in three data points per subject, one for each model. Each data point represented either a correct, flat, or sharp response to each model for each subject. Three trained evaluators independently analyzed forty percent of the total response with the Korg tuner. Interjudge reliability was $r = .93$ (227).

The investigator reported that results from the study indicated the following:

1. Adult female, adult male, and child vocal modeling had an effect on the pitch matching ability of students in grades one through six.
2. There were more total correct responses to the child model, with the least number of correct responses elicited by the male model.
3. Incorrect responses were likely to be flat for the female and male models and sharp for the child model.
4. Boys in the study tended to sing flat more often with the male model than with the female or child model.
5. Subjects in grades one and six sang the highest percentage of flat response for all three models.
6. For the female model only, there was an increase in accuracy from grades one to five, with a decrease in accuracy in grade six.
7. For the male and child models, there was not consistent increase or decrease in accuracy among the six grade levels (229).

The investigator recommended that specific research is needed in the area of male vocal modeling in order to establish guidelines for male teachers in the elementary music classroom. It was also recommended that studies involving male normal range as opposed to male falsetto and different ranges of male voices were indicated (230).

VOCAL PITCH ACCURACY ASSESSMENT

Brophy (1997) investigated if children's singing games were a valid and reliable for use in authentic assessment of vocal pitch accuracy on the scale degrees "sol-mi" and "do-re-mi." Subjects for the study were first-grade, second-grade, and third-grade students (N = 236) who attended a public elementary school. Primary classes at the school combined either first-grade and second-grade or second-grade and third-grade students together as separate class units. The students attended music classes in those combinations. Subjects in the study were in six grade one/ two combination classes and six grade two/ three combination classes (61).

As part of their regular instruction, students at the school were trained to sing the pentatonic scale in the following sequence: "sol," "mi," "la," "do," "re." Children in the grade one/ two-combination classes experienced song materials that concentrated on (but were not limited to) "sol," "mi," and "la" in order to develop their vocal pitch accuracy on those pitches. Students in the grade two/three combination classes experienced expanded song repertoire that included "do," "re," and "mi," and concentrated on, but were not limited to, those pitches until the entire pentatonic scale could be sung accurately. The song repertoire experienced in those grades included numerous singing games, during which there were alternating sections of group and solo singing in a playful context (62).

Two singing games were chosen to test vocal pitch accuracy of "sol" and "mi" in the grade one/two combination classes, and two singing games were chosen to test vocal pitch accuracy of "do," "re," and "mi" in the grade two/three combination classes. The games were selected because they required the child to sing a solo on the pitches being assessed as part of the playing of the game. The "sol-mi" games were *Doggie, Doggie, Where's Your Bone?* and *Aunt Dinah. Aunt Dinah* was a call and response game, with the lead child singing first echoed by the class; all solos in the game contained only the pitches "sol-mi." The *Doggie, Doggie, Where's Your Bone?* game was a question and answer game, with the lead child in the middle of a circle answering the questions sung by the class in the circle. The first solo incorporated the pitches "sol-mi"; however the second solo included the note "la" as part of its melody. The phrase pitches of the solo were "sol," "mi," "la," "sol," "mi"; the "la" was not assessed for pitch accuracy (62).

The "do-re-mi" games were *The Closet Key* and *Charlie Over the Ocean.* The *Closet Key* game involved only the pitches "do-re-mi," sung first by the group and echoed as a solo by the child who had "found" the key. *Charlie Over the Ocean* was an echo game, with the lead child singing first, echoed by the class. The majority of *Charlie Over the Ocean* remained strictly on "do-re-mi"; however, the second phrase incorporated low *sol* one time. The phrase was "do-do-do-do-do"-low "sol"; the low "sol" was not assessed. Because of the authenticity of the singing material used in the assessments, the presence of additional pitches other than those being assessed was assumed to be a contextual necessity by the investigator and the melodies were not artificially altered to eliminate them (62).

Prior to the main study being undertaken, a pilot study was completed and reliability among four judges was assessed to be $r = .73$. Although slightly low for reliability purposes, this was determined to be acceptable by the investigator given the authentic, nontechnological nature of the measurement being used for vocal pitch accuracy (62).

The students attended music class at their normally scheduled time and were taught the games by their teacher as part of a regular music lesson. When the assessment began, the students were told that their teacher would be listening for how well they sang the identified pitches ("so-mi" or "do-re-mi") as they played the game. When the group sang, the teacher played a simple harmonic accompaniment at the piano that was consistent for each repetition. The solos were unaccompanied. Tonality was consistent between songs, with *Dinah* and *Doggie* performed in D major (the target pitches occurring on A4 and F#4), and *Charlie* and *Closet Key* performed in F major (the target pitches occurring on F4, G4, and A4) (63).

During each solo testing, a "+" or "-" was marked by the appropriate child's name on the class list to indicate whether or not the solo was sung accurately. To be determined accurate, the singer had to produce "accurate intervals and be at the correct pitch level" (63). Text errors, melodic errors in the song that preserved the pitches being tested, or stopping and starting over were not considered as factors in this determination (63).

Each set of songs was treated as two forms of the same criterion-referenced test. *Aunt Dinah* and *Doggie, Doggie* were considered to be alternate forms of a test for "sol-mi" accuracy; *Charlie Over the Ocean* and *The Closet Key* were considered to be alternate versions of a test for *do-re-mi* accuracy. Subjects were classified as follows: (a) consistently accurate—those who obtained a "+" on both songs; (b) consistently inaccurate—those who received a "-" on both songs; and (c) inconsistent—subjects who obtained a "+" on one song and a "-" on the other song. Data was tabulated for consistency of pitch accuracy between songs (listed as accurate, inaccurate, or inconsistent) and accuracy by individual song (63).

As a measure of reliability, Subkoviak's classification statistic was calculated to determine the consistency with which the sets of two songs tested the subjects' accuracy in singing the target pitches. A coefficient kappa (k) was then determined to correct for chance classification and to determine the consistency of classification based on the test forms. The calculations were completed by grade level and by class groupings (64).

Results indicated that of the two forms of the "sol-mi" test, classification consistency (both Subkoviak's po and k) was higher for second grade subjects ($po = .81$ and $k = .49$,

respectively) than for first grade subjects (*po* = .77 and *k* = .44). Of the two "do-re-mi" tests, classification consistency was higher for third grade subjects (*po* = .82 and *k* = .46) than for second grade subjects (*po* = .75 and *k* = .46). It appeared to the investigator that those songs, when used as tests of vocal pitch accuracy, provided more consistent results with older subjects (65).

The investigator reported that percentages of accuracy determination by individual songs also increased with each grade level. Of the "sol-mi" tests, a higher percentage of second-grade subjects were determined accurate than first-grade subjects on both songs. Of the two "sol-mi" tests, the *Doggie, Doggie* game was performed accurately by a greater percentage of subjects at each grade level as well as by the half class groupings. Of the "do-re-mi" tests, a higher percentage of second-grade subjects were determined accurate on *The Closet Key* than on *Charlie Over the Ocean*, whereas the opposite was true of the third-grade subjects. A greater percentage of the two-thirds classes sang the *Charlie* game more accurately than *The Closet Key*. Overall, the older subjects were classified as accurate on the songs more often than the younger subjects within each class grouping (65).

The investigator concluded from the study the following:

1. Authentic children's singing games can be used as a measure of vocal pitch accuracy in elementary children.
2. The children enjoyed the games, and appeared to sing alone easily and confidently within a play environment.
3. The teacher must be very quick to judge the accuracy of individual singers due to the relatively short amount of time during which the solo is sung (65).
4. The sets of songs used in the study to test for "sol-mi" and "do-re-mi" intonation appeared to be consistent enough in their classification of the targeted pitches to be reliable for use by classroom teachers.
5. At each grade level, the *po* results were high enough (*po* = .75 to

po = .82) to indicate that their use by practicing teachers could help reliably identify students needing more instruction in vocal pitch accuracy.

6. The opportunity to hear individual children during the games provided a comfortable atmosphere for the teacher to provide brief one-on-one instruction to students who were not matching pitch.
7. Further interpretation of the data by the investigator revealed that the lower *k* results indicated that the two songs did not test the pitches very consistently on the basis of the songs themselves, and that those sets of songs may not have been alternate forms of the same test (66).

Several factors were deduced by the investigator to account for the inconsistency in test results between the sets of songs:

1. Since the games were authentic ones that sometimes included pitches other than those targeted, the additional pitches may have had an effect on the vocal accuracy of students that was unaccounted for by the study.
2. Variations in individual subject familiarity with the song and the accompanying game may have impacted upon the observed results.
3. Various movements inherent in the game may have resulted in a kinesthetic impediment to accurate singing.
4. Variations in the elements of the songs, such as length, melodic contour, rhythm, number of pitches in the solo, and vocal range could have affected pitch accuracy.
5. Individual personality differences in subjects may have caused some students to sing less accurately within the class environment.
6. Since the assessments were made on separate days, daily influences in the subjects' lives, such as rest, nutrition, outside school pressures, and classroom activities engaged in prior to the music lesson may have influenced individual accuracy.
7. The teacher's accuracy judgments may not have always been correct. This could have been the result of the teacher's individual level of ability at pitch accuracy judgment, and normal intrusions in the classroom environment during the test administration (66).

Content validity of the study was established by the presence of the targeted pitches in the children's solos. The reportedly high consistency levels obtained between songs were interpreted as a measure of concurrent validity. The games were considered to be valid measures of the target pitches (66).

The differences between the observed consistency classification (*po*) and coefficient kappa (*k*) calculations were indications to the investigator that the songs should not be

used as sole measures of vocal pitch accuracy on the targeted pitches. They suggested that the assessment results could be useful in a larger, multiply-contexted assessment such as a portfolio or developmental profile (67).

The investigator made several recommendations for future research:

1. Other authentic songs should be similarly tested that do not employ pitches other than those being targeted for assessment.
2. Subsequent studies should involve more practicing teachers—rather than limited to one teacher's use of the games.
3. Other methods of recording observations such as unobtrusive tape recording for later assessment or hand-operated microcomputers should be examined.
4. A longitudinal study over a period of one school year should be conducted where the singing games are assessed at specified times to track student progress with respect to vocal pitch accuracy (67).

VOCAL RANGE AND DEVELOPMENT

Wilson (1970) investigated whether significant individual differences exist in the vocal range and development of children's voices at all grade levels from first through sixth (6). Subjects for the longitudinal study were sixty-nine students, thirty-seven girls and thirty-two boys, who entered first grade in Eugene, Oregon, during the school year 1957-1958. Extreme cases of physical or emotional deviations from the normal were excluded from the group. Administrators of the school district believed that the children represented families of the upper middle class socio-economic level with no extremes of economic background (9).

Results of a testing program conducted in the school district during the school year 1958-1959 indicated that the group rated above average in IQ. Identification of gifted children was based upon the following factors: (1) IQ above 120; (2) high achievement record in the classroom; and (3) teacher recommendation. Identification of the slow

learners was based upon: (1) IQ below 80; (2) poor achievement record in the classroom; and (3) teacher recommendation. The percentage of gifted identified at Fairfield school was approximately ten percent above national norms. The percentage of children selected from the second grade for those two classifications was the same as the whole population (9).

The study continued with the same group from first grade through sixth grade with a follow up test in the twelfth grade. Subjects available for testing each year were: second grade, thirty boys and thirty girls; third grade, twenty-five boys and twenty-three girls; fourth grade, twenty-three boys and twenty-one girls; fifth grade, twenty-two boys and twenty girls; sixth grade, sixteen boys and eighteen girls; and twelfth grade, thirteen boys and thirteen girls (9).

Students in the Aymer Jay Hamilton Laboratory School (AJH), Chico State College, Chico, California, were additional subjects. All students from first grade through sixth grade who attended the school from 1964 to 1969 were included in the study. The total number of students tested at each grade level were: first grade, sixty-three boys and sixty-two girls; second grade, sixty boys and fifty-four girls; third grade, fifty-six boys and forty-nine girls; fourth grade, fifty-six boys and fifty-six girls; fifth grade, sixty-one boys and sixty-three girls; and sixth grade, forty-nine boys and sixty-eight girls. Some of the subjects were tested at different grade levels for five years. Other students remained in the school for varying periods of time (10).

The student population at AJH was described as similar to that of Fairfield that due to the heavy weighting of college faculty children, a large proportion of the children had cultural advantages such as travel, opportunities to hear concerts, and opportunities for

private instruction in music. There were no extreme cases of students with emotional problems reported. The two groups were similar in IQ samples, and so it was assumed that the two groups were similar (13).

Voice tests were administered in September of each year during 1957 to 1962 for the subjects at Fairfield School and 1965 to 1969 at AJH. Each year before beginning the testing program, the tester conducted music classes for two or three weeks to establish rapport with the children. Children in the longitudinal study and those at AJH who were enrolled in the school for several years during the testing program had an advantage in being familiar with the tester and the testing situation. This variability factor could not be avoided but was considered in the interpretation of the data (14).

Each child was tested individually in a sound-proof room. The tester informed the child that the purpose of the test was to keep a record of his voice so that he could see how much his voice had grown each year. The children were assured that the test was not connected with any kind of grading process. The children at Fairfield School were told that they were an experimental group and that the information was to be used for a research study. It was thought that such information may produce a "Halo effect," and so the students at AJH were not given this information. All students at AJH were tested for various musical skills each year and the vocal range test was an incidental part of the testing program. No group was singled out for special attention (15).

In testing for vocal range, the tester started with a pitch selected by the child. The child was then asked to match pitches in sequence from this point in an ascending and descending chromatic scale with the aid of the piano. When necessary, the tester sang the pitch to assist the child. The syllable "oo" was used for the ascending scale and "ah" was

used for the descending scale. Those syllables were selected by the investigator because they appeared to be the easiest for the children to sing (15).

The entire procedure was approached as a game to see how high and low they could sing. As an additional aid to the child for pitch direction and self confidence, the tester held the child's hand and moved it up and down with the direction of the pitch. If a child became tense during the testing process, the tester took time out to get the child in a relaxed mood before resuming the test (16).

In October of each year, a tape recording was made of the longitudinal group singing a song. The song *Good Morning* was used from first grade through the fourth grade. Because of the maturity of the students, the chorus of *O Susannah* was substituted in the fifth and sixth grades. On one day, each child sang the song at a pitch of his own selection. The following day, he was asked to sing it at a starting pitch of C5, the pitch printed in the songbook from which it came. The starting pitch was played on a standard pitch pipe (16).

A recording of each child was made in the classroom. New students each year were also recorded so that they would not feel left out. Those students provided additional data for pitch selection of children (17).

Because of the necessity of scheduling the music teacher's time, the tapes were made in the classroom. Since she was not the music teacher in 1962-1963, she was able to make the recordings for that particular year individually (17).

The investigator observed that each child used the starting pitch of the preceding child unless it was obviously too high or too low for him. The sixth grade tapes avoided this problem. At AJH it was possible to give the singing test individually; but because of the

number of students involved and difficulties encountered in the use of the tape recorder, tapes were not made of those subjects. In place of the tapes, the tester recorded the information directly on the test form. The subjects at AJH were asked to sing a song at a pitch selection that was comfortable for them. In the first grade they sang *The Farmer in the Dell* and in other grades *Row Row Row Your Boat*. Those songs were selected because they were well known to the students. If a child did not know either song, he was allowed to sing a familiar song of his own selection. The starting pitch used by the child and the accuracy of the melodic line were recorded (18).

The tester later observed that *The Farmer in the Dell* was a poor selection because it was difficult to discern whether the child had accurately sung the quick upbeat. For this reason, the investigator cautioned that the lower note recorded for pitch selection for the first grade should be interpreted with some reservations. It was said that the tessitura of the song was a fourth above the starting note and the pitch for that part of the song could be assessed with accuracy. It was observed that children with limited ranges usually skipped the upbeat or else sang it on the same note as the first strong beat of the song (18).

In the fall of 1969, an attempt was made to locate as many of the longitudinal subjects as possible. The perspective high school voice teachers individually tested each of the subjects located. For each student, they recorded voice range and voice classification. They judged the quality of the voice and indicated the presence or absence of any bad vocal habits. Each student was asked to sing *Row Row Your Boat* in a personally-selected key. The tester recorded the key selected and judged the accuracy of the melodic

line. Each student was asked to fill out a questionnaire regarding any choral experience since elementary school and attitude toward singing (18).

Information regarding the child's physical, emotional, and home background was recorded on a form prepared specifically for the study. The information was obtained from school record cards, health examination forms, interviews with the classroom teacher and interviews with each child (19).

Various statistical techniques were used to summarize the data. Graphs, tables of upper and lower voice limits and common pitches, and standard deviations were used to determine the extent of individual differences and rate of vocal development. Percentages were used for comparing pitches matched by the children, song ranges of the book and children's voluntary pitch selection. The significance of mean pitch and mean range span difference between boys and girls was determined by *t* tests. *Chi Square* was used to test differences in children's voice range and their adult voice classification (20).

The investigator made the following conclusions in regards to the study:

1. Marked individual differences in children's voice compass and span existed at all grade levels.
2. Children's voices developed at different rates at all grades.
3. Training did not decrease individual differences in voice ranges, but the importance of considering individual differences for unison singing decreased as pitches common to most children increased.
4. The range of pitches that children considered comfortable for singing was considerably lower than the range of E4 to E5 that was traditionally recommended by music educators and considerably lower than the average range of songs in the music series examined (164).
5. The average range of songs in current music series examined was considerably higher than the range that the majority of the children matched at all grades.
6. The average voice range of boys was slightly lower than the average voice range of girls at ages, six, seven, ten and eleven.
7. The percent of children who matched pitches in the low range was higher for boys than for girls.

8. The percent of children matching pitches in the high range was higher for girls than for boys.
9. Marked differences in tessitura of children's voices existed both between and within genders.
10. The range span of boys and girls was not significantly different, except at age eleven preceding puberty when some boys' range span became narrower.
11. The tessitura of the child voice of adult altos and basses may be lower than that of the child voice of sopranos and tenors.
12. Traditional methods and procedures for singing with children were in direct conflict with research on the child voice (not specified).
13. Music educators in general had ignored research on the child voice.
14. Singing below the range of E4 to E5 does not harm the child voice if voices are not forced (165).
15. Providing for individual differences and adjusting the pitch range to the children for classroom singing improved attitudes toward singing and elminated non-singers. The improvement was particularly noticeable with boys.
16. Self-confidence affected vocal range.
17. Differences in musical training increased individual differences in vocal range.
18. Elementary music teachers need to be informed more about vocal techniques as applied to the child voice and the psychology of learning.
19. Elementary music teachers need to know the vocal range of each child and need to be able to transpose and arrange music in order to provide successful singing experiences for all children (166).

The investigator recommended the following in regards to future research:

1. Determine the extent of correlation between self-concept and vocal range.
2. Determine the variables that affect the development of vocal range of children's voices.
3. Measure the difference in tessitura of children's voices.
4. Measure the correlation between tessitura of the child voice and his later adult voice classification.
5. Test the experimental method for singing in a controlled experimental design (166).
6. Examine the comparative tessitura of children's voices in families.
7. Provide more longitudinal data on the development of children's voices (167).

Levinowitz, Barnes, Guerrini, Clement, D'April, and Morey (1998) investigated:

1. The reliability of Rutkowski's original SVDM, *Singing Voice Development Measure*, to measure students' use of the singing voice in grades one through six.
2. Whether the use of the singing voice is developmental in grades one through six.
3. The dependability of children's use of their singing voice when singing a song in a major mode and when singing a song in a minor mode.
4. The expectation for the use of the singing voice in students from grades one through six (38).

One-hundred-seventy students from five elementary schools in southern New Jersey participated in the study. The students were said to represent a diverse ethnic, intellectual, and socioeconomic population. Three of the schools were considered suburban, one school was designated rural suburban, and one school was urban (38).

The study was designed and implemented by part-time graduate students who were also full time elementary general music teachers. The design and implementation of the study was under the direction of the faculty member who taught *Choral Procedures* class at Rowan University (38).

Prior to the study, all of the graduate students completed reading assignments from Kenneth Phillips's book *Teaching Kids to Sing* (1996) and from research literature regarding the development of the child's singing voice and the use of the singing voice during the elementary school years (38).

Five of the six graduate student coinvestigators were responsible for audiotaping the students who would consitute the sample for the investigation. Each of the five coinvestigators were assigned the following: (a) students in at least two grade levels between grades one and six, and (b) approximately forty students from those designated grade levels to be audiotaped. The procedure was designed to create the sample because

the job descriptions of the five coinvestigators were said to be extremely divergent as to the grade levels that they instructed (38). Therefore, the five coinvestigators taped approximately forty students each. To enhance the generalizability of the results, students were sampled from at least two schools for each of the grade levels from first through sixth (39).

One month prior to the audiotaping of the students, each of the five coinvestigators introduced as new songs the two criterion songs in similar rote-song procedures to all their students in their respective schools. The criterion songs were sung for four class sessions in all music classes (39).

One of the criterion songs, *In the Sea,* was in minor tonality; the other song, *Row, Row, Row Your Boat*, from traditional folksong literature, was in major tonality. The primary consideration for both songs, when sung in the designated criterion key of D, was that they had to encompass a range of one octave from the D4 to the D5 and therefore contained notes above the "lift" or register break. A secondary consideration for the choice of the two criterion songs was finding a song that was short in length and a song that could be longer in length, enabling the students to audiate at a deeper level during their performance (39).

All audiotaping by the coinvestigators began during the same week. Singing each of the two criterion songs, each student was tape recorded individually by his or her respective music teacher during a regular general music class (39).

The order of the songs was randomly determined by each of the coinvestigators. For each student audiotaped, the teacher set the tonality of the song by playing a I-V-I

progression on the guitar or piano. Then, the teachers sang "Ready? Sing" on the starting pitch and took a preparatory breath to direct the student to sing (39).

The tape-recorded performances were independently rated by the six coinvestigators using the original SVDM. The performances were attended to and judged in random order to avoid a grade level bias (39).

Results of the interjudge reliability coefficients for grade, major song, minor song, and both songs ranged from $r = .43$ to $r = .98$ (41). It was reported that the reliabilities for both criterion songs from the performances of children in grade six were lower than the reliabilities that were calculated from the performances of the younger children. The investigators hypothesized that it could have been that the use of the singing voice of students in sixth grade could not be appropriately measured because of their preference for chest voice singing. The investigators concluded that therefore, it seemed that a measurement tool other than the SVDM should be used with students in grade six and above (40).

The researchers failed to find a statistically significant difference in the use of children's singing voices in grades one through six. The investigators said that many children in elementary school grades vacillated between the speaking voice and the singing voice. They concluded that it seemed that a child's use of his or her singing voice did not develop due to maturation as did the range of his or her voice. It was reported that the data confirmed the notion that the use of one's singing voice is a learned, complex skill (41).

A statistically significant main effect for song across all six grade levels was found in favor of the major song. The investigators noted that the minor song was longer in length and could have contributed to this finding (42).

It was found by the investigators that 16.60 percent of the students could use their singing voice with some facility below the register break. It was also observed that students who fell within the "presinger," "speaking range singer," and "uncertain singer" categories were between seventy-five to ninety percent of the population. The investigators reported that students in grades five and six tended to use their speaking voice range although research indicated that at this level the child's voice has reached its peak developmentally. The data indicated that the older students preferred to sing in the lower adjustment/chest voice to emulate role models in the media (43). The investigators hypothesized that the fifth and sixth graders may have felt self-conscious singing in front of their peers. Also noted were a number of changing male voices which were hypothesized as causing the frequent chest voice register use by the older students. The numbers of these were not reported (44).

VOCAL REGISTRATION

One study in which the interaction of vocal register and range has been taken into consideration is that by Brown (1988), who investigated the effects of self-selected pitch and prescribed pitch with a model on the vocal range of children in grades one, three, and five. In the first assessment, subjects were asked to sing a familiar song at a pitch level of their own choosing. Range was then measured, ascending and descending, from the final pitch of the song. In the second assessment, the investigator provided a vocal model in

the head register, and guided subjects with vocal activities (speech to song) that helped them to locate their "upper" voices. The subjects then repeated the familiar song, beginning on the pitch B4 above middle C4. Range was measured as before, beginning this time on the pitch G4 above middle C4, the last note of the song. Results of the analysis of the data for vocal range demonstrated a significantly wider range for singing in the higher vocal register. As a group, subjects, when permitted to choose their own singing level, chose a lower range and register for singing. When exposed to a higher vocal register, subjects were able to make the transition quickly to this "new" voice, resulting in a wider vocal range. Had the subjects not been instructed in the use of the head register, the results of the vocal range assessment would have been quite different, that is, lower. Brown recommended that further research as to vocal range should make provision for the registration factor as it affects optimum singing range (53).

Children sang individually in small groups. The tester started the children singing at a note which they could reproduce easily and then worked up and down from that pitch. The tester first played each note on the piano, then sang the tone, and asked the subjects to reproduce it. If subjects appeared to have difficulty at their range extremes, the note was attempted up to eight times. If they were still unsuccessful, they were presented notes which were quite beyond the immediate note (54).

It was reported by the investigator that mean and median numbers of tones sung by boys and girls at yearly and half-yearly age levels were computed using the scores obtained in the last of the three tests. Mean scores reportedly increased with each year of age, with a slight decrease at age ten. Median scores also increased with age but less gradually than mean scores. For example, the median range for two-year-olds was five

tones (D4-A4), for four-year-olds was nine tones (B3-C5), for six-year-olds was fourteen tones (A3-G5), and for ten-year-olds was sixteen tones (G3-G5). It was reported that a wide variety of ranges was found at all levels. A wide spread between the twenty-fifth and seventy-fifth percentiles suggested to the investigator that neither the average nor the median score represented typical subjects (71).

Data tapes were replayed for analysis on the device with which they were recorded. This was connected by means of a patchcord to a *Kay Visi-Pitch Model 6087*. The *Visi-Pitch* displayed frequency and intensity values on an oscilloscope screen in real time (56).

The investigator made the following conclusions based on the study:

1. When beginning vocalization in the "loft" or head voice register with a vocal model, subjects sang with significantly wider range spans (4.68 semitones) than when they began vocalization at their own chosen pitch level.
2. When beginning vocalization in the loft register with a vocal model, subjects had significantly higher upper range extremes (three semitones) than when they began vocalization at their own chosen pitch level.
3. When beginning vocalization in the loft register with a vocal model, subjects had significantly higher mean pitches of singing range (two semitones) than when they began vocalization at their own chosen pitch level.
4. When beginning vocalization in the loft register with a vocal model, subjects showed less individual difference in span and mean pitch of singing range than when they began vocalization at their own chosen pitch level.
5. There were no significant differences in span of range between genders with or without vocal model.
6. Only when subjects began vocalization at their own chosen pitch level did girls have significantly higher mean pitches of singing range than did boys (80).
7. There was no significant difference in the lower range extremes of boys and girls, with or without a vocal model.
8. In both spontaneous speech and singing, girls' voices were significantly higher than boys'.
9. In speaking, girls used a wider range of pitches than did boys.

10. Only when beginning vocalization in the loft register with a vocal model, did subjects in higher grades have significantly wider range spans and produce significantly higher upper range extremes than did subjects in lower grades.
11. There were no significant differences between grade levels in subjects' mean pitches of singing range.
12. There were no significant differences between grade levels in subjects' lower range extremes when singing with or without a vocal model.
13. The higher the grade level, the lower the modal pitch of the speaking voice.
14. The higher the grade level, the narrower the span of pitches used in speech.
15. When choosing their own beginning pitch, the majority of subjects sang at pitch levels near middle C4.
16. After singing with a vocal model and doing vocal exercises, the majority of subjects sang within a semitone of B4 (a seventh above middle C4).
17. When asked to pick their own beginning note, most subjects began singing within a distance of four semitones above the pitch of their speaking voice (81).
18. The correlation between the subjects' pitch of the speaking voice and chosen beginning pitch of singing was positive but only moderate.
19. Subjects with lower speaking voices did not have narrower singing ranges (82).

Cleall (1970) investigated the singing ranges of more than 1,200 subjects ages four years to adult in Surrey County, England. The procedure involved encouraging each subject to sing "as long as it seemed ... that there were more notes to come" (107). Range was determined by the singing of scales. Interpretation of his findings was difficult because nothing more was reported of his methodology. For example, although he mentioned his criterion, it was not clear whether pitch-matching was part of the task. Cleall reported that he accepted notes which were "actually and incontrovertibly pitched, no matter what their quality" (107). Cleall mentioned that only vocal range was being tested and not quality of singing (108).

Cleall reported median ranges divided by gender and academic ability groupings as follows:

1. Boys ages five to eight years: G#3-G#4
2. Girls ages five to eight years: A#3-B4
3. Boys, top group ages eight to eleven: F#3-D5
4. Girls, top group ages eight to eleven: F3-E5
5. Boys, lower group ages eight to eleven: F#3-B4
6. Girls, lower group ages eight to eleven: F#3-D5
7. Boys, top group ages eleven plus: E3-F#5
8. Boys, lower group ages eleven to voice change: F3-D5
9. Girls, both groups ages eleven to eighteen: E3-E5 (108)

Cleall reported differences in range between genders and between academic ability groupings of the same gender. He did not report any analyses of variance. Forty-six of the subjects were retested (eight to eleven years) two-and-a-half years later and it was reported that no pattern was found in how subjects' ranges had changed. No correlation coefficient between the first and second scores was reported. Later (1979) on the basis of those findings, Cleall concluded that the children did not possess the vocal ranges included in the current songbooks of the time and vocal ranges of the individual subjects differed greatly (1).

Remedial Techniques Used with Inaccurate Singers

Various remedial strategies to teach singing have been tested and identified. Following is a review of literature of some extant and relevant research related to remedial techniques used with inaccurate singers.

ADDITIONAL DRILLS AND EXERCISES AND SONGS

Joyner (1969) conducted a study to investigate what he called the "monotone problem" with maximum speed and effectiveness. In the study he examined group and individual characteristics of monotones which he hoped would help to determine the basic nature of their deficiencies and indicate the type of remedial training methods that should be used (115).

Joyner defined the monotone singer as someone who "consistently failed to produce the tonal configuration of a melody in a recognizable manner" (115). Thirty-two teacher-assessed monotones from a boys' grammar school were given a detailed singing examination designed to test for monotonism and subsequently form an experimental group of monotones. The average age of the subjects was twelve years of age (115).

The main singing test was the performance of the *British National Anthem*, initially in its usual key of G major, and then in a series of stepwise downward transpositions to G major, one octave lower. Each performance received a grading of "A," "B," or "C" according to the following rubrics:

A = No major pitch errors or departures from the melodic outline.
B = Generally erratic in pitch, in spite of moments of tunefulness.
C = No hint whatsoever of melodic outline or pitch coincidence (115).

Grades were given for the performances in G major. Six subjects received a grade of "A." Eight subjects received a grade of "B." Eighteen subjects received a grade of "C." It was reported that a downward transposition of the tune to the key of Bb resulted in a greater number of tuneful performances. Fifteen subjects scored a grade of "A." Nine subjects received a grade of "B." Eight subjects received a grade of "C." Certain subjects achieved the same grading at either pitch, their performances remaining

unaffected by transposition. Six subjects remained at the "A" grade. Two subjects remained at the "B" grade. Eight subjects remained at the "C" grade. The performances of sixteen subjects were affected by the transposition from G to Bb major. Four subjects went from "C" to "A." Six subjects went from "B" to "A." Six subjects went from "C" to "B." Thus, it was reported that the original thirty-two subjects fell naturally into four divisions on the basis of their different levels of tuneful singing. The four divisions were the following:

1. Normal Singers: Able to sing in tune at both pitches.
2. Grade A Monotones: Tuneful at low pitches, untuneful at usual one.
3. Grade B Monotones: Erratic at both pitches, slightly better at low one.
4. Grade C Monotones: Always completely untuneful (117).

There were six subjects who had shown individually that they were capable of giving an accurate account of a melodic outline sung at the usual pitch. The investigator said those subjects demonstrated the following characteristics in the classroom:

1. They experienced periodic moments of singing out of tune.
2. The general quality of voice production was reported as being rather poor, their tone being thin and weak and having little resonance or flexibility.
3. Difficulty was reported in negotiating the transition into the head register when ascending the scale, and singing in this register was difficult and unfocused (117).

It was hypothesized by the investigator that it was possible that the six subjects were unable to hear themselves singing out of tune in a large group of singers and thus self-correction was not possible. It was observed by the investigator that those subjects generally maintained tuneful singing most consistently where the music was soft and low in pitch rather than loud and high (117).

The investigator eventually decided that the experimental term "real monotone" should apply only to those who sang consistently out of tune, regardless of the pitch level

employed. That definition excluded the ten subjects who were regarded neither as normal singers nor real monotones, but treated as a separate experimental group and labeled grade "A" monotones. The final group of real monotones consisted of the sixteen subjects found under the headings of grade "B" and "C" monotones (117). The purpose behind the different monotone classifications was to isolate an experimental group of real monotones and compare its performances in various facets of singing ability with those of control groups of normal singers (117).

The investigator postulated that in order to sing in tune a person must be able to do at least three things: (a) to tell one pitch from another, thereby gaining a correct mental concept of the rise and fall of a melodic outline; (b) to recall successions of pitches organized into melodic patterns in order to be aware of what comes next; and (c) to have a vocal instrument capable of reproducing the succession of pitches in a melody; and (d) that instrument must be able to make an immediate and accurate response to his intentions (118).

All thirty-two subjects took the *Bentley Pitch Discrimination* and *Tonal Memory* tests. Their raw scores suggested to the investigator that those less capable of singing in tune also possessed poorer pitch discrimination. A statistical analysis of the significance of the differences between the various means gave some support to this argument. The method used was the critical t ratio ($p < .05$) (118).

Some indication was reported that in the raw scores the more able singers possessed the better musical memories; however, the distributions of the scores was skewed on the high side, and calculations revealed no significant differences between any two means (118).

Each of the thirty-two subjects was tested in a song of his own choice and various ascending and descending scales. Notes were made of tone quality, comfortable range, and ease of production (118).

It was reported that the testing revealed a general poverty of voice production among the subjects. The worst performers were the grades "B" and "C" monotones, six of whom had drastic breaks in the voice around G4 to A4. In general, the grade "A" monotones experienced similar problems but to a lesser extent. Their voices, unlike those of the grades "B" and "C" monotones, did not tend to collapse near the above pitches; they did, however, continue further up the scale only through great physical effort, producing little tone of a very strained quality. The production of notes in this register was far from being natural, spontaneous, and easy (119).

The researcher reported that there was evidence that the six monotones who proved to be nearly normal singers suffered from a mild degree of the same vocal inflexibility. Those six sang with less ease than their average normal-singing contemporary at the same school, tending to make an uneasy transition into the head register and shying away from the high notes (119).

The voice production tests suggested a strong association between untuneful singing and voice production difficulties, and thus raised the consideration that monotones might be tone-deficient rather than, or as well as, tone deaf. It was reported by the investigator that the grade "A" monotones sang in tune where the melody lay within their comfortable range. He noted that, however, most of the grade "C" monotones had no comfortable range at all, experiencing difficulty in producing tone on any pitch. The grade "B" monotones were also reported as having very small comfortable ranges (119).

The investigator stated that the quality of vocal tone depends on the amplification and harmonic enrichment of the sound produced by the vocal cords through various resonators in the body. The investigator's explaination of vocal tone production was the following:

1. The sound beam originating at the larynx will reach the appropriate resonators only if it is of the correct width and height for each note sung.
2. Those dimensions are determined by the amount of vibrator mass or vertical thickness of the vocal cords employed. In a naturally good voice, the vocal cords adopt the correct vertical thickness for every pitch sung, thereby producing a sound beam of the correct dimensions for finding the most efficient resonator areas. Thus, a large and rich tone is produced with comparatively little effort.
3. In a poorly produced voice the vocal cords fail to reduce their vertical thickness with rising pitch, an over broad and foreshortened sound beam is produced which lacks the proper harmonic reinforcement and is of poor quality.
4. It is a tendency of those with voices thus limited to attempt to produce more tone through increased breath pressure and stronger support from neck and laryngeal muscles. This results in tone of even coarser quality produced with a great sense of strain, and makes it unlikely that the victim will experience the proper working of his singing voice.
5. A break occurs in a voice when the vocal cords retain excessive thickness in ascending the scale where the thinning process should have been operative.
6. Under extreme pressure, the cords suddenly lose their thickness, a split-second interruption of the sound beam takes place, and the tone snaps into a different resonating zone.
7. Until nature asserts itself in this way the thinning process has been minimal, indicating an extreme lack of flexibility in the vocal apparatus.
8. A break may be very pronounced and almost percussive in effect.
9. The physiological reason for the break is to be found in the slightly extra thinning which the cords have to undergo to make the transition from G4 to B4.
10. Resulting from this extra thinning is the production of a taller and narrower sound beam which naturally projects into the mouth, nasal passages, and sinuses—the so-called "head resonators." Thus singing difficulties tend to become critical at the point at which the above process should take place (120).

The investigator hypothesized why grades "B" and "C" monotones did not sing in tune. He said their vocal instruments were "manifestly unfit for normal usage, and generally uncoordinated as a unit." He added that in songs sung "at their normal pitch, the disconcerting effect of the break made their efforts useless." He noted that even where songs were transposed downward to avoid the break, the grade "B" and "C" monotones still lacked sufficient coordination of the chest voice to participate successfully. It was observed that all their efforts to sing resulted in physical discomfort and strain, and the probability that this, coupled with the ridicule of normal-singing contemporaries, had endowed some of them with a negative attitude toward singing (120).

The investigator noted that the voices of the grade "A" monotones were imperfect, but only critically so above the pitch where the head voice came into play. During the course of testing they were noticed to aim at high notes, but failed out of what was perceived to be purely technical reasons. In that way, the investigator noted that they contrasted with the "B" and "C" grades of monotones who generally showed little or no intention of aiming at the right note (121).

The next step was to compare the performances of monotones and normal singers in the basic psychological skills of melody recognition and reproduction, tonal memory, and pitch discrimination. The test used was the *Bentley Measures of Musical Abilities*. The investigator's use of the test involved a three-fold comparison, since a separate group of grade "A" monotones was made (121).

One testing sample included 134 normal singers, twenty-six real monotones, and twenty grade *A* monotones. The average age was M = 11.8 years and the age range

was 11.2 to 12.6 years. Differences between the means of the normal singers and real monotones were significant at the $p < .01$ level for both pitch discrimination and tonal memory. Differences between the means of the grade *A* monotones and the real monotones were also significant at the $p < .01$ level for both tests (121).

The investigator stated the following results:

1. Monotones revealed characteristic deficiencies in both pitch discrimination and tonal memory when compared with normal singers.
2. Grade "A" monotones associated themselves in terms of those abilities with normal singers rather than real monotones.
3. The partial monotonism demonstrated by the grade "A" monotones was a direct product of their inability to "produce the voice normally."
4. Grade "A" monotones clearly differed from the real monotones who were deficient in pitch discrimination, tonal memory, and voice production (121).

The investigator suggested the following theories to define the real monotones and their singing abilities:

1. Physical sensations aid the comparison of two different pitches.
2. Real monotones have an unusually inflexible laryngeal mechanism, or lack the power of making normal use of it.
3. The child who, early in life, lacked the ability to make any definite and physical response to melodic stimuli, also lacks, as a consequence, a reliable method of checking or reinforcing the correctness of his initial attempts to sing in tune.
4. A child who starts with comparative deficiencies in pitch discrimination and tonal memory may be little inclined or even unable to exercise his voice over a wide pitch range. His vocal apparatus then remains undeveloped with respect to singing due to lack of practice. This, in turn, might further inhibit his mental processes (122).

As part of the research project, remedial training was given to eight real monotones. Methods included trial and error procedure, especially in the early stages. A definite attempt was made to investigate the findings concerning the relative deficiencies of monotones in pitch discrimination, tonal memory, and voice production. Sufficient

common factors arose from case to case to permit the selection of one subject, described below by the investigator, which was regarded as typical (122).

The typical subject was a primary school boy, eleven years and two months of age, who had scored only twelve out of twenty on pitch discrimination and four out of ten on tonal memory in the Bentley tests. His training consisted of a daily twenty-minute session given four times a week for a total of fifteen weeks (122).

Initial sessions were spent in trying to teach the subject how to pitch a single note—middle C4 on the piano—consistently in tune, with the aim of using this as a reference note for learning intervals and tunes. After about twenty sessions, it was reported that the subject was still finding the task beyond him. Great difficulty was reported in producing the note, and even when this was accomplished an immediate repetition was reportedly liable to be inaccurate. When he eventually produced the correct note, the investigator reported that the subject appeared to be unaware of the fact, and that the reinforcement of correct responses had no effect. The subject was described as having a very small span of attention (122).

During the training it was noticed that the subject's responses were "of a curiously brittle and ill-defined nature, characterized by considerable physical exertion but little confidence or sense of purpose" (123). It was reported that his attacks lacked focus and did not appear to be under his direct control. It was decided to attempt improvement in the subject's voice production and to examine what effect it had on his ability to recall musical material. Following were some of the main exercises employed:

1. Open his mouth as wide as possible and rest his tongue unobtrusively on the floor of his mouth, checking this with a mirror, thus creating resonating space and opening the throat.
2. Utilize the lower part of the lungs to conserve breath and to oppose the outward flow through the extension of the rib cage, thus providing the natural support for the voice and encouraging jaw, neck, and throat muscles to relax.
3. Consciously direct the sound forward to the head resonators by raising the upper lip and singing a very sharp "ee" sound while holding the nose. The experimenter furnished constant examples of undesirable sounds followed by improved ones, which the subject was asked to identify and criticize (123).

At first progress was reported as being slow due to the subject's self-consciousness. The investigator reported that in the fifth voice training session a breakthrough occurred when the subject suddenly produced a single sound "of much improved quality" (123). Immediate reinforcement was given, and it was reported that the subject easily produced further repetitions at the next training session. A new attempt was made to establish middle C4 as a reference tone, with stress laid on the ease of vocal attack. It was reported that after intensive practice the subject began to show new powers of focusing on that note and repeating it at will. Practice was then given in singing the interval C4-D4 on the sound "ah." The subject was told to sustain the reference note for as long as possible, supporting it strongly with the breath. He was then asked to begin the same exercise, this time making a glissando up to D4. The investigator said that the glissando helped the initial quality to be carried up to the new note (123).

The subject learned to sing up and down the following scale: C4, D4, E4, F4, G4, and to imitate short tunes based on consecutive notes. A normal legato replaced the glissando and the subject's vocal quality, although still comparatively poor, was reported as being smoother and fuller. The subject reportedly possessed a new sense of confidence,

purpose, and direction, and had been brought a stager nearer to becoming a normal singer (123).

The investigator concluded the following:

1. Parallel development took place between the ability to make definite, easy, and accurate vocal responses to musical material and the ability to recall the same material.
2. Direct training in pitch discrimination and tonal memory was less effective than the training of the vocal instrument (124).

Among seventy-two real monotones examined by the investigator, continuity was found to exist from case to case. However, it was reported that four subjects bore no apparent relation to the majority, and no workable system of remedial training was successful. All four were reported as being problem children at their schools because of their general incapacity for any kind of school work. Two of them had difficulty in speaking clearly, their consonants being slurred. It was concluded that their general ability to form mental concepts was very low, and that their appreciation of pitch change and tonal configurations was so much lower than that of the average normal singer or monotone as to mark them as subnormal (124).

The investigator reported that training sessions held with grade "A" monotones indicated that their partial monotonism was a direct product of faulty voice production. It was reported that each of the five subjects was always able to recognize when he was not singing in tune, and further testing revealed that each possessed an adequate ability to recall tonal configurations. Training was directed to the extension of the subjects' comfortable ranges. The exercises employed were similar to those used with real monotones with certain additions, which included glissandos over an upward leap of an

octave. The average gain in operative range after thirty forty-minute sessions was reported as being between seven and nine semitones (124).

The investigator made the following recommendations:

1. No teacher has sufficient time to administer remedial training to ten-year-olds and eleven-year-olds whose responses are resistant to change. Preventative action should thus be taken in grade school (ages five to seven) to ensure that bad vocal habits are not allowed to persist.
2. Potential monotones can be helped through early training in the easy command of the vocal instrument over a gradually-extended pitch range (124).

The Roberts and Davies (1975) experiment was to see whether the musical performance of monotones could be improved by a program of remedial training. A survey of poor-pitch singing carried out in the Chester area collected information on 745 boys and 226 girls rated by their teachers as monotones or droners. Monotone singers were defined as being "Always completely untuneful with little variation in pitch" (228). Monotone boys amounted to approximately eleven percent of the male sample and monotone girls to approximately three percent of the total female sample. From the 971 children, a sample was drawn of ninety children six-plus to eight-plus years of age who attended five schools. In the sample there were forty-nine boys and forty-one girls. In addition, a sample of thirty normal singing children ages six-plus to eight-plus years of age was taken from another school, which was said to be comparable to the other schools in terms of the population served by the investigator (228).

The ninety monotones were assigned at random to three groups of thirty children each. The normal singing children made a fourth group. The control group had no special treatment except for the usual music lessons at the school. The traditional training group received extra singing lessons amounting to one hour a week in two half-hour sessions

for eight weeks. Those children studied thirteen traditional and contemporary folk songs with piano, guitar, and percussion accompaniment. The purpose of the group was to act as an extra control in the event there was improvement due to the child's getting extra instruction and attention rather than due to the nature and content of the remedial training. The remedial training group received a special program of individualized and group instruction amounting to one hour a week in two half-hour sessions for eight weeks (229).

The general aim of the remedial training was to encourage the child to become aware of pitch fluctuations in the speaking voice, to extend these, and to transfer such skills to pitch production in a musical context. Some of the specific techniques used were developed by the researcher; others were developed from Gould's (1965) work. The following techniques were included:

1. The use of speech devices to extend the speaking range; e.g. shouting "hooray" or "hello."
2. Finding a note that the child could produce and sustain. This was his "personal" note. The child was taught to sing his name on this note, and find it on the piano. An attempt then was made to expand the child's repertoire of notes from this one note up to five notes.
3. Tone-matching exercises; i.e., a note was played on the piano within the child's vocal range. The child tried to match it by singing.
4. Humming everyday sounds: ambulance, cuckoo, and animal sounds.
5. Echoes: the child imitated first spoken and then sung words. A variety of other methods was used as well, including percussion work and feedback exercises--the child sang a note and this was played back to him through headphones (229).

All three groups of monotones were given a battery of tests of musical recognition and production both before and after the treatment. Training sessions were carried out by the schools' usual music teachers for the control group and, by final year students

specializing in music at a college of education for the traditional and remedial groups (230).

A test battery was constructed to assess performance of the children before and after the various treatment conditions. The battery was made up of four paper and pencil tests of musical recognition and five individual tests of musical production. All instructions and test items were presented on tape. The musical recognition tests were given in a group; the production tests were given individually. The child's response was tape recorded for later scoring by a panel of judges. The musical recognition tests were as follows:

1. Single note recognition (ten items). Two notes were played on a piano. The child was asked to write whether the second played note was the same as or different from the first. Intervals tested included the unison at C4, F4, G4, D4, B5, and intervals of a semitone, tone, major third, minor third, fourth, and octave.
2. Interval recognition (twelve items). The child was asked for same-different judgments for two-note, three-note, or four-note sequences.
3. Melody recognition (ten items). The child was asked for same-different judgments for one-bar, two-bar, three-bar, four-bar, or eight-bar melodies.
4. Rhythm recognition (ten items). Same-different judgments were required for rhythmic sequences ranging in difficulty from easy quarter note minimum values up to sequences with triplets.
5. Single note production. A note was played to the child twice, and he was asked to sing it. The range explored was from A4 to C5.
6. Interval production (eight items). Sequences of two notes, three notes, or four notes were played to the child twice; the child was asked to sing them.
7. Melody production (two items). A tune was presented to the child twice, and the child was asked to sing it.
8. Rhythm production (six items). The child was presented with a box and hammer. Each item was played on tape twice, and the child was asked to tap the rhythm out. The difficulty level ranged from simple crotchet and quaver patterns to quaver and triplet rhythms. The child's response to each item as assigned a score of "zero," "one," or "two," according to whether the response was completely incorrect, partially correct, or completely correct, respectively.

9. Free song. The child was asked to sing any song or nursery rhyme he knew and was given a starting note within his own range. The melody was tape recorded and scored on a five-point rating scale (0 = tune completely unrecognizable even if words correct; 1 = part of tune recognizable at correct pitch; 4 = correct performance) (230).

A panel of three experienced school music teachers listened to the pretest and posttest tapes of the ninety monotones and the pretest tapes of the thirty normal children. The panel did not know: (a) which treatment group the child was in or (b) whether the test session was a pretreatment or posttreatment session (231).

In tests one to four (the paper and pencil tests), each item was scored correct or incorrect according to whether the same-different judgment was correct. On the single-note production tests, a point was awarded for each note within the range the child was judged to be able to sing. On tests six to eight, the child obtained a point for each response the panel judged to be correct. Test nine was the five-point rating of the child's free singing (231).

The pretest battery was examined to see whether the tests were reliable and whether they discriminated between children rated as monotones and those rated as normal singers. The items within each pretest were arranged in order of difficulty and pairs of items adjacent for difficulty obtained. Two comparable subtests were formed by randomly allotting one of each item at a difficulty level to the separate subtests. Each child was scored separately on the subtests, and split-half reliability coefficients were calculated, correcting for attenuation. The coefficients were high according to the investigator and those for the recognition tests were reported as lower. The exact data was not provided by the investigators (231).

The following results were reported by the investigators in regards to the production tests:

1. There was statistically significant discrimination between the test scores of monotones and normal singers for all tests except that of single-note recognition.
2. The production tests were better discriminators than the recognition tests.
3. The most discriminating test was single-note production.
4. The tests of interval production and free singing were moderately good discriminators.
5. Of the recognition tests, interval recognition discriminated best between the groups (232).

The pretest scores of the three groups of monotones were examined to see if their scores on the tests were equivalent prior to the training procedures. A one-way analysis of variance was carried out for each test separately. It was reported by the investigators that in no case was a significant difference between the groups found ($F < 1.0$, $df = 2$). It was concluded by the investigators that the three groups of monotone singers were of equivalent ability prior to the experimental treatments (232).

The effects of the three training procedures were examined using a two-factor, split-plot analysis of variance. Factor "A" was treatment group (remedial, traditional, or control procedures); Factor "B" was occasion of test (pretest or posttraining score). A number of specific comparisons among means were made. The researchers were also interested in the comparison of the remedial and traditional groups in order to see whether the specific content of the remedial program affected the children's change in score over and above any effects attributable to extra attention (232).

The results of the comparisons follows:

1. There was reportedly a significant effect of occasion of test for seven of the nine tests in the battery.
2. Inspection of the means showed this to be an increase in score from pretest to posttraining occasion.
3. The only tests not showing significant improvement over the period were the single-note recognition test and rhythm recognition.
4. For no test was the main effect of Factor "A," or the treatment group significant, and the interaction between treatment group and occasion of test reached significance only for test six, the one on interval production.
5. Turning from the omnibus analysis of variance to the specific comparisons of interest, it was observed that the remedial group showed a greater change in score on both the test of single note production and the test of interval production.
6. The remedial versus training comparison was significant in both cases which indicated that the specific content of the remedial program was exerting some influence.
7. The contrast of the two extra attention groups with the control group also reached significance for test six on interval production.
8. On test two, interval recognition, the two attention groups differed significantly from the control group, but not from each other (235).

The investigators made the following conclusions based on the study:

1. It was possible to effect some improvement in pitch production among children rated by their teachers as monotones.
2. All groups of children showed a significant increase in score over the eight-week period on test of interval and melody recognition and on single note, interval, melody, and rhythm production.
3. The children's singing test nine on free song, also improved over the period (236).

The investigators hypothesized that the improvement might have been due either to maturational factors operating over the period between pretests and posttests, or to specific practice effects in the test situation (236).

It was reported that the remedial training group showed greater improvements than the other two groups on two tests—single-note production and interval production. On single-note production, the improvement for the remedial group was reportedly greater

than the mean improvement of the other two groups; specifically, the remedial group showed greater improvement than the traditional group. This pattern of improvement was more marked for test six, on interval production. The investigators reported that for the aspects of voice production, the remedial program did seem to have improved performance. The remedial program worked in detail on extending this range via the concept of the child's "personal" note. The interval production test required the child to sing short sequences of notes presented. Once more, training in interval production was given in the remedial procedure. The child started with his "personal" note and intervals were presented based initially on this note. When this was mastered, the same interval and word was presented up one semitone (G4 and Eb4) and so on up to A4 and F4. The investigators reported a lack of transfer from traditional singing lessons to performance on the test battery (236).

The investigators found it disappointing that no differential improvement was found for the remedial group in the test of free song because this was the measure that was mentioned as coming closest to what most people would regard as singing. No group showed significantly greater improvement than the others. The traditional group was better than the control group only on the test of interval recognition (237).

The conclusions drawn from the remedial training study were said to be reported with caution, due to the small number of children involved and the limited reliability and discriminating power of the tests of the battery. The test results suggested, that poor pitch singing is not associated with recognition defects. The monotone group showed no significant differences from the normal singing group on test one, on single-note recognition (237).

The study was said to be too short and crude to indicate which element of the remedial training brought about the increased vocal range on test five and the improved production of intervals on test six. It was reported that the extra singing lessons given to the traditional training group were largely ineffectual (238).

Feierabend (1985) researched the effects of specific tonal pattern training on singing and aural discrimination abilities of first grade children. The subjects (N =70) were four first-grade classes from an elementary school in suburban Philadelphia. The socio-economic status of the community ranged from lower middle-class to predominantly upper middle-class (17).

Prior to instruction, a listening test and a singing test were administered to the four classes. The listening test items were constructed of tonal patterns derived from major tonic and dominant functions. Students were shown a picture of two stars and were told that the two pictures were the same. Students were then shown a picture of a star and a circle and were told that the two pictures were different. Two identical tonal patterns were then played and the class was asked if the patterns were the same or different. Forty items were included in the test; twenty items were matched pairs of tonal patterns and twenty items were unmatched tonal patterns. Matched and unmatched pairs were randomly ordered on the test (18).

One week later a singing test was administered to each first-grade class. Students were tested individually. Items on the singing test were derived from those items which served as matched pairs on the aural discrimination test. In addition, items on the singing test were in the same order as they were presented on the aural discrimination test. Students were asked to listen to individual tonal patterns and were asked to sing exactly what they

heard. Two practice items were provided at the beginning of the test. All items were said to be within the children's singing range. Both the stimulus and response were recorded (19).

Three of the first-grade classes served as experimental groups and the fourth class served as a control group. Each experimental group was instructed for approximately five minutes every morning for seven weeks. Experimental instruction consisted of the researcher singing two-note, three-note, and four-note patterns on a neutral syllable alternating between tonic and dominant functions. After the class repeated each tonal pattern, the researcher sang the next tonal pattern without pause. Experimental Group One was presented patterns which were said to be easy to perceive aurally regardless of singing difficulty. Experimental Group Two was presented patterns which were said to be easy to sing regardless of aural difficulty. Experimental Group Three was presented only those patterns which were both easy to sing and easy to discriminate aurally. The control group received no special instruction in echoing tonal patterns. In addition to the instruction given to the experimental groups, all four first-grade classes continued to attend two thirty-minute general music classes each week. Instruction in the general music class consisted of singing songs with piano or recorded accompaniment and movement activities (19).

Criteria for levels of singing difficulty were derived from a preliminary study by the investigator which consisted of two-note and three-note pitch patterns of various interval sizes. Criteria for levels of aural difficulty were derived from research by Gordon (1978) (20).

The recorded responses from the singing test were evaluated by three judges using a five-point rating scale which was developed in the preliminary study. Inter-judge reliability for the rating scale was determined for the singing pretests and posttests. Reliability ranged from $r = .72$ to $r = .86$. Test-retest reliability for the students' singing performance was reported ar $r = .92$ (21).

To determine if any of the various treatments had a significant effect on either singing or aural discrimination abilities, scores on the pretests and posttests were derived for the listening test and for the singing test. Four one-way analyses of covariance were performed; one to identify differences in singing abilities, and three to identify differences in aural discrimination abilities derived from total listening scores on unmatched pair items. Pretest listening or singing scores served as the covariate for the corresponding posttest in each analysis (21).

For the three experimental groups and the control group, pretest scores on the singing test were correlated with: 1) the total pretest score on the listening test; 2) the pretest score for the matched pairs on the listening test; and 3) the pretest score for the unmatched pairs on the listening test. The same procedure was followed using posttest scores. Correlations between pretest and posttest listening and singing test scores were calculated to investigate how the various treatments affected the comparative relationship between pretest and posttest singing abilities and aural discrimination abilities (22).

The investigator reported the following results:

1. None of the treatments had a significant effect on singing and/or aural discrimination abilities of first-grade students.
2. Observed group means on the singing test scores and listening test scores improved from pretest to posttest for all groups.

3. Of the three experimental groups, students echoing patterns that were easy to aurally discriminate made the greatest observed gain between singing pretest and posttest means.
4. Students echoing patterns that were easy to sing made the greatest gain between total listening pretest and posttest means.
5. Observed means derived from scores on the matched pair items of the listening test improved most for students who echoed patterns that were easy to sing.
6. Observed means derived from scores on the unmatched pairs increased most for the control group (35-38).

The investigator reported that from the results of the study, it could not be concluded whether groups of first grade children which received: 1) instruction in patterns which were easy to aurally discriminate; 2) instruction in patterns which were easy to sing; 3) instruction in patterns which were easy to both aurally discriminate and easy to sing; or 4) no pattern instruction, learned to aurally discriminate or to sing patterns better. Changes in mean scores and changes in the correlations between singing and listening test scores suggested the following:

1. Echoing tonal patterns that were easy to aurally discriminate but were varied in singing difficulty decreased the relationship between singing and listening abilities.
2. On the basis of the observed means, students who echoed patterns varied in singing difficulty could be expected to develop better singing skills than students who echoed patterns that were easy to sing.
3. Echoing tonal patterns that were easy to sing but were varied in aural discrimination difficulty increased the relationship between singing and listening abilities.
4. Echoing patterns that were easy to sing regardless of aural discrimination difficulty, may improve students' aural discrimination abilities better than if they echoed patterns that were easy to aurally discriminate (47-49).

Richner (1976) conducted a study comparing what he defined as "three common methods of elementary school music instruction" and an experimental remedial method

in order to study the effects of those methods on the ability of inaccurate singers in grades three, four, and five, to reproduce pitches and sing melodic phrases (1).

The term "accurate singer," was used in the study to designate singers who consistently sang in tune in both individual and group situations. The term inaccurate singer was used in the study to designate a person who sang out-of-tune in both individual and group situations. Inaccurate singers were initially identified as those singers who failed to sing the correct pitches when singing songs with their classmates. They were then tested individually and more specifically identified, for the purposes of the study, as singers who could accurately reproduce no more than twenty of the pitches on the *Pitch Reproduction Test* (5).

The following null hypothesis was tested:

There is no statistically significant difference, at the $p = .05$ level of significance, in the ability of inaccurate singers in the fifth grade, to vocally reproduce pitches and sing melodic phrases between the various treatments (24).

Treatment I consisted of the music instruction which was normally provided by the classroom teacher. Treatment II consisted of two twenty-five-minute music classes per week which emphasized all areas of music education, taught by an elementary music specialist to each classroom. Treatment III consisted of two twenty-five-minute classes per week consisting entirely of singing songs, taught by an elementary music specialist to small groups of inaccurate singers. Treatment IV consisted of two twenty-five-minute classes per week of specialized, remedial voice training, taught by an elementary music specialist to small groups of inaccurate singers (24).

The inaccurate singers for the study were selected from students in the third, fourth, and fifth grades in one suburban school district. Music instruction was the responsibility of each classroom teacher (25).

The student population consisted of approximately 5200 students in grades Kindergarten through twelve. The student groups were comprised of grades Kindergarten through five in the elementary schools, grades six and seven in a middle school, grades eight and nine in a junior high school, and grades ten through twelve in the senior high school (25).

The population from which the sample was drawn, consisted of the inaccurate singers in the third, fourth, and fifth grades (27). A total of 843 children, all of the third, fourth, and fifth grade children in the four schools, were involved in the initial screening. The initial selection of inaccurate singers was made by the music specialist and classroom teacher on the basis of listening to children as they sang familiar songs with their classmates (28).

The population of inaccurate singers, as identified in the initial screening of the four schools used in the study, consisted of a total of 287 children. It was noted that the percentage of inaccurate singers was approximately the same at each grade level (28). More boys than girls were identified as inaccurate singers in the initial screening. School I reported twenty-eight percent of girls and seventy-two percent of boys to be inaccurate singers. School II reported twenty-two percent of girls and seventy-eight percent of boys to be inaccurate singers. School III reported thirty-five percent of girls and sixty-five percent of boys to be inaccurate singers. School IV reported eighteen percent of girls and eighty-two percent of boys to be inaccurate singers (29).

After the initial screening, a list was prepared of the inaccurate singers interviewed prior to the test. The study was explained to each student, and the detailed and requirements of his/her participation were outlined. All except two students indicated a willingness to participate (29).

After the random selection and individual interview, each child was given the pretest individually. Some of the selected children were not used in the sample because they could correctly reproduce more than twenty out of the fifty-five pitches on the individually administered pretest. Although those children sang inaccurately in the classroom, they were not used as a part of the sample in order to restrict the study to those students who sang inaccurately in both group and individual situations (30).

The sample used in the study was comprised of seventy-seven inaccurate singers—approximately twenty-seven percent of the population of inaccurate singers identified in the initial screening. In the sample, eighteen percent of the inaccurate singers were girls and eighty-two percent were boys (30).

The inaccurate singers in the sample at School I, and School II, were not separated from their classes except for the testing. At School III, and School IV, the inaccurate singers in the sample were included in groups of ten selected from each grade level for the music instruction. A few accurate singers were included in each group in order to avoid the negative self-image which might have resulted from membership in a group composed only of inaccurate singers. The number in each of those groups was limited to ten because Treatment IV, the remedial help for inaccurate singers, required some individual work. Treatment III, the treatment that was based on singing songs, was kept

to a maximum of ten in order to be a control to the individual attention factor in the smaller group of Treatment IV (34).

Measurement was done by a test devised for use in the study: the *Pitch Reproduction Test* (34). *The Pitch Reproduction Test* was designed to utilize the strobotuner, an electronic device that eliminated the need for subjective judgment about pitch accuracy. In order to use the strobotuner, an individually administered test was needed (35).

The first fifteen items, Part I tested the immediate vocal reproduction of one sound at a time. The only tonal memory necessary for this portion of the test was the memory for the one pitch. Therefore, Part I measured pitch reproduction without the loss of pitch information that might have resulted from longer sequences. Part I of the test used the strobotuner to check the accuracy of each of the pitches as the student responded vocally (36).

Parts II and Parts III of the test measured the student's ability to sing melodic phrases accurately. That portion of the test required memory of tonal patterns. The same instrument was used for both the pretest and the posttest. Repeating the test with all groups minimized assumptions concerning the comparative difficulty of melodic phrases and assumptions concerning the relative familiarity of the songs used in the test (36).

A trial test of thirty-two samples was checked for split-half, odd-even reliability. The correlation between odd and even responses was $r = .98$. The estimated *Spearman-Brown* reliability for the total instrument was $r = .98$ (36).

The study used a four-group design with subjects randomly selected for the treatments from the population of inaccurate singers within each of the four schools. The independent variable in the study was the method of instruction (36).

The four treatments were assigned randomly to the four different schools used in the study. The first method of instruction, Treatment I, was assigned to the first school selected from a list of random numbers. The inaccurate singers who were identified and tested at that school received no music instruction except that which would normally be provided by their classroom teachers. That group was an important factor in the study because it was representative of the self-contained classroom approach to music education used in many school districts statewide and nationwide. The research plan was discussed with the teachers at the school, but no details of the remedial treatment were revealed. The classroom teachers were instructed to continue teaching music in their usual way (37).

The second method of instruction, Treatment II, was assigned to the second randomly selected school. Treatment II consisted of two regular music classes of twenty-five minutes each, taught to all classrooms of third, fourth, and fifth grade children by a music specialist. The inaccurate singers received no individualized attention and were not separated from their classes except for the testing. A balanced program of activities including rhythmic, singing, instrumental, note reading, listening, and creative activities was offered in those music classes (37).

The third method of instruction, Treatment III, was assigned to the third school. Treatment III consisted entirely of singing songs. The inaccurate singers, along with several of their singing classmates, were assigned to three small groups of nine or ten—one group for each grade level. For two twenty-five-minute periods each week, those small groups of children sang a wide variety of songs under the direction of a music specialist. All of the singing was done by the group; the individuals in the group did not

sing alone. The children in the Treatment III groups received no vocal help and no special instructions; the time was used only for singing. The songs were always sung with words; no neutral syllables were used (37).

The fourth method of instruction, Treatment IV, was assigned to the fourth school. Treatment IV consisted of remedial vocal instruction which was given for two twenty-five-minute periods each week. The inaccurate singers in the sample, along with several independent singers, were assigned to groups of nine or ten. There was one group for each grade level (37).

Vocalizing techniques were used with the entire group and with each individual in the group for approximately twenty minutes of each period. The exercises began with locating whatever pitch the student could sing on the piano. Instructions to sing "more loudly" helped the students to sing higher pitches. As the child's range began to expand, additional exercises which included the new pitches, helped to establish those pitches. The vowel sounds of "u" and "o" were used most frequently in the exercises (38).

After single pitches could be reproduced within the range necessary for a given song, the transition from exercises to songs was encouraged by the singing of song melodies with neutral syllables. Songs which contained stepwise patterns were used at the beginning of the transition. Those songs were followed by songs that contained small and familiar skipping patterns. The final five minutes of each period were used for singing songs with words. Those songs were favorites, often selected by the children, and were sung for enjoyment (38).

The investigator reported that in order to minimize the effects of testing, the same instrument was used for both the pretest and posttest in all of the schools. Data was

collected individually from each child in the sample. One student at a time was taken from his classroom to the testing room where the strobotuner was explained to him and the *Pitch Reproduction Test* was administered (39).

The accuracy of each subject's vocal reproduction on each of the fifty-five test items was recorded by the dichotomous rating of "match" or "no-match" by the music specialist who administered the test. The test administrator was a highly qualified elementary music specialist with eleven years of experience in teaching music to children of that age group. The same teacher did the individual testing in all of the schools for both the pretest and the posttest (39).

The pitches for Part I of the test were played on a chromatic pitch finder. If the student's reproduction of the pitch was accurate enough to light the correct band on the strobotuner, a correct response was noted for that item on the test sheet. The degree of accuracy necessary to completely stop the rotation of the lines on the strobotuner was not required for a correct response (39).

The melodic phrases, Parts II and Parts III of the test, were played two times on the piano before the student responded. The test administrator also played each phrase on the piano while the student attempted to sing it. Each pitch sung correctly by the student was noted by the administrator on the test sheet (40).

Results of the study were reported by the investigator as follows:

1. The analysis of covariance indicated that significant differences existed between the treatment groups at the fifth grade level and at the third grade level.
2. There were no significant differences between the treatment groups at the fourth grade level.
3. Six separate comparisons made between the treatment groups at the fifth grade level indicated that the inaccurate singers who received

Treatment IV, the remedial voice training, improved significantly, at the $p = .01$ level, in comparison to all other treatment groups.

4. The comparisons made between the treatment groups at the third grade level indicated that the inaccurate singers who received Treatment III and those who received Treatment IV improved significantly, at the $p = .01$ level for Treatment III, the $p = .05$ level for Treatment IV, in comparison with the inaccurate singers in Treatment I. No other comparisons were significant at the third grade level.
5. The null hypotheses which stated that no significant differences existed in mean scores between the treatment groups were rejected for the fifth grade level, and for the third grade level (54).
6. The null hypothesis which stated that no significant difference existed in mean scores between the treatment groups at the fourth grade level was accepted.
7. The *t*-values computed separately for the "A" and "B" sections of the test indicated that the Treatment IV group improved significantly in the singing of melodic phrases with both syllables and words in comparison with the Treatment I group. Therefore, the null hypothesis which stated that no significant difference existed in the ability of inaccurate singers to sing melodic phrases with words or syllables, between children in the various treatment groups, was rejected (55).

The following conclusions of the study were made by the investigator:

1. The remedial voice instruction, Treatment IV, had a significant positive effect on the ability of inaccurate singers in the fifth grade to reproduce pitches. This conclusion followed from the statistical comparisons of the adjusted mean scores of the four treatment groups. The improvement made by the inaccurate singers in the remedial treatment group was significant at the $p < .01$ level in comparison with each of the other fifth grade treatment groups.
2. There were no significant differences in the mean scores of inaccurate singers between the treatment groups at the fourth grade level. Each of the four treatment groups improved in pitch reproduction ability between the pretest and posttest, but the gain of any one group was not significant in comparison with the other fourth grade treatment groups.
3. Treatment III, the group in which singing was emphasized, and Treatment IV, the remedial voice instruction, had significant, positive effects on the ability of inaccurate singers in the third grade to reproduce pitches. The improvement made by the inaccurate singers in the Treatment III group was significant at the $p < .01$ level, and the improvement made by the Treatment IV group was significant at the $p < .05$ level in comparison with the third grade Treatment I group (55).
4. There was a significant difference in the ability of the students to sing melodic phrases with syllables and the ability of students to sing these same phrases with words, in the total sample and in Treatment I. This

conclusion followed from the statistical comparison between the mean scores of Part A and the mean score of Part B for each of the treatment groups and for the total sample.

5. The inaccurate singers in the Treatment IV group showed significant improvement in the ability to sing melodic phrases with syllables in comparison with the Treatment I and the Treatment III groups. The inaccurate singers in the Treatment IV group showed significant improvement in the ability to sing melodic phrases with words in comparison with all other treatment groups (56).

The follow recommendations were made by the investigator for further study:

1. The remedial method, Treatment IV, should be tested with children from other socio-economic and geographic backgrounds and with larger numbers of children.
2. A study be conducted which measures attitudes of students toward singing and correlates those attitudes with the various methods of classroom and remedial music instruction.
3. Investigate why some children sing inaccurately in group situations even though they have the capability of singing accurately in individual testing situations.
4. A longitudinal study should be conducted to determine whether the improvement made by inaccurate singers during the treatment period would continue, remain stable, or regress upon termination of the treatment (56).

BREATH CONTROL MANAGEMENT

Phillips (1983) investigated the effectiveness of group breath control training on physical and vocal singing measures among subjects in grades two, three, and four. Pretest and posttest training data for each subject on four dependent measures (vocal range, vocal intensity, tonal duration, and pitch accuracy) were analyzed for the effects of the treatment (reflected in measures of abdominal and thoracic movement and vital capacity) between groups and among grade levels (129).

The subjects consisted of all forty-four girls and boys in grades two, three, and four at a private Christian school in Ravenna, Ohio. Twelve were in second grade, twenty were in third grade, and twelve were in fourth grade. The subjects were randomly assigned

within each grade level to either the control or experimental group. Each group consisted of twenty-two subjects, with each grade level being equally represented in each group (129).

Health checks for normal hearing and adverse physical problems were conducted via student records, and no subjects were eliminated for either cause. An absence rate of twenty percent class attendance was established, and no subjects were lost to excessive absence. The parental/guardian written permission form was used to determine via a question to the parents or guardians, that none of the subjects had previous vocal training experience. No specific type of vocal training other than that of the normal song training instruction, and two weeks each at the beginning and end of the study were used for testing procedures and data gathering. The study began in January, and ended in June. Any training sessions missed for regular or emergency school closings were rescheduled (130).

In the semester preceding the experiment, the subjects met together as a choir for two, one-half hour sessions per week. This was a non-select group, and all students were required to participate. For the semester of the experiment, the subjects were randomly assigned to either the experimental or control group, which then met separately for two, one-half hour sessions per week. The additional time needed for this scheduling was made possible through the cooperation of the classroom teachers. The experimental and control groups met consecutively, and the schedule order was alternated between the two days per week of instruction. There was a total of thirty-six sessions for each group (131).

The choir format was continued during the course of the experiment. The reasons given to the subjects for separating the original group into two sections were to create a more manageable classroom environment, and to allow for smaller group work. Both groups were advised that they would be learning the same songs, with a combined concert at the end of year as the common goal. The subjects were not informed of the fact that they were participating in an experiment, nor that the experimental group was receiving any additional type of training than that received by the control group (131).

The psychomotor training procedures combined both psychological and physiological approaches for voice training, and emphasized physiology. Physical conditioning exercises were the first part of the training sequence as used in this study. These were intended to activate the body for singing. This was accomplished through the "muscle movers" (stretching, bending, shaking, etc.), and posture development exercises. Emphasis on "the body as an instrument" was carried through all of the exercises for conditioning. Phillips noted that posture is directly related to breath control in singing and thus, was emphasized as a preliminary to, and a corollary of the breathing exercises (132).

Pretests and posttests of the four dependent measures (range, intensity, duration and pitch accuracy) as well as vitalometer and pneumograph readings were obtained from each subject. A vocal range test was administered in a separate session by the investigator and a female assistant who was also the regular music teacher. The other five measures (intensity, duration, pitch, vital capacity and body-wall displacement) were taken in a second session during the two-week pretest and posttest periods (137).

All singing trials were tape-recorded using two four-track stereo reel-to-reel tape recorders, and two low impedance dynamic microphones. The taping was done at seven- and a-half cps., using high fidelity, low noise recording tapes. One recorder was used to record the pitch accuracy trials, whereas the other was used to record the intensity, duration, and range trials. In this way, a separate tape of pitch accuracy responses was available for the judges to rate (138).

The data from this study were subjected to a two-way multivariate analysis of covariance (MANCOVA) as a part of a total multivariate analysis of variance (MANOVA) package. The posttest means of the four dependent variables (vocal range, vocal intensity, tonal duration, and pitch accuracy) were analyzed separately for the effects of three covariates (pretest dependent mean, pretest and posttest means for the three breathing measures). The posttest mean for each dependent measure and the three covariate means constituted a set of variables in this multivariate procedure. The two independent variables, breath control training and grade level, constituted a two-by-three design with two groups (experimental and control) and three grade levels (second, third and fourth) (155-160).

The results of the study were reported as the following:

1. Breath control training had significant effects on the singing ability of students in grades two, three and four.
2. The effectiveness of psychomotor breath control training could be effective in approaching child vocal training.
3. The exercises and vocalises used in the training helped develop abdominal/diaphragmatic/costal breathing.
4. The absence of a significant grade-by-training interaction and the means of the three breathing measures for the experimental group indicated that subjects at each grade responded to the breath training in the same manner.
5. It was found that abdominal expansion increased, thoracic expansion decreased, and vital capacity remained stable.

6. The means of the breathing measures for the control group changed in the opposite direction of those in the treatment group.
7. The control group's abdominal contraction increased, thoracic expansion increased, and vital capacity increased.
8. The experimental group moved in the direction of the breath control training goal of more abdominal expansion, but the control group moved in the opposite direction with a greater expansion of the chest upon inhalation for singing.
9. Those differences between the experimental and control groups were found at all three grade levels (180).
10. All four dependent measures (vocal range, vocal intensity, vocal duration and pitch accuracy) were found to be significant between groups.
11. Of those, both groups improved on vocal range and pitch accuracy. In those cases, however, the means on the posttests for the experimental group were higher than for the control group. This was true also for the vocal intensity measure in which the experimental group increased from pretest to posttest, whereas the control group decreased (181).
12. The dependent measure of vocal duration was also significant between groups. However, the experimental mean decreased from pretest to posttest while the control group mean increased. This finding suggested that breath control training was not effective in improving the vocal duration measure for the experimental group (181-182).
13. Psychomotor breath control training was an effective means by which to improve the vocal range, vocal intensity and pitch accuracy of elementary students. This suggested that such training be considered in addition to the song approach in teaching singing in the elementary grades.
14. Psychomotor breath control training can be implemented successfully to improve the singing ability of elementary students in grades two, three and four (183).

The investigator made the following summary conclusions based on the study:

1. Group breath control training for singing of children in the second, third and fourth grades had a significant effect on changing breathing habits from chest to abdominal/diaphragmatic/costal breathing.
2. Group breath control training for singing of children in the second, third and fourth grades had a significant effect on singing ability, especially on measures of vocal pitch range, vocal intensity, and pitch accuracy (187).

The investigator made the following recommendations for further study:

1. Extending the training portion of the study to one or more years may help to determine the maximum extent to which breath control for singing can be developed in children.
2. Investigating the possibility of beginning breath control training with children younger than second grade.
3. Investigating the effects of breath control training on children's vocal quality.
4. Investigating the effects of breath control training during exhalation in singing in addition to inhalation before singing.
5. Investigating the effects of breath control training on laryngeal control mechanisms for singing in children (187).

Gackle (1987) examined the effects of selected vocalises on the improvement of tone production in the junior high school female voice. The following questions were addressed: (1) Can tone be improved through the use of specific vocalizes to reduce extraneous breathiness and increase vocal line? (2) Can vocal development and more efficient use of the female adolescent voice be facilitated through systematic training within the choral setting? (66)

The sample population (N = 129) was comprised of seventh, eighth, and ninth grade female students from three junior high schools in the Dade County Public Schools. Forty-five percent were black, thirty-seven percent were Hispanic, seventeen percent were caucasion, and one percent were "other." The ethnic balance in the study was said to represent the school community (68).

A nonequivalent control group design was employed. The control group (N = 56) received all aspects of the singing procedure except the specific vocal training for breath management, vowel unification, and resonation which was used with the experimental group (N = 73) (69).

The treatment variable was the selected vocal exercises for proper management of breath, resonation, and unification of the vowel stream for increased connection of vocal line. The treatment was administered by the regular music teacher in each school, using the daily lesson plans provided by the researcher (69).

The experimental group met for six consecutive weeks for a total of thirty sixty-minute segments, for a total of thirty hours of instruction. During each class period, the group received approximately twenty minutes of training using the selected vocal exercises outlined in the method. The exercises were based on those used by the researcher and a summary of vocal techniques used by various selected choral conductors (70).

Choral selections were chosen by the researcher for their inherent pedagogical potential regarding style, musical line, and specific diction problems. The researcher said that an important aspect of the warm up process was the incorporation of vocalises created directly from specific parts of the music (71).

Five dependent variables were employed in the study, four quantitative measures and one quantitative measure. The four quantitative variables included a measure of the mean speaking pitch, singing range, phonation duration, and pitch perturbation. The quantitative variable was a performance measure which was based on judges' composite ratings of perceived tone quality and connection of vocal line. Pitch perturbation was quantified through the use of a *Kay Elemetrics Visi-Pitch Model 6087* interfaced with a microcomputer (73).

The researcher made the following conclusions based on the study:

1. There were no differences in total performance, mean speaking pitch, or singing range as the result of the use of selected vocal techniques for proper breath management, resonation, and vowel unification.
2. There were differences in phonation duration and pitch perturbation as the result of the use of selected vocal techniques for proper breath management, resonation, and vowel unification (116).

The researcher made the following recommendations for future study:

1. Replication of the study with an extended treatment time for as much as eighteen weeks in other geographic locations throughout the country. Also changing the testing procedures to control for various physiological fluctuations that effect the female adolescent voice.
2. A longitudinal study of the female adolescent voice to determine various stages of vocal development with respect to physiological maturation. An inter-disciplinary approach should be encouraged, utilizing the professional expertise of laryngologists, synocologists, and educators, primarily vocal music educators.
3. Experimental research concerning the use of vocal pedagogy methods with all voices—child, adolescent, and adult.
4. Development of recordings, films, video tapes and other media concerning vocal methods, the adolescent voice, and rehearsal techniques for use in teacher training.
5. Experimental research with the child and adolescent voice utilizing technological tools such as the *Visi-Pitch* spectrographic analysis, and other objective measures as correlates of aural perceptions of voice.
6. Experimental research concerning the effects of smaller class strategies or even studio training with the female adolescent voice utilizing professional voice instructors rather than school music teachers (118).

A study by Aaron (1991) used the *Visi-Pitch* in analyzing the pitch accuracy of inaccurate singers in grades, four, five, and six. Aaron employed a treatment with an experimental group that emphasized posture and breathing exercises (91).

The four musical characteristics assessed in the study as dependent variables included two measures of vocal production (pitch accuracy and range), and two psychological measures (pitch discrimination and tonal memory). Vocal coordination training, and gender served as independent variables. The instructional sequence for the experimental

group was based on the Phillips (1983) study, and combined four common treatments used with inaccurate singers: vocal coordination instruction (general physical coordination, posture, breathing, phonation, resonant tone production), and instruction in gaining flexibility crossing the register "break," vocalises, pitch shaping instruction, and additional song singing. The control group instruction included rote vocalises, pitch shaping instruction and additional song singing, but no vocal coordination instruction. Identical vocalises were sung by both groups, but subjects in the experimental group were given continued vocal coordination coaching during the exercises, while subjects in the control group were not (91-92).

Subjects were students in grades four, five, and six from one public elementary school. Upper elementary grades were used because it was stated by the investigator that many inaccurate singers in the third grade and lower outgrow their singing disabilities (92).

The public school from which the subjects were drawn were said to consist of a diverse socio-economic population. The average *Iowa Tests of Basic Skills* scores for grades four, five and six were reported at at percentiles forty, fifty, and forty-four respectively in the distribution of pupil scores (92).

During the first two weeks of the study, students in grades four, five and six were pretested on vocal pitch accuracy, vocal range, pitch discrimination, and tonal memory. Subjects chosen as a result of a vocal pitch accuracy test score greater than thirty-three cents (one-third of a semitone) were identified as inaccurate singers and assigned at random to an experimental or a control group. The remaining accurate students also were assigned at random to the two groups. Aaron reported that the inclusion of accurate singers was beneficial in avoiding a negative attitude among subjects (93).

Two procedures were conducted to determine if it was necessary to eliminate any subjects from the study. An eighty percent attendance criterion level was established. At the conclusion of the treatment one subject was eliminated from the study because she did not meet the attendance criterion. Health records were also examined to determine if any students were afflicted with hearing loss, speech disabilities, or other organic disorders that may affect singing. No students were eliminated from the study for those causes (93).

Of the 118 students in grades four, five, and six, 109 agreed to be tested. Seventy-five were identified as inaccurate singers. Data was collected for seventy-one subjects. In addition to the subject eliminated for poor attendance, one subject moved away during the semester, and two students were absent during the final week of data collection (93-94).

Two weeks prior to the pretesting, the investigator visited all of the potential subjects in their classrooms to practice the pretest song phrases and to discuss the students' role in the study. He explained that materials were being developed to assist children in improving their singing and that students were needed to take part in pretests, singing instruction, and posttests. Students were told that the test results would not have any effect on their music grades (94).

During the week testing began, the investigator practiced the pretest phrases in each music class involved in the experiment, discussed the study with the students, showed them the recording equipment, and checked to see if they had any questions. The two classroom visits also served to allow the students to become more comfortable with the

In scoring the *Test of Vocal Pitch Accuracy*, the frequencies of eight target pitches isolated from the three test phrases were measured electronically, and the deviations from the absolute pitch were noted in cents (100). Different sets of phrases (set A and set B) were used in the pretests and posttests because it had been observed that singing inaccuracies during the pretest may carry over to the posttest if the same melodies are used (Gould, 1968: Phillips, 1983: Roberts and Davies, 1975). The pairs of phrases were composed by the investigator using similar ranges and intervals. The parallel forms reliability and the internal consistency of the phrases (sets A and B) were calculated during a pilot study. At that time, twenty-four students from grades three through seven sang the phrases (sets A and B) and pitch deviation scores were calculated. The parallel forms reliability computed by correlating the total scores for each form of the test was $r = .90$. This coefficient represented the reliability of either form of the test (101).

Estimates of internal consistency (for the eight target pitches of each set) using *Cronbach's Coefficient Alpha* were $r = .81$ and $r = .87$ for forms A and B respectively. Coefficient alpha represented the average correlation among all possible split halves of a test (101).

The pitch models for the *Test of Vocal Pitch Accuracy* were included in a tape-recorded protocol that consisted of the investigator giving instructions and playing the phrases on an electric piano (102).

The pitch accuracy test was presented and recorded on two tape recorders using high output studio recording tape. All of the singing tests were conducted individually in a private room in the school (102).

investigator. Permission slips were given to all students prior to the pretesting and were collected by the regular music teacher and classroom teachers (94).

The duration of the study was twenty weeks. Sixteen weeks were given to actual instruction, and two weeks at the beginning and end of the study were used for pretests and posttests. Subjects in both the experimental and control groups received ten minutes of special instruction from the investigator during each regular music class. All music classes met twice weekly. Because of parent/teacher conferences, in-service days, special assemblies, and school cancellations due to weather, the classrooms received varying amounts of instruction. The two fourth-grade classes both received twenty-five sessions, and the sixth grades received twenty-five and twenty-seven sessions. Though the sessions per classroom varied slightly, the number of sessions given to each control group and experimental group was identical (95).

Four pretests and posttests (vocal pitch accuracy, vocal range, pitch discrimination and tonal memory) were administered to all students in grades four, five and six. The pitch accuracy and range tests were administered individually because subjects sang alone during those tests, and the pitch discrimination and tonal memory tests were given to all students during their regular music classes as the tests were designed for group administration (100).

The author-designed *Test of Vocal Pitch Accuracy* was given first. The test was used to identify the inaccurate singers appropriate for the study, and to determine the pretest and posttest pitch accuracy scores for data analysis. This test consisted of singing three complete song phrases using lyrics (100).

The *Test of Vocal pitch Accuracy* was scored by the investigator using a *Visi-Pitch* (Kay Elemetrics Corp., Model 608705). The investigator estimated a reliability coefficient of $r = .98$ as a result of rescoring sixteen individual pitches of a pilot tape. The *Visi-Pitch* measured and calculated the fundamental frequency of a complex tone. Given musical input, the contour of the music was converted to a visual contour on a monitor. High pitches appeared as a row of dots near the top of the screen and low pitches appeared similarly near the bottom. Up to eight seconds of tones could be displayed at one time. The Visi-Pitch was interfaced with an *Apple IIe Computer*, and the operator moved two vertical cursors to the beginning and end of the pitch being measured. The computer calculated the mean pitch isolated between the cursors and displayed the frequency in Herz (103).

The tape recorder was patched into the *Visi-Pitch*; the tape was played and the frequencies of the eight target pitches were determined in Herz. Using *A Table Relating Frequency to Cents* (Young, 1952), the investigator transformed the frequencies to cent deviations. The pitch accuracy score used for data analysis consisted of the average number of cent deviations for all eight notes scored. Subjects with an average deviation of thirty-three cents (one third of a semitone) or larger were considered inaccurate singers (103).

The *Vocal Range Test* was given only to students identified as inaccurate singers. Subjects sang major triads using the neutral syllable "ah." Beginning on the pitches C4, E4, and G4, the triad was moved up by half-steps to the highest note of the subject's musical range, then, beginning again on C4, E4, G4, the triad was moved down to the lowest note of the subject's musical range. Three trials were administered in each

direction. A trial was terminated when the student stopped singing, or when a strained or forced sound became evident. The goal of the *Vocal Range Test* judging was to determine the "musical" frequency range. The "musical" frequency range was defined as the range from the highest musical tone a child could match down to the lowest musical tone he could match accurately (9-10). All range tests were scored by an assistant to the investigator (104).

The pitch discrimination subtest from the *Colwell Achievement Test* (Test 1, Part 1) was used for the dependent measure of pitch discrimination. This test was employed because no intervals smaller than a semitone were included, tones were played on musical instruments, it was designed for group administration, national and regional norms were included, and the manual gave high reliability and validity coefficients for the total test. The test manual noted a reliability coefficient of $r = .83$ for this part of the test. The *Wing Standardized Test of Musical Intelligence* (Test 3, Memory) was used because model melodies were followed immediately by altered melodies, all examples were played on a musical instrument, and all examples were composed specifically for the test. As limited reliability information was reported in the manual, *Cronbach's Coefficient Alpha* was included in the data analysis (104-105).

Posttests were administered the week after treatment was completed and lasted for two weeks. Only subjects identified as inaccurate singers were administered the *Test of Vocal Pitch Accuracy*, and the *Vocal Range Test.* Song phrases (set B) for the pitch test were practiced twice during the last two weeks of treatment, and modeled individually during the protocol with each subject before each test. All students were readministered the pitch discrimination subtest (Test 1, Part 1) from the *Colwell Music Achievement Test* and the

tonal memory test (Test 3) from the *Wing Standardized Test of Musical Intelligence* during regular music classes (105).

The investigator reported the following data from the study:

1. The data for the study were analyzed with four separate 2 x 2 analyses of covariance using a general linear model approach. Independent variables for each analysis included group (experimental and control) and gender. Dependent variables included two vocal production measures (pitch accuracy and range) and two measures of musical psychological acuity (pitch discrimination and tonal memory). The covariate in each analysis was the corresponding pretest score for each dependent measure (110).
2. Frequencies of percentage of inaccurate singers by grade level and gender were reported: (a) Grade 4: (76.4%); Boys (100%); Girls (55.5%); (b) Grade 5: (53.8%); Boys (66.6%); Girls (45.8%); (c) Grade 6: (77.7%); Boys (78.5%); Girls (77.3%); (d) All (68.8%); Boys (82.2%); Girls (59.4%) (113).
3. Null Hypothesis 1: There will be no independent variable (gender and treatment) by covariate (pretest) interaction effects for pitch accuracy."
   The null hypothesis 1 was rejected at the .01 alpha level. A significant group by gender by covariate interaction was found for pitch. Vocal coordination instruction was significantly more effective than the control group instruction for all boys in this sample of inaccurate singers from gardes four through six. This effect was demonstrated strongest with highly inaccurate singers of this gender. The experimental instruction was equally effective as the control group instruction for girls of moderate singing inaccuracy and less effective than the control group instruction with girls who were highly inaccurate singers (119).
4. Null Hypothesis 2: There will be no significant group by gender interaction for vocal pitch accuracy.
   The null hypothesis 2 was rejected at the $p < .05$ alpha level. The girls in both groups attained low pitch inaccuracy scores on the posttest (indicating high accuracy), while the boys only in the experimental group attained low pitch inaccuracy scores on the posttest (122).
5. Null Hypothesis 3: There will be no significant treatment group main effect for vocal pitch accuracy.
   The null hypothesis 3 was rejected at the $p < .05$ alpha level. The results of the ANCOVA revealed a significant group effect for pitch accuracy. An examination of the adjusted posttest means indicated that overall the experimental group sang more accurately than the control group on the pitch posttest (123).
6. Null Hypothesis 4: There will be no significant gender main effect for pitch accuracy.

The null hypothesis 4 was rejected at the $p < .05$ alpha level. The results of the ANCOVA revealed a significant gender effect for pitch accuracy. An examination of the adjusted posttest means indicated that overall the girls sang more accurately than the boys on the pitch posttest (123).

7. Null Hypothesis 5: There will be no independent variable (gender and treatment) by covariate (pretest) interaction effects for vocal range.
   The null hypothesis was retained at the $p < .05$ alpha level. This result indicated that entry ability did not effect posttest scores within group or gender (124).
8. Null Hypothesis 6: There will be no significant group by gender interaction for vocal range.
   The null hypothesis was retained at the $p < .05$ alpha level. This result indicated that the effect of the treatment did not vary for girls and boys (124).
9. Null Hypothesis 7: There will be no significant treatment group main effect for vocal range.
   The null hypothesis was rejected at the $p < .05$ alpha level. This result indicated that the experimental group sang with a wider vocal range than the control group during the posttest (125).
10. Null Hypothesis 8: There will be no significant gender main effect for vocal range.
    The null hypothesis was retained at the $p < .05$ alpha level. This result indicated that the adjusted posttest means showed no significant differences between genders in vocal range (125).

After a semester of instruction, Aaron made the following conclusions and summaries:

1. The effects of vocal coordination instruction on vocal pitch accuracy varied for gender and entry level ability.
2. Vocal coordination instruction was more effective in improving vocal pitch accuracy for boys than for girls.
3. Highly inaccurate boy singers benefited the most from vocal coordination instruction.
4. Girls achieved improved vocal pitch accuracy with either vocal coordination instruction or a general program of vocalise, pitch shaping instruction and song singing.
5. Vocal coordination instruction was effective in improving the vocal range of inaccurate singers.
6. Pitch discrimination was unaffected by vocal coordination instruction.
7. The effects of vocal coordination instruction on tonal memory varied for entry level ability.
8. The experimental treatment was effective in improving tonal memory for subjects with low pretest scores (132).

The investigator made the following recommendations for future study:

1. Effective methods of adolescent vocal instruction are strongly needed particularly for boys experiencing the change of voice.
2. Continued investigation of other vocal coordination instruction programs isolating different exercises and different sets of exercises may help to determine the efficiency of different components of such programs.
3. Investigating the effects of vocal coordination exercises with subjects of varying abilities may be worthwhile as the instruction also may have beneficial effects for subjects of average and above average vocal pitch accuracy.
4. The effects of the instruction on other variables such as tone quality, ability to sustain tone, ability to project tone, rhythmic accuracy and diction may yield valuable results.
5. Continued study of the relationship between vocalise and vocal coordination instruction is important.
6. The development of a standardized test of vocal pitch accuracy is another important area.
7. Continued investigation into the relationship of psychological music skills and singing ability (147).

Phillips and Aitchison (1995) investigated the effects of breath-control management instruction on elementary general music students' singing performance. The subjects were 303 students in grades four, five, and six in an Iowa community. The population was twenty-five percent Hispanic and four percent Laotian. Data was analyzed from 269 subjects; thirty-four students were eliminated due to absence, health problems, or failure to complete all of the testing (188).

Each grade consisted of four general music classes made up of approximately twenty-five students each. Students had been assigned at random to each grade. Intact classes at each grade level were assigned as either treatment or control. The treatment group consisted of two classes at each grade level (N = 127) and the control group consisted of two classes at each grade level (N = 142) (188).

General music classes met for forty minutes twice a week. One of the investigators saw each of the six classes in the experimental group once a week, during one class, for approximately fifteen minutes of psychomotor skills instruction in singing. This instruction time was extracted from that normally reserved for singing, which was approximately one-third of the time. On the second day of class per week for the experimental group, the regular music teacher continued singing activities with references to the treatment instruction. The control group received the normal singing curriculum from the regular music teacher, who used a song approach without formal vocal instruction (188).

The treatment period began in October of the school year and ended the first week the following May. Each class in the experimental group received the vocal skills instruction once a week from the investigator for twenty-seven weeks. This amounted to approximately 405 minutes of instruction, or nineteen percent of the total time each week for general music. Every attempt was made to keep the amount of time for singing the same for all classes (188).

The psychomotor skills instruction in singing for the experimental group consisted of the vocal method by Phillips (1992). There were five broad goals in this method, in which students learned to sing through activities related to (1) posture and respiration; (2) phonatory/speaking voice development; (3) resonant tone production; (4) diction; and (5) expression. Each of those broad goals consisted of objectives that related directly to the exercises and vocalizes of the method. In all, there were ninety instructional objectives and activities. Those were then divided into six levels of instruction of

increasing difficulty. The six levels were grouped into three learning sequences according to the year or grade in which the instruction began (188).

This study used the second learning sequence in the method because it was recommended for use when beginning the method with intermediate-age students. The subjects in the experimental group experienced six exercises in each of the five main areas of the method for instruction in a total of thirty exercises and/or vocalises by the end of the study. Approximately one new exercise was introduced each week, with other exercies being reviewed. When appropriate, each treatment period or class singing lesson was followed by the singing of songs from the regular general music curriculum (189).

The main investigator was introduced to students in the experimental group as a music teacher who was part of a project with the nearby university. The students were unaware that they were in an investigation (189).

The study used a variation of the *Posttest-Only Control-Design* by Campbell and Stanley (1963), with intact classes of randomly-assigned subjects. A 2 x 2 x 2 (group, gender, grade) factorial analysis of variance (ANOVA) was used in each of the statistical analyses (189).

Individual posttesting was administered following the treatment period. All students were administered the measures, but only the scores for those students who were present for all testing measures and a minimum of ninety percent of the treatment instruction were used in the statistical analyses (189).

The treatment emphasized the subjects' breathing mode for singing. Therefore, three dependent measures of respiration were assessed by one of the investigators. The first

two measures were taken simultaneously using a Respiradyne instrument for airflow evaluation. Subjects were asked to take a breath and to blow as much air as they could and as hard as they could into a mouth tube that was connected by a smaller plastic tube into the *Respiradyne* calculating instrument. Each subject used a nose clip to keep air form escaping through the nostrils. The investigator logged two statistics: (1) vital capacity, and (2) peak flow as expressed in liters exhaled per minute (189).

The third respiration measure consisted of asking the subject to inhale a full breath and to exhale for as long as possible. The investigator calculated the expiratory duration using a stopwatch and logged each of three trails in number of seconds of duration (189).

Upon finishing the respiration measures, each subject was taken to another office located adjacent to the music classroom. The two singing measures were explained to the student. The lower vocal range was tested by singing downward from middle C as low as the subject could sing comfortably. The upper range was tested by singing upwards an arpeggio from middle C as the subject could sing comfortably. The procedures were repeated until the investigator was assured that the vocal range was accurate. Scores for highest and lowest pitches and total range were logged for each subject using a standard range measurement form, which used a numbering scale of one to forty-nine for half-steps spanning three octaves (190).

The second measure of vocal performance was an assessment of pitch accuracy. The investigator played twice a three-note tonal pattern in standard quarter note rhythm on an electric keyboard, and then asked the subject to repond by singing the pattern on *loo*. There were four patterns using pitches within the octave of middle C4 to one octave above. The scale degree patterns were "five-three-one," "five-eight-seven," "six-four-

two," and "three-five-one." The investigator said that those were standard patterns found in children's song literature, and with which the children were familiar from singing in class. As the subject responded by singing the pitches of each pattern, the investigator subjectively marked on a form those pitches sung incorrectly and totaled those sung correctly, with a possible score of twelve (190).

The results suggested that the treatment was more successful for boys than for girls in producing use of the breathing musculature as measured by peak flow. The investigator hypothesized that older boys may be more motivated by exercises that are physical in nature, and boys may be developing more physical use of their muscles at this age. The finding that fourth-grade girls in the treatment group were superior to those in the control group suggested that the treatment was successful, to some degree (194).

The second measure of respiration, vital capacity, attempted to determine if the volume of air one is able to inhale-exhale had any impact upon breathing for singing. No main effect for group was found. The main effect for gender in favor of all boys suggested that the majority of boys were larger in stature than the girls. However, it was only the boys in the sixth grade who scored significantly better than sixth-grade boys in the control group on vital capacity. Mean scores for treatment boys in fourth-grade and fifth-grade were only slightly higher than those for the control group (194).

The measure of expiratory duration in seconds represented the skill of controlling the air column upon exhalation. A main effect for group in favor of the treatment condition was confounded by the group-by-grade interaction; only the fifth-grade subjects in the treatment group scored significantly higher than did those in the control group. Therefore, the exercises to develop control of air emission tended to work only with fifth-

graders, although the treatment mean for fourth-graders was slightly higher than that of the control group. The treatment and control means for sixth graders were almost identical (194).

The investigator said that singing with an ever-increasing vocal range is an objective of vocal music education. He added that the vocalizes in the treatment condition were used to exercise the upper and lower registers of the voice so as to maximize vocal range. He said for the highest pitch sung, the results showed the treatment group to sing, on the average, higher than the control, and girls, in general, to sing higher than boys. It seemed to the investigator that the treatment was successful in extending the subjects' highest pitch sung. It was found that the highest pitch did not increase by grade; students in all grades tended to sing as high as D5 (194).

The investigator said that the treatment condition seemed to have been successful in having students increase their lower vocal range. He said however, the group-by-gender interaction showed that only the boys in the treatment group sang significantly lower than those in the control (194).

Based on the results of the study, the investigator made the following conclusions:

1. Breath support may be improved with instruction.
2. Vital capacity seems to be affected more by age than by instruction.
3. Breath control may be improved with instruction, especially among fifth-grade students.
4. Highest pitch of vocal range may be improved with instruction.
5. Lowest pitch of vocal range may be improved with instruction especially for boys.
6. Total vocal range may be improved with instruction, especially for boys. Pitch accuracy appears to be greater for girls than for boys and male pitch accuracy may not be helped with instruction when they are reluctant to sing in the treble range (195).

Collins (2000) examined the effects of the presence or absence of breath-management exercises during vocal instruction on the vocal accuracy and attitudes toward singing of sixth-grade choral students (2). Two vocal skills were assessed to determine the effects of the independent variable of choral instruction on the singing ability of the students, specifically the dependent variables of vocal range and pitch accuracy. A third dependent variable, students' attitudes toward singing, was also examined as a secondary research objective. The effects of gender on the dependent variables were also determined. The independent variable involved the manipulation of the presence or absence of breath management exercises. The experimental group received vocal instruction that involved the use of breath-management exercises while the control group received vocal instruction that did not include breath management exercises. The specific instructional procedures were modeled after approaches presented in Fracht and Emmett's (1960) book *You Too Can Sing* and Phillips's (1996) book *Teaching Kids to Sing*. *World of Music* (1991) served as the text for the vocal instruction of the control group, and the core instruction for all groups. The *Making Music* and *Reading Music* sections from the textbook were chosen in place of breath management exercises for the vocal instruction of the control group. The investigator noted that *World of Music* recommended those sections for teaching students the basic skills of reading music. Those sections were only used during the warm up portion of each class session for the control group. Other sections in *World of Music* were used for the core instructional portion of each session. *World of Music* was selected by the investigator because it was the adopted music text of the school system where the study was administered (38).

Subjects for the study (N = 92) originally consisted of sixth-grade students from eight classes at four Kindergarten-eighth grade elementary schools. The schools were located in the Yadkin County School System, North Carolina, and were selected by the investigator for their convenience and subject congruence. The investigator served as the teacher at the four schools. Students were originally assigned to classes by the respective schools' homeroom teachers without consideration of abilities or other factors that might have induced contamination in the research design. The eight intact classes were then assigned to experimental or control group classifications by the order of class meeting time. All classes that met during the first period were designated, by the investigator, as the control groups, and all classes that met during the second period class were designated as the experimental groups (39).

Students were instructed two times a week, on an alternating schedule, so that the groups received the same amount of instructional time during the study. Two sessions at each school, one at the beginning and one at the end of the study, were used to gather data using an attitude questionnaire, and one session at each school, outside of the regular class time at the beginning and end of the study, was used for gathering data on vocal range and pitch accuracy. The duration of the study was nineteen weeks, with sixteen sessions devoted to instruction. The treatment began in August 1998 and continued through October 1998 (39).

The investigator administered a questionnaire modeled after an instrument used by Phillips and Vispoel (1990) to determine whether any of the students had received formal vocal instruction beyond that received in the respective schools. No students with formal

vocal instruction experience beyond the regular school setting were found; therefore, the investigator did not eliminate any student (41).

The testing of vocal range was administered privately in a separate classroom. Students were tested individually during the first week and last week of the instructional portion of the study. During the testing of vocal range, each student was first instructed how to sing into the microphone of a *Visi-Pitch.* The *Visi-Pitch* instrument recorded the frequencies of the tones sung into the microphone and converted the signal of each production tone to a digital frequency display. Each student was instructed to sing in the key of C major ascending the keyboard on the syllable "la" until they felt their singing was strained, uncomfortable, or non musical, as had been explained and demonstrated. An electronic keyboard was played as the students sang "la" on the ascending scale. The *Visi-Pitch* measured each step of the student's singing as it occurred, converted each sung step to frequencies, and provided comparable range performance indices as designated by the *Visi-Pitch* for the scale frequencies. This allowed the researcher to determine the highest pitch each student was able to sing. The singing of "la" while descending in the key of C major was then demonstrated, followed by each student's responding as described above. Again, the *Visi-Pitch* measured each step of the student's singing, converted each sung step to frequencies, and provided appropriate range performance indices for the scale frequencies. This allowed the researcher to determine the lowest pitch each student was able to sing. The range difference of the highest to lowest pitch each student could sing was recorded in Hertz and converted to cents using Young's (1952) conversion table—thus, providing the score indicating each subject's range (42).

The testing procedure for pitch accuracy was similar to the vocal range test. Each student was reminded of the correct way to sing into the microphone of the *Visi-Pitch*, followed by a demonstration of singing six ascending and descending triads including: (1) C4, E4, G4; (2) C#4, E#4, G#4; (3) D4, F#4, A4; (4) D#4, G4, A#4; (5) E4, G#4, B4; and (6) F4, A4, C5. Then, one triad was played at a time on the electronic keyboard as the *Visi-Pitch* measured the keyboard pitches. Each student was instructed to listen, and at the command "go," sing the six triads on the syllable "la," chromatically ascending the octave. The *Visi-Pitch* measured each note the students sang, converted each note to frequencies, and provided comparable pitch performance indices designated by the *Visi-Pitch* for the frequencies. The frequencies were converted to cents using Young's (1952) conversion table (43).

The attitude questionnaire was administered to classes as a group the first day of the study. The questionnaire was modeled after Phillips and Aitchison's *Self Perception of Singing Abilities Measure* (1995). Adjustments were made so that information could be obtained regarding how students felt about their own singing rather than how much they liked to sing in music class (44).

There were ten equally weighted statements on which the students indicated agreement or disagreement by marking a five-point scale indicating how each student felt about his or her own singing ability. The same test was used as a pretest and a posttest (44).

The chorus classes met for two forty-minute periods per week. To assure equality among groups, four classes, one class per school, were designated as experimental groups, and four classes, one class per school, were designated as control groups. The

researcher taught chorus to all groups. There was a total of sixteen instructional sessions per group (44).

All class periods consisted of two segments, vocal instruction and core instruction. The experimental and control groups received vocal instruction at the beginning of each class. The experimental group's instruction time included breath-management instruction at the beginning of each class period. This instruction was modeled after exercises presented in Fracht and Emmetts's (1960) book, *You Too Can Sing*, and Phillips' (1996) book, *Teaching Kids to Sing*. During this time, the control group sang warm-up exercises and warm-up songs chosen from the *Making Music* and *Reading Music* sections of *World of Music*. The Silver Burdett and Ginn music series, *World of Music* (1991), also served as the text for the core instructional portion of the classes. This segment included the last thirty minutes of each class and was identical for all groups (45).

The treatment sequence for the experimental group consisted of breath-management instruction. Included in this instructional process were posture exercises, breathing exercises, and vocal exercises which incorporated a psychomotor approach emphasizing breath-management instruction. Physical conditioning exercises began the instructional sequence; these were structured to teach students the proper posture for singing. Breathing exercises intended to establish abdominal-diaphragmatic-costal breathing followed. Vocalises emphasizing breath control were used to help students achieve improved vocal control (46).

The first set of exercises used in the experimental group took approximately two to three minutes of each class period and served the purpose of teaching students proper posture for singing which included activities in posture modeling, stretching, flexing, and

twisting (46). The second set of exercises used in the experimental group took approximately three to four minutes of each class period and served the purpose of teaching students proper breath-management techniques for singing. Imagery was used to encourage students to lower the diaphragm and expand the ribs and abdominal wall for inhalation (46). The third set of exercises involved approximately three to four minutes of each class period and included singing and the use of proper breath management. Activities included students singing a variety of vowel sounds and consonants both staccato and legato, practice in sustaining tones, singing at different dynamic levels, singing a variety of intervals, ascending and descending one and two octaves, and beginning a tone properly. The control group received vocal instruction that included warm-up exercises and warm-up songs for approximately two to three minutes, but did not include any breath management exercises. The singing exercises for the control group also included songs from the *Reading Music* section of *World of Music*. Specific skills such as sight-reading, singing intervals correctly, recognizing different scales, phrasing, and changing dynamics were the focus of those exercises. Singing songs to accomplish those skills was the main focus of the warmup activities for the control group and took approximately eight minutes (47).

The questionnaire modeled after Phillips' and Vispoel's 1990 study was used to determine if any students had received formal vocal instruction beyond the classroom school chorus program. Pretest and posttest scores for two dependent variables, vocal range and pitch accuracy, were obtained from each subject individually by means of the *Visi-Pitch*. Pretest and posttest scores for the secondary variable of attitude toward

singing were obtained from each subject individually by administration of the *Attitude Toward Singing* questionnaire (48).

Students' singing ranges and pitch accuracy, were determined by the use of a *Visi-Pitch*. The students were instructed to sing along with the keyboard in the key of C major ascending and descending the keyboard on the syllable "la." As previously instructed, all students began on C4 and ascended the keyboard in the key of C major until each chose to stop according to each individual's vocal comfort. Beginning on C4 again, all students descended the keyboard in the key of C major until each chose to stop according to each individual's vocal comfort. The students sang while the researcher played the scale using an electric keyboard. The range difference of the highest to lowest pitch each student could sing was recorded in Hertz and converted to cents, using Young's (1952) conversion table, as the score indicating each student's vocal range. One score for each student was used for the analysis (48).

The vocal range data between the experimental and control classes were gathered and means were calculated for the pretest scores revealing a mean range of M = 1550 cents for the experimental group and a mean of M = 1450 cents for the control group. A *t* test between the means of groups revealed no significance ($t = .74, p = .462$). Since the results were not significantly different between groups on the pretest for vocal range, the use of the scores for analysis of covariance was unnecessary. Collapsing data across groups, the calculation of means from the pretest for independent samples of gender revealed a mean of M = 1380 cents for the females and a mean of M = 1580 cents for the males. The results were not significantly different between the males and females

($t = 1.57, p = .12$) on the pretest for vocal range; therefore, the use of the scores for analysis of covariance was unnecessary for the grouping variable of gender (49).

The pitches were recorded in Hertz. The students were instructed to sing the ascending and descending triad as the keyboard was played. Students sang into the microphone of the *Visi-Pitch*, and their singing was recorded in Hertz. Students repeated this process for ascending and descending triads throughout one octave ranging from C4 to C5, with C5 being the last pitch in the triad that began with F4 (49).

For consistency of analysis, the data, which were recorded in Hertz, were converted to cents using Young's (1952) conversion table. The difference between the target pitches played on the electronic keyboard, which were also measured on the *Visi-Pitch*, and the pitches each subject sang were recorded in cents as the scores indicating each student's pitch accuracy skill (50).

The pitch accuracy data was analyzed across the experimental and control classes, and means were calculated for the pretest scores revealing a mean of M = 360 cents difference between the pitches played on the keyboard and the actual pitches sung by each student in the experimental group, and a mean of M = 405 cents difference between the pitches played on the keyboard and the actual pitches sung by each student in the control group. There was no statistically significant difference between groups on the pretest for pitch accuracy ($t = .66, p = .51$), therefore, the investigator said that the use of those scores for employing analysis of covariance on post-treatment data was unnecessary. Collapsing data across groups, the calculation of means for independent samples of the males and females for pitch accuracy revealed a mean of M = 272 cents for the females and a mean of M = 445 cents for the males. A statistically significant

difference between the males and females was revealed on the pretest for pitch accuracy ($t = 2.27, p = .03$), with females matching pitch more accurately, thereby supporting the use of these scores for employing analysis of covariance on post-treatment data (50).

The questionnaire measuring the students' attitudes toward singing contained ten statements of equal value. The students agreed or disagreed by marking a five-point scale. The attitudes toward singing data were combined for the experimental and control classes and means were calculated for the pretest scores. The results ($t = .35, p = .73$) were not significantly different between groups on the pretest for attitudes toward singing, and so the use of those scores for analysis of covariance was unnecessary. Collapsing data across groups, the means of males and females was calculated. Scores between the means of males and females ($t = .80, p = .43$) revealed no statistically significant difference between the males and females on the pretest for attitude toward singing, thereby making the use of those scores for employing analysis of covariance on post-treatment data unnecessary (51).

Analysis of variance was computed to examine posttest differences between the experimental and control groups for vocal range and between males and females for vocal range. Analysis of variance was also computed to examine posttest differences between the experimental and control groups for pitch accuracy. ANCOVA was computed to examine posttest differences and between males and females for pitch accuracy. Analysis of variance was computed to examine posttest differences between the experimental and control groups for attitudes toward singing and between males and females for attitudes toward singing (52).

After pretest group means and standard deviations for the dependent variables were calculated between the two groups, *t* tests were computed on group means of the pretest performance measures serving as dependent variables. Pretest group means and standard deviations for the dependent variables were calculated by gender as well, and *t* tests were computed on the means of the pretest performance measures serving as dependent variables (58).

The analysis of variance for vocal range indicated to the investigator that one and two-tenths percent of the variance in posttest scores were accounted for in the pretest. The analysis of variance for vocal range between males and females indicated that three percent of the variance in posttest scores was accounted for in the pretest. The analysis of variance of posttest measures for pitch accuracy skills between groups indicated that the pretest score accounted for less than two percent of variance in posttest scores. Analysis of covariance was computed for pitch accuracy measures between males and females and results indicated to the investigator that the pretest accounted for 58.40 percent of the variance in the posttest scores (62).

The investigator summarized that although all of the groups showed improved vocal skills as well as improved attitudes toward singing, there were no significant differences between instructional groups for measures of vocal range, pitch accuracy, or attitudes toward singing measures for subjects who had received breath-management instruction and subjects who had received instruction by means of a traditional song approach. However, the females in both the experimental and control groups significantly matched pitch more accurately than the males in both groups ($p = .01$). The females also significantly demonstrated more positive attitudes toward singing than did the males in

both groups ($p = .02$). The investigator stated that although there was a statistically significant difference between the pitch accuracy skills of males and females, and the attitudes toward singing of males and females, there was no evidence that breath-management instruction, as investigated in the study, significantly affected the singing skills of sixth-grade choral students (67).

During the course of the study several observations were made. A positive attitude change toward singing in all groups was observed regardless of the method of instruction used. However, it was noticed that the group who received the breath-management instruction was more enthusiastic about the warm-up period that the group who received instruction during the warm-up period using the traditional song approach. During the ten-minute period the experimental groups also were observably more attentive, and maintained more correct posture during the core instruction part of each session for a longer period than the group who received the traditional song approach instruction (77).

The investigator made the following observations in regards to the study:

1. Breath-management instruction appeared to be a complex task that may require skill development over time.
2. Breath-management may be so interdependent with students' abilities to coordinate the vocal mechanism, which typically improves with age, that breath-management instruction alone may not produce significant differences between individuals' vocal performance and attitudes toward singing.
3. Students' physiological development in conjunction with increased abilities to coordinate their vocal mechanisms and combined with proper breath management development may result in desired performance qualities that breath management alone cannot accomplish (78).

The investigator made the following recommendations for addition research:

1. Repeat the study with multiple grade levels, and for an extended period of time to maximize the treatment.
2. Study grades four, five, and six for an entire year, using the breath-management instruction employed in the study.
3. Investigate attitudes toward singing as age increases.
4. Concurrently compare singing skills of students at different age levels (78).
5. Investigate objective and scientific information regarding breath-management instruction to refine vocal instruction techniques (79).

CONDITIONING OF A PITCH RESPONSE

Cobes (1969) investigated the use of conditioning of a pitch response using uncertain singers. A total of forty-five fourth grade, fifth grade, and sixth grade subjects were identified as uncertain singers based on a pretest using a strobotuner. To control for the motivational effects of reinforcement, Cobes used a three-group experimental design in which a control group, a "shaped" group, and a "yoked" group were employed (3). In the control group, subjects were reinforced only for correct respones. In the "shaped" group, subjects were reinforced for successive approximations to the correct pitch. In the "yoked" group, the reinforcement was not contingent upon the correctness of the response but upon the correctness of the response of a paired subject in the "shaped" group. For example, if a subject in the "shaped" group was reinforced for correct responses on items "eight," "ten," and "twelve," then the paired subject in the "yoked" group was reinforced for responses on those same three items regardless of whether the responses were correct or incorrect (30).

The specific hypotheses tested were:

1. Performance of uncertain singers in singing the correct pitch is facilitated significantly more by the method of successive approximations than by reinforcement of correct pitch only.
2. Performance of uncertain singers in a successive approximation group should be significantly better at matching pitches than that of their counterparts in a yoked group (3).

The investigator described each of the experimental groups as follows:

> Subjects in Group S were reinforced for successive approximations to the correct pitch. The size of a successive approximation interval was at least a semitone or even greater, depending on the deviance of the subject's response from the stimulus or criterion tone. Thus, the size of the first approximation interval depended on how much a subject departed from the criterion pitch in his initial performance. After the subject emitted three consecutive pitches within the interval, a new approximation was used. Again, as indicated above, the change in intervals was a semitone or more depending on the subject's response. For example, when the experimenter stopped reinforcement for the first approximation, the subject might have changed his response one-half, one, or more semitones in the direction of the criterion tone. After three responses in the new range, the experimenter reinforced the closest of these tones to the criterion pitch. If at any point in the conditioning process, the subject's performance dropped (i.e., four consecutive responses outside of the successive approximation interval), reinforcement was given for the immediate past interval. In the final "approximation" series, reinforcement was given until the subject emitted ten criterion pitches in succession. The acquisition phase was terminated when the subject reached the criterion or when he failed to improve (i.e., make the transition from one approximation to another) (30).
>
> Subjects in Group Y performed the learning task before their yoked partners since each subject in the yoked group was to receive the same number of reinforcements and on the same trials as the subject he was yoked to in Group S. Therefore, reinforcement for a subject in Group Y was not contingent upon the topography of the response but rather on the response given by his counterpart in Group S. Termination in Group Y during the acquisition phase resulted from a lack of improvement. The subjects in Group Y remained in the acquisition phase at least as long as their yoked partners in Group S (31).
>
> Group R was used, to replicate the teaching situation where positive reinforcement is typically given, on a continuous and immediate basis, only for emitting the correct pitch. The subjects in this group were retained in the experimental condition until ten consecutive correct pitches (criterion tone plus or minus twenty cents) were emitted or until the subject emitted at least thirty consecutive unreinforced pitches and showed signs of frustration, e.g., when he

> would no longer respond. At the termination of the acquisition phase of the experiment, each subject in this group was given five to ten reinforced trials in order to alleviate the apparent frustrations experienced by subjects in this treatment group (29).

All subjects remained in the experimental condition until either the criterion of ten consecutive pitches were emitted or the experimenter terminated that part of the experiment. After their respective experimental treatments for emitting pitches, all subjects were given a transfer task of three pitches in addition to the criterion pitch. A relearning task was then administered to all subjects who had reached criterion (23).

The instrument for analyzing the pitch emitted by subjects was a strobotuner. The instrument provided an accurate reading in cents (1/100 of a semi tone) by comparing the frequency to be measured with internal frequency standards based upon the equally tempered musical scale. The procedure for measuring the sharpness or flatness of a tone involved moving the control knob and pointer to the right or left until the mobile pattern on the strobotuner disc was brought to a stand-still. An audio generator was used for presenting the stimulus tone (24).

The stimulus tone was played via a speaker versus headphones so as not to interfere with the subject's hearing of self-vocal performance. Two yellow lights, on the speaker cabinet, which, when lit, were used for reinforcing subjects's responses. At the bottom of the cabinet was a white light which, when illuminated, signaled the subjects to emit the stimulus tone sounded by the audio generator. A wooden screen separated the investigator from the subject in order to avoid facial or other expressive cues that might potentially affect the subject's response. The investigator used a record sheet to record the pitch sung, number of reinforcements, trials on which reinforcements were delivered, and total number of trials for the experiment (26).

Three tests were used in conjunction with the experimental conditions: (1) The pretest; (2) the transfer task; and (3) the relearning task. The purpose of the pretest was to identify uncertain singers. Subjects were seen in an individual session which lasted approximately five minutes. They were first asked to hum or sing a note that was comfortable and which could be emitted without strain. The investigator then determined the "scaled pitch" (a note of the C major scale) of this note with the aid of the audio generator, calibrated to the equal tempered chromatic scale where A = 440. If the tone emitted was not a scaled pitch, the closest scaled pitch was chosen by the investigator. The pretest was comprised of the following pitches in the given order: one step above, two steps above, one step below, and two steps below the pitch initially emitted by the subject (26).

The investigator instructed the subject to listen to a tone that would come from the speaker and as soon as the tone stopped, to emit (using the syllable "lu," "ah," or "hum") the same note into the microphone. Each of the four pretest pitches was then presented to the subject. The audio generator was used to present the pitch and the strobotuner was used to determine the accuracy of the subject's response. If a subject emitted a tone at least a half step sharp or flat, it was counted as one uncertain pitch. The subject was classified as an uncertain singer if at least three uncertain pitches out of four were emitted (28).

The transfer task, which followed the pretest task and a minute rest period, was used to test for transfer effects. It consisted of the same four pitches used in the pretest for any single subject and in the same order. The subject was required to match those pitches

using the same procedure as in the pretest. If a subject emitted a tone at least one half step sharp or flat, it was counted as one uncertain pitch (28).

The relearning task, which followed the transfer task and a minute rest period, was used to test for retention. It consisted of matching, once again, the subject's criterion tone. Only those subjects who had reached criterion in the acquisition phase participated in the relearning task. Those subjects were required to emit the criterion pitch two consecutive times, using the same reinforcement schedule and procedure as in the learning task (28).

All subjects were first given the pretest individually. From three to twelve days later each subject, or uncertain singer, was tested individually in the experimental acquisition phase with directions similar to those given in the pretest. They were asked to listen to the tone and as soon as the tone stopped, to emit the tone they heard, using the syllables "lu," "ah," or "hum," as long as the white signal light on the speaker box was illuminated. Subjects were told that if the investigator felt they were doing a good job, the two yellow lights would flash on and it was up to them to get the lights to flash on as often as possible. After questions were answered, the experiment began. Between trials twenty-five and twenty-six, fifty and fifty-one, seventy-five and seventy-six, there were thirty-second rest periods. Each subject's criterion tone was one step below the tone he had initially emitted in the pretest. The procedural differences for the three groups were in their reinforcement schedules (29).

The investigator reported that in summary, the method of successive approximation, as used in Group S, was superior in conditioning subjects to emit pitches to a yoked procedure (Group Y), where the number of reinforcements was the same as in the shaped

procedure. The method was also reportedly superior to a control procedure (Group R), where reinforcement was given only for the correct response (48).

An analysis of the data revealed unusually low means for all groups and no significant differences between the three groups. Accordingly, this suggested to the investigator that with the limitations of the procedure, only a specific response was conditioned in the shaping group. A comparison of the number of trials to originally learn to emit a given pitch and the number of trials to relearn this pitch revealed that there was a significant reduction in the number of trials to relearn. In this analysis only subjects from the shaping group were employed since only those subjects reached criterion (60).

IMAGERY TRAINING

Kramer (1985) investigated the effects of two different music programs on third-grade and fourth-grade children's ability to match pitches vocally. The purpose of the investigation was to establish a creative teaching method which incorporated auditory, kinesthetic images, visual mediating strategies, and mnemonic association to improve the inaccurate singer's ability to match pitches vocally (19).

The study employed a quasi-experimental design procedure that consisted of (1) an experimental component and (2) a correlational component. The experimental component was used to determine the relative effects of two different music instruction programs on the abilities of third-grade and fourth-grade children to match pitches vocally. The correlational component was used to determine the significance of the relationship between measured singing ability and age, grade level, IQ, measured music aptitude, math, reading, and vocabulary scores. The two instructional programs used

were the regular instruction as outlined in the *New Providence K-5 Music Curriculum Guide* and the *Gould Specialized Program in Singing* (58).

A total of 201 third-grade and fourth-grade students comprised the sample for the study. All students were pretested by means of an abbreviated form of the *Gould Speech and Song Response Test*. A singing score was obtained and the students were classified as having or not having difficulty in matching pitches vocally. Those having difficulty were classified as inaccurate singers and those not having difficulty matching pitches vocally, as accurate singers (59).

Concerning the correlational component, indicators of age, gender, I.Q. (*Otis-Lennon Mental Ability Test, Form J*), musical aptitude (*Gordon Musical Audiation Aptitude Test*), and reading, math, and vocabulary (*Stanford Achievement Tests*), were investigated to find any significant correlation between other skill areas and the ability to match pitches vocally (59).

A total of 201 students from the third-grade and fourth-grade in two public elementary schools originally served as subjects in the study. Each grade level in each of the two schools consisted of two or three separate, self-contained classes. The students were randomly assigned to the control and intervention groups. There was a minimum of five inaccurate singers in each section (61).

Within the study there were four third-grade classes and five fourth-grade classes. When those classes were assigned to intervention and control groups, they made up a total of ten classes. The third grade met once each week for thirty minutes. The fourth grade met once each week for forty-five minutes. The study was conducted for ten consecutive weeks. During week one, teacher A, a vocal music instructor, taught section

number one, the control groups. During week one, teacher B, also a vocal music teacher, taught section number two, the intervention groups. During each succeeding week, teachers A and B alternated the sections to be taught so that both instructed the control and intervention groups an equal number of times. A log was kept by each teacher (63). Training was provided to each teacher prior to the study in the area of the inaccurate singer: (1) Definition; (2) identification; (3) classification; and (4) imagery strategies (65).

The control groups, comprised of two third-grade classes and three fourth-grade classes (N = 98), received the regular instruction as outlined in the *New Providence K-5 Curriculum Guide*. Songs were taken from the primary text, *Making Music Your Own*, by Silver Burdett Publishing Company, and supplementary materials as listed in the guide. Singing was accompanied by the teacher on the piano, with recordings, or by the teachers singing along with the students. Songs included a wide variety of keys, major and minor as well as pentatonic tonalities, intervalic complexities, and extended vocal ranges. Words of the songs rather than neutral vowels were used in singing (66).

Student participation in the varied activities of the guide included reading, writing, and singing. Physical movement in response to the songs was limited to clapping and some bodily movement within the room. Spatial awareness related to the musical notes, voice, or text was not often promoted (66).

The intervention group, which consisted of two third-grade groups and three fourth-grade groups (N = 103), received activities from *Gould's Specialized Program in*

Singing. The goal of the program was to develop the student's singing ability to match pitches correctly through music skill awareness and practice. Activities in this group included:

1. Awareness exercises from the first five games of *Put Your Mother on the Ceiling* by DeMille (1967).
2. Supplementary music/image modality strategies that included physical, visual, and vocal tracing of a rainbow; the imitation of the *whoo* sound of a ghost with the mental image of the ghost flying, and the kinesthetic and visual sensations of tracing high and low note progression with accompanying arm and hand motion formed an important portion of the intervention method (67).
3. Singing in the Gould program was preceded by speech-to-song materials and exercises. Songs used in the study were taken from the primary text, *Making Music Your Own* by Landeck, Crook, and Youngbery (1971). Secondary texts used were *Music Through the Year, Grade 3,* and *Music Across Our Country, Grade 4* by Wolfe, Krone, and Fullerton, 1960. Singing was typically performed a cappella and, when infrequently accompanied, done so only by means of chords rather than through melodic replication.
4. The teacher gave beginning pitches by voice or pitch pipe sounds.
5. Musical examples were given by means of the singing voice.
6. Vocal ranges were limited to C4 to D5, encompassing the interval of a seventh above C4.
7. Songs sung were performed on the neutral vowel sound "oo" or "loo" until the melodic material was mastered. Once melodically correct, the words of the text were introduced.
8. In addition to singing strategies, vocal and physical relaxation exercises were used to start each lesson.
9. Imagery games from *Put Your Mother on the Ceiling* by DeMille, 1967, were also used and were included in seventy-five percent of the lesson.
10. In order to correlate music and imagery, activities through extensive bodily movement, imagination games, charts, and pictures were also incorporated (68).

Two dependent variables were assessed as outcomes of the treatment:

1. Individual vocal tone matching ability.
2. Imaging ability (68).

In addition, correlation significance between measured singing ability and age, grade level, I.Q., measured music aptitude, math, reading, and vocabulary scores was assessed.

Two measures of imagery were used both before and after instruction. *A Shortened Form of the Betts Questionnaire Upon Mental Imagery* by Sheehan (1967) was used to define the degree of image vividness that each student possessed. To determine the students' degree of image controllability, the *Gordon Test of Visual Imagery Control* was used (68).

Tests in tone-matching ability were administered to all students in the third-grade and fourth-grade. An abbreviated form of the *Gould Speech and Song Response Test 1 and 2*, was used for the purpose of identifying the inaccurate singer and for the purpose of measuring the vocal ability of each student before and after the experiment. This second test was given by a taped recording with the same male voice as in the first tape recording. However, in this test the pitches to be replicated were sung an octave higher, using the male falsetto voice. The additional musical measure incorporated was the *Gordon Musical Audiation Test*. This test was used as a means of obtaining an indicator of music aptitude possessed by each student (69).

The Gould test was formulated for the purpose of determining the ability of a child to match and remember pitches. The Gould test was administered by a music specialist on an individual student basis. Tape recordings of all individual testing were made and the recordings spot-checked by a music specialist for correct evaluation by the test instructor (70).

In order to make the time management for administering the pretest and posttest of the *Gould Speech and Song Response Test* practical, an abbreviated form of this response test

was used. The abbreviated form of the response test was determined by proportionally reducing the total number of points in each category of the original test (70).

The *Gould Speech and Song Test* (Gould, 1968) was given to 103 subjects in grades one, two, and three in the Western Illinois University Campus School. The basic purpose of the test was to test for singing accuracy. Reliability coefficients and standard errors for the test were computed for each grade and for all grades to determine the extent to which the test was internally consistent. The reliability coefficients for the Gould test were as follows:

1. Grade One: $r = .92$
2. Grade Two: $r = .95$
3. Grade Three: $r = .96$
4. All Grades: $r = .94$ (71).

Validity coefficients were computed for the test by correlating the gain scores obtained by each pupil, posttest score minus pretest score, with the improvement scores assigned the pupils by a panel of judges from before and after tapes of their singing voices. Validity coefficients obtained from the judges' scores from the original test administration were:

1. Grade One: $r = .51$
2. Grade Two: $r = .89$
3. Grade Three: $r = .67$
4. All Grades: $r = .66$ (71).

The data needed to make an assessment of image vividness was obtained through a self-reporting questionnaire completed by each individual student. The questionnaires were identical and contained thirty-five questions dealing with the visual, auditory, tactile, kinesthetic, olfactory, gustatory, and organic imagery modalities. *Sheehan's QMI*

was a shortened version of the *Betts Questionnaire Upon Mental Imagery* of 1909. The condensed form of the *Betts QMI* by Sheehan (1967) shortened the original test to thirty-five items to be rated on a scale of "one" to "seven." This scale rated "one" as being "perfectly clear and as vivid as the actual experience," with "seven" being identified with "no image present at all" (73).

Factor analysis had been used as the major means of determining validity. Extensive psychometric analysis of a S*hortened Form of the Betts Questionnaire Upon Mental Imagery* demonstrated that it measured a general ability to image. A cross-validation between the long and short forms of the *Betts QMI* and *Sheehan QMI* showed a correlation of $r = .99$ (74).

As in the *Sheehan QMI*, the *Gordon Test of Visual Imagery Control* was administered to each self-contained class as a unit. The researcher administered the test to all classes in an identical manner. The three-part system of answering was defined and then each question was read aloud to the students. The students circled their selected answer. All students progressed at an equal rate of time though the questionnaire (76).

Other indicators obtained for the correlational aspects of this study included I.Q., the *Otis-Lennon Mental Ability Test, Form J*, gender, age, grade, and reading, math, and vocabulary scores (*Stanford Achievement Tests*) (76). To determine if there was a significant difference in vocal ability as measured by the abbreviated form of the *Gould Speech and Song Response Test 1 and 2*, a series of one-way analysis of covariance were applied to each student (81).

The following findings were reported by the investigator:

1. The initial data received from the prevocal test one, indicated a total of 102 students (51%) classified as inaccurate singers. The percentages of inaccurate singers were reported as being evenly divided between the third and fourth grades.
2. The findings of the pretest two indicated that there were sixty-one students (30%) categorized as inaccurate singers. The percentage of inaccurate third-grade subjects was 30.29%. The percentage of inaccurate fourth-grade subjects was 29.52 (81).
3. For the pretest/posttest one and two, the treatment was highly effective in improving the child's ability to match pitches vocally. Statistical analyses revealed significant differences between the treatment and control groups at (p=.01) (84).
4. Examination of the means for the entire group of students in the vocal pretest and vocal posttest also revealed an improved vocal ability over the ten-week period of intervention. In the analysis of means, it could be seen that the treatment effect was slightly stronger for the fourth-graders. For third-graders, it was seen that the treatment effect was not statistically significant.
5. Statistical analysis of the vocal test two also revealed over a ten-week period, an increase of +5.10 points in improved vocal ability for the entire group of students. The comparison of means for the same test for grade three denoted an increase of +1.97, suggesting an improved level of vocal ability notably lower than the group mean. The mean for the fourth-grade, however, was M = 7.87, and suggested a stronger level of vocal improvement over the group mean.
6. In comparing the inaccurate singer segment class to the group mean and to each other, data suggested that the intervention method did improve the ability of the inaccurate singer to match pitches vocally (87).
7. Of the twenty-five subjects in the inaccurate singer segment within the control group, data demonstrated an increase in ability of +9.76.
8. Of the thirty-six subjects in the inaccurate singer segment within the intervention group, data suggested a high level of improvement in the vocal ability of the inaccurate singer segment within the intervention group. It was also noted that every student in the inaccurate singer intervention group increased his or her individual score from the vocal pretest to the vocal posttest (90).
9. The investigator reported that the ability to produce genuine sensory or perceptual experiences in the absence of the original stimulus condition could be measured in degrees of vividness. In this regard, the examination of the data indicated that there was no treatment effect with respect to the QMI.
10. Examination of the means indicated that the fourth-grade students had demonstrated an increase in imagery ability.

11. The means for the entire group indicated an increased score for grade three, for the control group, and for the inaccurate singer segments within the control and intervention groups. The means also indicated a decrease for grade four, and the intervention group (92).
12. The measurement of the degree of mental control a subject had over the manipulation of an image was administered to determine the degree of controllability a student had over his or her visual imagery. The examination of the data indicated that the treatment and grade level were not significant.
13. Statistical analysis revealed that the treatment and grade were not significant, but the interaction of treatment and grade was significant. The only decrease of mean score was indicated for the inaccurate singer segment within the intervention group.
14. With regard to image vividness, students generally reported scores indicating a high degree of image vividness ability on the pretest. The findings showed that the treatment was not having an effect with respect to the QMI.
15. Concerning image controllability, significance was not found on the treatment or grade level. There was significance in the interaction of treatment and grade level. The control group grade three was found to be significantly different from control group grade four and highly significantly different from the intervention group grade three. This indicated that, as related to controllability, the treatment had a significant effect on third graders. The inaccurate singer segments within the control group demonstrated decreases in controllability while the inaccurate singer segment within the intervention group demonstrated an increase in controllability. Within the inaccurate singer segments, the increase in controllability was due to the fourth-grade students while the decrease in ability was due to the third-grade students (96).
16. The correlational comparison component included indicators of age, gender, grade level, I.Q. and the Stanford Achievement Test scores for reading, math, and vocabulary. The examination of the data indicated a significant correlation between I.Q. and music aptitude, and a highly significant correlation between I.Q, and the prevocal test one. There was not, however, a significant relation between I.Q. and the vocal pretest two scores.
17. Demographic information indicated that the mean I.Q. for all subjects was not affected by their placement into treatment groups. The data also revealed a highly significant correlation between I.Q. and the *Stanford Achievement Tests* (97).
18. Statistical analysis indicated that the significant correlation between I.Q. and music aptitude was found for tonal ability and rhythmic ability. The highly significant correlation between I.Q. and prevocal one was noted. The lack of significance between I.Q. and prevocal two was noted.
19. The examination of data with regard to music aptitude was said to be of considerable importance. In the comparison of differences between

groups, the means for the tonal and rhythmic tests were significantly lower for both the inaccurate singer classes in the control and treatment groups when compared to the tonal and rhythmic means of the overall group (100).

20. It was also noted that the relationship between the tonal and rhythmic ability scores demonstrated a highly significant correlation.
21. A significant relationship between the tonal and rhythmic scores and the post-vocal scores was found. This correlation was highly significant of the control group and significant for the intervention group.
22. In the examination of data in the comparison of gender and age with the areas included in the study, gender was not revealed as being significant in any of the comparisons except in the aspect of image controllability (101).
23. Visual preimage scores had significant correlation with rhythm while the data of the postimage scores indicated significantly negative correlations with age and grade. In the precontrollability and postcontrollability image test scores, there was a significant correlation between rhythm. All other correlations were significantly negative with age, grade, and math. Improvement in controllability did show a significant correlation to gender. Statistical analysis revealed a significant correlation between rhythm and controllability (102).

The researcher made the following conclusions based on the study:

1. Environmental forces in the form of music education programs can be effective in teaching students how to sing.
2. Creative methods of singing that include direct and indirect image strategies provide the child with associative functions that increase learning comprehension (123).
3. The *Gould Specialized Program in Singing* that included image strategies was effective in increasing the ability of third-grade and fourth-grade inaccurate singers to match pitches vocally.
4. The intervention method did affect patterns of learning and comprehension for the elementary student.
5. Results of the study suggested that, in part, the use of image strategies did provide an alternate means to understanding when there seemed to be a verbal mediating deficiency demonstrated by the child (124).

The researcher suggested that recommendations for future research might:

1. Design creative methods of teaching singing that include image strategies.
2. Factor out the variables in various singing methods that are effective in assisting the inaccurate singer.
3. Explore how method variables affect the development of singing ability.
4. Investigate and increase understanding of how children learn.
5. Develop an instrument to measure the degree of paired-associate and mnemonic use in individual learning methods.

6. Repeat the study on various elementary grade levels.
7. Further investigate the effects of verbal mediating deficiency in children's learning.
8. Investigate how children that are not inaccurate singers seemingly develop their ability to sing naturally (126).

MELODIC ECHO TRAINING

Stauffer (1986) investigated the effects of melodic and harmonic context provided during systematic melodic echo training on the development of singing skills and aural discrimination abilities of children in the primary grades. The four models of training included: (1) no context; (2) melodic context; (3) harmonic context; and (4) both melodic and harmonic contexts (7).

Secondary purposes of the study were: (1) to prepare a systematic melodic echo training program and produce the program on tape with the appropriate treatment modes; (2) to develop measurement instruments which collect quantitative data on the singing voices of the subjects; (3) to study the effects of the treatment modes on the singing skills and aural discrimination abilities of the subjects; (4) to investigate the relationships between singing skills and aural discrimination abilities; and (5) to investigate the relationships of selected demographic control variables, grade, gender, previous musical experience, attitude, and interest in musical activities, to the singing skills and aural discrimination abilities of the subjects (7).

The study was conducted in the Hartland School District, Harland, Michigan, during the fall of 1984. The procedures of the study were to (1) collect demographic data and reading ability data; (2) collect pretest data regarding the singing skills and aural discrimination abilities of the subjects prior to instruction; (3) administer an instructional program designed to improve singing skills and develop aural discrimination abilities;

(4) collect posttest data regarding the singing skills and aural discrimination abilities of the subjects following instruction; and (5) analyze data to determine the effectiveness of the instructional program and identify factors which may contribute to the positive growth of singing skills and aural discrimination ability (58).

Two standardized measures were chosen for the study: (1) *Primary Measures of Music Audiation* and (2) *Wide Range Achievement Test-Revised.* Two further data collection instruments were designed by the author: (1) *Melodic Echo Test* and (2) *Test of Singing Ability*. In addition, an existing survey instrument was revised to collect information regarding musical experience, interests, and attitudes of the subjects—*How I Feel About Music* survey (60).

The *Primary Measures of Music Audiation (*PMMA*)* was a battery of tests designed to measure the developing musical aptitude of children in kindergarten through third-grade. The battery was composed of two tape-recorded subtests—Tonal and Rhythm. In both subtests, the subjects listened to two musical patterns and identified them as "same" or "different" by circling pictures of same or different faces on an answer sheet (60).

The PMMA manual included essential reliability information for both subtests and for a composite score derived from the subtest scores. Split half reliability coefficients for the subtests ranged from $r = .72$ to $r = .92$ for children in kindergarten through third grade. The range of reliability coefficients was $r = .85$ to $r = .92$ when kindergarten was eliminated. The test-retest coefficients ranged from $r = .60$ to $r = .76$ when kindergarten coefficients were included, and from $r = .66$ to $r = .76$ when kindergarten coefficients were not included (61).

The PMMA battery was employed in the study as both a pretest and posttest measure. Results were used to examine the effects of the instructional program and the harmonic and melodic contexts provided therein on the singing skills and aural discrimination abilities of the subjects (61).

The *Wide Range Achievement Test – Revised* (WRAT-R) was an academic achievement instrument designed to study the development of reading, spelling, and arithmetic codes. The WRAT-R was standardized using a national sampling model stratified by five factors—age, geographic region, gender, race, and metro/non-metro residence (62).

The reading subtest was used for the study. The test measured the subject's achievement in recognizing and naming letters, and recognizing printed words. The test was administered individually. Four types of scores were derived from the WRAT-R: (1) raw score, (2) standard score, (3) percentile rank, and (4) grade level equivalent (62).

Content validity for the WRAT-R was determined by item separation analysis and was found to be $r = .99$. Reliability was established through the test-retest method. For a group of children ages seven to seven-and-a-half and ten to ten-and-a-half years of age, the reliability coefficient was $r = .96$ (63).

The WRAT-R reading subtest was used during the pretest phase of the study in order to determine if any academic differences existed between groups. The results were also used to determine if a relationship existed between academic achievement of the subjects and the effects of the instructional program (63).

The *How I Feel About Music Survey* was an attitude and interest inventory designed to collect information about the musical background of the individual, including music in the home, performance and/or instructional experience, and preference for musical

activities (63). Subject responses were used to determine if experience, attitude, and interest in musical activities were related to the singing skills and aural discrimination abilities of the subjects (64).

The *Melodic Echo Test* (MET) was an instrument designed by the author to measure the ability to imitate melodic patterns by singing. The test was piloted the summer prior to the study and necessary adjustments were made. The test consisted of twenty prerecorded melodic patterns which a subject listened to and then imitated. Each pattern was heard only once. Patterns consisted of three, four, or five pitches within a four-beat rhythmic framework. The subject was allowed four beats in which to initiate a response. The test included patterns in both major and minor tonalities, and various combinations of ascending and descending melodic steps and skips. The test was administered individually. Administration time for each individual was approximately four minutes. The subject's responses were tape recorded and evaluated at a later time by judges working independently (65).

The *Test of Singing Ability* (TSA) was an instrument designed by the author to diagnose and measure the singing ability of the subjects participating in the study. The TSA required each subject to sing portions of the melody of five familiar songs. The test was administered to individual subjects in a five-minute time interval. Test responses were tape recorded and evaluated later by judges working independently. The TSA was also tested and adjusted accordingly in a pilot study (76).

The children participating in the study had neither previous nor supplementary music instruction through the schools (80). The subjects for the study were the children attending the first, second, and third grades of two public schools in Hartland, Michigan.

Four classes of each grade level were used to investigate all possible treatment combinations of the study. Heterogenous, intact classes were used (81).

The treatment effects were the addition of melodic and/or harmonic context to the basic sequences described above. Melodic context was achieved by providing a replication of each pattern during the student response measures. Thus, the subjects heard the pattern again while they were singing the response. A synthesized sound was used to record the melodic context treatment. Harmonic context was provided by adding a synthesized harmonic background to the entire sequence. Thus, the four treatment contexts derived were: (1) no context; (2) melodic context; (3) harmonic context; and (4) both melodic and harmonic contexts (84).

All instructional tapes were recorded by the author and an assistant at the University of Michigan. The author's voice was used for the sequence stimulus patterns. The synthesized sound for the melodic context was provided by an *Alpha Syntauri* system. The harmonic and rhythmic backgrounds were adapted from pre-recorded tracks (84).

The treatments were administered during a twelve-week instructional period in the fall of 1984. At the beginning of each week, the author presented a twenty-five minute general music lesson in each class. The lesson consisted of four components: (1) a rhythmic movement sequence and/or rhythmic speaking activity; (2) an instrument demonstration; (3) review of one taped melodic echo sequence set and introduction of a new sequence set; and (4) a song activity. The lessons were identical for each class by grade level except for the context provided during the tape-recorded melodic echo sequences (88).

In addition to the instruction provided by the author during the class lessons, the classroom teachers were provided with copies of the treatment tapes which they played twice during the week for the subjects. The teachers participated in a demonstration/training session prior to the start of the instructional period, and were given specific instruction in leading the student singing without using their own voices. In addition, the teachers were provided with log books in which they recorded (1) the days and time of day during which the tapes were played, (2) the equipment used to play the tapes, and (3) comments and student reactions (88).

During the first week of school, a consent letter and the *How I Feel About Music Survey* were sent to the parents of every student in the first, second, and third grades. The teacher orientation/instruction session was held, and log books were distributed (89).

At the end of the first week of school, the PMMA was administered as a pretest to all groups and administration of the WRAT-R was started. The first week of instruction was followed immediately by pretest administration of the MET and TSA. The remaining eleven weeks of instruction followed without incident. Immediately following the twelve weeks of instruction, posttests (PMMA, MET, TSA) were administered (89).

The following results were reported based on the data:

1. Posttest mean scores for the Tonal subtest of the PMMA were compared through analysis of variance. The results revealed no significant differences in aural discrimination as measured by the Tonal subtest between the four main treatment groups. When treatment groups within each grade level were considered, no significant differences were found. Similarly, multivariate analysis of variance revealed no significant main effects or interactions for the melody or harmony treatments. The null hypothesis was retained. It was noted, however, that the mean posttest scores were significantly higher than the mean pretest scores for every main treatment group and for the majority of the within-grade-level groups (211).

2. Posttest mean scores for the MET were compared through analysis of variance. The results revealed no significant differences in the ability to imitate melodic patterns between the four main treatment groups. Similarly, when treatment groups within each grade level were considered, no significant differences were found in posttest mean scores. The null hypothesis was retained.
3. A significant interaction effect of harmony and grade was identified in the multivariate analysis of covariance of MET scores. An interaction effect of melody, harmony, and grade was also found. Further analysis revealed that an interaction between melody and harmony was evident for third-grade subjects, but not for first-grade and second-grade subjects (212).
4. As in the PMMA Tonal subtest, it was noted that mean posttest scores were significantly higher than mean MET pretest scores for the four main treatment groups and for two-thirds of the groups within grade levels.
5. Posttest mean scores for the TSA were compared through analysis of variance. Results revealed no significant differences in singing ability as measured by the TSA between the four main treatment groups. Analysis of covariance for each grade level also revealed no significant differences between treatment groups. Similarly, multivariate analyses of variance and covariance revealed no significant main effects or interactions for the melody and harmony treatments. The null hypothesis was retained. Again, TSA posttest scores were significantly higher than TSA pretest scores for most treatment groups.
6. Posttest scores for the Tonal subtest of the PMMA, the MET, and the TSA for girls and boys were compared using student *t* tests to determine differences between girls and boys in the same treatment groups. No significant differences were found between boys and girls in aural discrimination ability within any treatment group. Significant differences were found in posttest scores of both MET and TSA for both the "melodic" and "both contexts" groups. All significant results were reportedly in favor of girls (213).
7. Correlations computed for pretests and posttests on all dependent variable measures indicated that the ability to discriminate tonal patterns aurally, the ability to echo melodic patterns, and the ability to sing rote songs were strongly related. Some relationships were also found between the pretest and posttest measures and the reading achievement scores of the subjects. There were few relationships between the "Experience," "Attitude," and "Interest" scores of the subjects and the musical skills investigated in the study.
8. In every dependent measure (PMMA, MET, TSA), significant differences were found between the musical abilities of grade level groups prior to treatment. Differences favored older children, and were greatest between first-grade and third-grade subjects. Following treatment, grade level differences were either less significant or not evident. Posttest mean scores were higher than pretest mean scores on all measures for all grade level groups, indicating that subjects in all grades improved in musical

abilities during the instructional period. Evidence suggested that progress was more rapid for younger children than for older children (214).

Based on the study, the following conclusions were drawn by the investigator:

1. The ability to discriminate tonal patterns aurally was facilitated by instruction regardless of context provided during systematic instruction based on imitation of melodic echo patterns.
2. The ability to imitate melodic patterns vocally was facilitated by instruction regardless of context provided during systematic melodic echo training. Older students benefited more from harmonic context than younger students, and that may be related to the development of harmonic discrimination skills.
3. Singing ability improved with training regardless of context provided during systematic instruction based on melodic echo patterns. There was evidence indicating that instruction may negate initial differences in singing ability.
4. There was evidence to suggest that the ability to discriminate tonal patterns aurally, the ability to echo melodic patterns vocally, and ability to sing rote songs were strongly related.
5. There was a tendency for girls to score higher than boys on singing tasks, but not on aural discrimination tasks which did not require a vocal response.
6. There was evidence to suggest that some relationship existed between musical skills and reading achievement in primary grade children.
7. There was little evidence of relationships between musical skills of primary grade children and their previous musical experience, attitudes about music, and interests in musical activities.
8. Musical skills developed with maturation; for children in the primary grades, the development of musical skills was more rapid in the first grade than in the third grade (215).

The investigator made the following recommendations for further research:

1. Although aural discrimination of tonal patterns and singing skills improved for most subjects in the study, it should be determined whether the improvement was a result of (a) systematic melodic echo training; (b) general music instruction provided to subjects who had no previous instruction; (c) maturation; or (d) a combination of those variables.
2. *The Melodic Echo Test* and *Test of Singing Ability* used in the study required students to sing a capella. It was possible that students who received training with a context were confused by the absence of that context in a testing situation. A study is needed in which the effects of context are measured in testing and training situations.

3. *The Test of Singing Ability* was composed of songs that most subjects knew prior to instruction. There may have been a tendency for subjects to sing previously known songs as they were learned regardless of intervening instruction that seeks to improve singing ability. A more accurate measure of singing ability may be obtained in a similar study if singing ability following instruction is measured using songs taught during the training period.
4. There were no attempts to address individual singing problems or to give instructions regarding vocal production during the twelve-week instructional program. Individual attention and specific instruction in good vocal production may further facilitate accurate singing. A study is needed which examines these variables.
5. A longitudinal study is needed to investigate the effects of melodic pattern training on the singing skills and aural discrimination abilities of primary grade children.
6. An item analysis of the PMMA Tonal subtest and the *Melodic Echo Test* used in the study might provide further insight regarding the relationship of aural musical skills and musical production skills (216).

PENTATONIC AND/OR DIATONIC PITCH PATTERN INSTRUCTION

Jarjisian (1981) investigated: (1) the comparative effects of instruction in: (a) pitch patterns which include half steps, (b) pitch patterns which include no half steps, and (c) pitch patterns of both types on diatonic and pentatonic rote singing achievement; (2) the effects of tonal aptitude on diatonic and pentatonic rote singing achievement; and (3) the effects of teacher, school environment, and socioeconomic status on diatonic and pentatonic rote singing achievement (9).

The students who participated in the study were drawn from two Philadelphia elementary schools. The students were enrolled in three first-grade classes in a black public school and three first grade classes in a white archdiocesan school (19).

The Tonal test of the *Primary Measures of Music Audiation* (PMMA) was administered in January of 1980 to all students. The students received regular music instruction for a period of sixteen weeks. The regularly employed music specialist taught the public

school students once weekly for a period of forty-five minutes; the investigator, also an elementary music specialist, taught the parochial school students twice weekly in thirty-minute periods (19).

Each class in each school was randomly designated Experimental Group T1, T2, or T3, and all students received musical instruction for the duration of the instructional segment of sixteen weeks. All experimental groups received pitch pattern instruction, ten to fourteen minutes weekly, which consisted of students echoing the teacher's a capella singing of the patterns on neutral syllables or with syllable names. The patterns were echoed in large and small groups and individually. Group T1 received instruction in diatonic patterns selected on the basis of range and size of intervals from Gordon's taxonomy (Gordon, 1980; 142-160). Group T2 received instruction in four tone pentatonic patterns derived from the sequence employed by proponents of the Kodaly approach. Group T3 received instruction in both diatonic and pentatonic patterns, their patterns comprising about half those of each of the other two groups while spanning the same melodic content (20).

In addition, all students learned by rote the same two songs weekly; one song was diatonic, the other pentatonic. Rhythmic activities, movement experiences, and listening opportunities were determined by the individual teacher and were identical for all three classes within each school (20).

During the third month of instruction, four criterion songs were taught to all participating students and were sung approximately the same number of times until the end of the instructional segment. Opportunities for individual students to sing alone were provided within the class setting. Two songs were pentatonic, one do-based and one la-

based; two songs were diatonic, one harmonic minor, and one major. The songs were selected for variety of meter and for patterns which incorporated various stepwise and skipping melodic configurations (21).

At the conclusion of the instructional segment, each participant was individually recorded singing the criterion songs. Before the recording session, the investigator visited each class, sang each of the criterion songs with the entire group, and explained the recording procedure to the children (22).

For each student, the investigator randomly determined the initial song to be sung. The investigator played a tonic triad on a fixed bell set and sang the starting pitch on "Ready, sing" in an appropriate meter and tempo. The songs were sung a cappella, as they had been during the instructional segment (22).

A five-point rating scale was employed by two judges, the investigator and another music educator, to evaluate the performances of the criterion songs. The judges rated the performances independently of each other. The rating scale was as follows:

1 = Use of singing voice.
2 = Maintenance of pitch center or general sense of melodic direction.
3 = Maintenance of pitch center and general sense of melodic direction.
4 = Accuracy in singing adjacent intervals or leaps.
5 = Accuracy in singing adjacent intervals and leaps (22).

Pearson product moment correlation coefficients were obtained to examine the interjudge reliability of the ratings for the songs. Seven three-way analyses of variance were conducted to investigate the problems stated earlier. For each analysis there were eight observations in each cell of the design, a number obtained through a process of random elimination to secure equality among cells. The "A" factor represented the three instructional treatments; the "B" factor represented high and low aptitude levels, the

fiftieth percentile on the Tonal test of PMMA being the dividing point; and the "C" factor represented the two school populations. A separate analysis was conducted for each of the four songs. Additional analyses were made of the pair of pentatonic songs, the pair of diatonic songs, and all four songs combined (23).

The investigator reported that the significant main effect for tonal aptitude was a predictable finding. It was found that high aptitude students performed significantly better than low aptitude students. One unexpected finding was the lack of significance for the "C" or school factor. The absence of a school main effect was particularly interesting to the investigator because of the seemingly large differences between the two schools. It was reported that there were no two-way or three-way interactions. The investigator explained that the absence of any interaction was an indication that the treatment effect was a strong one; that is, a combination of diatonic and pentatonic pitch pattern instruction was superior to either diatonic or pentatonic pitch pattern instruction regardless of tonal aptitude level or circumstances surrounding instruction (41).

With regard to the five-point rating scale employed, the investigator said it should be noted that a student earned a score of "two" if he maintained pitch center or a general sense of melodic direction and a "three" if he maintained both. It was explained by the investigator that because some children sing with direction before maintaing a pitch center and others develop the abilities in reverse order, it was not possible to place those achievements in the rating scale in a rank order. The investigator said therefore, the particular rating scale used did not allow the judges to distinguish between those students who maintained pitch center but not direction and those who maintained direction but not pitch center. Other questions raised by the investigator were whether significant

differences might have been found between the diatonic treatment group and the pentatonic treatment group if those two abilities had been separated for rating purposes (42).

The investigator made the following conclusions based on the study:

1. Young children's rote singing achievement is benefited by pitch pattern instruction which includes both diatonic and pentatonic patterns.
2. Instructional methods such as Kodaly's and Orff's in which pentatonic pitch patterns form the exclusive pitch pattern content in the beginning years of instruction may actually decrease children's potential for rote singing achievement.
3. Students at any tonal aptitude level can benefit from receiving a combination of diatonic and pentatonic pitch pattern instruction.
4. Instructional content and tonal aptitude are more powerful influences on rote singing achievement than teacher, school environment, or socioeconomic status.
5. Rote singing achievement can be expected to be responsive to the proper pitch pattern instruction regardless of the amount and frequency of music instruction or the socioeconomic backgrounds of the students (46).

PIANO HARMONIC ACCOMPANIMENT

A study by Atterbury and Silcox (1993) investigated the influence of piano harmonic accompaniment on the developmental singing ability of kindergarten students during a year of music instruction. The study used a pretest-posttest control group design with an additional posttest measure. The additional posttest was Gordon's (1979) *Primary Measures of Music Audiation* (41).

All kindergartners attending the *Kindergarten Center* in Gorham, Maine during the 1990-91 school year were the subjects of the study. Classes were randomly assigned to either the experimental (no piano accompaniment) or the control condition (piano accompaniment). Seven classes were designated as experimental, and eight classes were the control group. In the middle of the school year, the time of instruction--morning or

afternoon, was reversed, giving all subjects an equal amount of morning and afternoon instruction. There were ninety-six subjects in the experimental group and 109 subjects in the control group (42).

A pretest-posttest control group design was used with an additional posttest measure. During the first three weeks of school, all kindergartners were taught the test song *Pinto Pony*. The song contained four phrases, AABA, and had a range from D4 to A4 above middle C4. All classes were taught the song by rote during the first three weeks of school, and each child individually sang the song into a cassette recorder during music class in the first week in October as a pretest. The same song was sung again by each child during the first week of June, as the posttest. Gordon's (1979) *Primary Measures of Music Audiation* (PMMA) was administered to intact classes during the last two weeks of May (42).

The same lesson plan for the weekly thirty-minute periods was used for all classes with the exception of piano accompaniment during all singing. All classes were taught by the same teacher. Weekly instruction included ten to twelve minutes of activities to promote singing voice development, incorporating strategies such as echoes, singing games, and speech-to-song contrasts found in the current methods textbooks of the time. In addition to singing, instruction also incorporated experience with classroom percussion instruments, movement, and song stories. Instruction throughout the year was conceptually based, and there was considerable emphasis on steady beat. The concepts of register, duration, dynamics, and form were also emphasized (43).

The taped singing of the song *Pinto Pony* by all subjects was evaluated using the following rating scale:

1 = Presinger: Does not sing but chants the song text;
2 = Uncertain singer: Sustains tones, uses both speaking and singing voice, when singing uses a limited range of about a third;
3 = Partial singer: Sings some phrases correctly but not entire song;
4 = Singer: Sings entire song correctly in one key (43).

The scale was adapted from a five-point scale developed by Rutkowski (1986). A pilot study was conducted in which interrater reliabilty was established using two judges who were experienced elementary music educators. Interrater correlation for the pretest was $r = .75$, and for the posttest was $r = .86$ (43).

The PMMA was administered in the spring (43). PMMA scores were analyzed both as composite scores and as an aptitude group (high, average, low) score. The aptitude group scores were determined as follows: 1 = high aptitude (raw score sixty-one and above); 2 = average aptitude (raw score forty-three to sixty); 3 = low aptitude (raw score below forty-two) (43).

The posttest means were compared through a two-way analysis of covariance (ANCOVA), adjusted for unequal group size, using the pretest means as the covariates. The ANCOVA was used to control for initial differences observed between group means. No significant difference ($p = .59$) was observed between the groups in whole-song reproduction after a year of instruction (43).

The difference between posttest means for each PMMA group was compared through the analysis of covariance already mentioned. A significant difference in posttest means ($p < .01$) was found among the means of the aptitude groups. Further analyses using the

Duncan Multiple Range Test showed one group, that of high-ability students, to be significantly different from the other two groups (43).

The PMMA composite scores for each group were compared to determine if there were differences between the two groups following a year during which one group had unaccompanied singing instruction. The results of the analysis indicated no significance (44).

The researchers summarized that:

1. No significant differences existed in singing ability between one group of kindergartners who had piano harmonic accompaniment and one group with no accompaniment during one year of instruction.
2. Posttest song scores of high-aptitude subjects were significantly higher than either average-aptitude or low-aptitude subjects.
3. No significant differences existed in the composite scores of the PMMA between the experimental and control group (45).

The researchers made the following suggestions for further research:

1. Future investigators should expand the scale used in the study to give additional specificity. It was hypothesized that the narrow range (one to four) of the rating scale may have been responsible for the lack of difference observed in the statistical analysis. A more accurate scoring system might focus on selected phrases within the song instead of making an overall judgment regarding the entire song.
2. More extensive exposure (more than thirty minutes a week) and/or extended exposure throughout the grade levels is necessary for young children to significantly improve their singing ability (45).

PITCH ACCURACY AND HEAD VOICE GROUP INSTRUCTION

Wyatt (1993) investigated the effect of class instruction in singing pitches accurately combined with instruction in the use of head voice on the intonation and functional singing ranges of second-grade children. It was not the intent of the study to isolate one technique, but rather to use a combination of two approaches: exercises to encourage

singing in-tune combined with exercises to encourage the development of the use of the lighter vocal mechanism, the head voice (10).

Specific questions addressed included:

1. Will there be a significant difference in the pitch accuracy of children as a result of group training in pitch accuracy and in head voice, over a period of one semester, when compared to control groups whose training does not focus on pitch accuracy and head voice?
2. Will there be a significant difference in the functional range of children as a result of group training in pitch accuracy and head voice, over a period of one semester, when compared to control groups whose training does not focus on pitch accuracy and head voice? (10)

Subjects (N = 120) were children from the second-grade classes of a public elementary school in Columbia, South Carolina. From the five second-grade classes, two were randomly chosen to be the experimental group; of the three remaining classes, two were the control group, and one was randomly chosen to be a second control group (46).

A pretest/posttest was designed by the researcher, judges were selected, and pilot studies were set up. Judges were all experienced music teachers, well acquainted with children's voices. High interjudge correlation was obtained (r = .88 to r = .94) (46).

The purpose of the pretest/posttest was to measure each child's pitch accuracy in an individually recorded performance, and to identify each child's comfortable singing range (46).

For the pretest/posttest, each child sang a familiar song, beginning on a pitch given by the investigator and then sang it in successively higher and lower keys, ascending and descending chromatically. This continued so long as the child was singing comfortably. Judges were given chromatic scale forms on which they circled the pitches sung in-tune. Judges also rated the total performance according to several criteria concerning the use of the singing voice (47).

The judges used the *Chromatic Scale Form* as an aid in deciding the overall pitch accuracy rating of each child. On this form the judges were to circle individual notes that were out of tune. An adaptation of the Feierabend (1984) scale was used for that portion of the test. A separate score was given for practical range, which was adapted from Rutkowsky (1986) (48). Correlations among the three judges were as follows: Pitch Accuracy, from $r = .89$ to $r = .92$ and for range, from $r = .85$ to $r = .91$ (48).

The existing program emphasized concepts, but also included a balance of singing, moving, playing instruments, and listening activities. Some basic fundamentals of theory and study of composers' lives were also incorporated (48).

A *Discipline-based Arts Education* program had been adopted in South Carolina under which the music teacher was expected to integrate music into the other disciplines. As a result, teachers were expected to give the background of each song. Precise measurements of time were not taken, but it seemed to the researcher that, upon occasion, more time was spent on the background than on the actual singing of the songs (48).

The once-per-week fifty-minute music class typically began with the same song, *Hello Everybody*. Singing was done unaccompanied or with a recording. A variety of songs were sung with minimal repetition (49).

The teacher's technically proficient, light, soprano voice was an excellent model for the children. Although vocal technique was not a high priority in the classes, she did instruct them to use their "singing voices" (49).

During most of the second semester, when the study was in progress, the children worked on the music for the spring program. The second graders sang, danced, and a few played percussion instruments or recorder (49).

It was reported that since the music had to be memorized, there was much repetition of songs during the second semester. The investigator reported that the teacher taught the songs in the exact keys printed, but the keys in *It's a Small World* were lower and more appropriate for most of the singers than those in the music textbook (49).

The researcher concluded classroom visits by teaching a short music lesson during the last music class before the pretest was administered. The recording procedure was explained and the children were recorded individually and privately. Each child sang *Happy Birthday* starting on C4. The song was then sung in successively higher, then lower keys, ascending and descending chromatically as long as the singing was still comfortable. The tester gave each new starting pitch. If a child needed help, the tester sang a few notes to help get started (50).

As soon as the pretest had been administered to all subjects, the researcher began teaching the experimental approach to the experimental group for fifteen minutes at the beginning of the music class. Control Group One was administered the pretest and posttest only, with no other instruction by the researcher. Control Group Two received fifteen minutes of instruction from the researcher at the beginning of their music class; a balanced program without the experimental treatment but not unlike that of their regular music teacher was taught. All groups continued to receive their weekly music class from their regular music teacher (50).

The experimental approach was comprised of strategies to improve pitch accuracy and to develop the head voice. Those group strategies were performed during the first fifteen minutes of seven to eight music classes throughout the semester (50).

Repetition was a key element in encouraging pitch accuracy—repetition of keys, phrases, verses, notes, words, intervals, and songs (50). *Old MacDonald* was an example of a highly repetitive song used in the treatment; with the form, AABA, three of the four phrases were melodically the same, and the B phrase was mostly a repeated note. In addition, the song had many verses. The researcher said that those kinds of repetitions provided drill on the intervals of the song and allowed the children to focus on pitch (51).

Every class included some improvisatory echoes utilizing the descending minor third. The researcher claimed that this interval capitalized on the children's play experiences and was easy for them to sing in-tune (51).

Listening games and strategies to develop inner hearing were employed. Games were created to encourage the child's listening to his/her own voice, and to other voices. Students were asked to match tonal quality as well as pitch (51).

Songs were sung in a comfortable range for the children (which was reportedly lower than they usually sang). Descending minor third echoes on *oo* were sung in different keys, gradually exploring the upper range (51).

Solo singing was encouraged to enable the children to attend to their own voices. Head voice instruction included the following strategies: light, high, speech echoes; short, sung echoes employing the *oo* vowel; movement to reinforce pitch levels and alleviate tension; soft singing with less effort in the lower register; imitation of environmental sounds--sirens, whistles, and animals; and songs that began in the upper register and descended (52).

On the posttest the children's singing was individually and privately recorded in the same manner as had been done in the pretest. To avoid carrying the same mistakes

learned on known material, all groups sang *Happy Birthday*, the song of the pretest, for the posttest but additionally, the experimental group was tested on a song taught as part of the treatment, *Little Red Caboose.* Three experienced music teachers independently judged the taped performances of all pretests and posttests. Judges did not know whether they were judging pretests or posttests (53).

The researcher reported that the results indicated a significant difference in the pitch accuracy of the experimental group for the test of the song used in the treatment. No significant difference was found in the pitch accuracy of the experimental group on a song they had known previously. There was no significant difference between experimental and control groups for range (61).

The following conclusions were made by the investigator:

1. It was not found that second-grade singers underwent improvement in intonation, as a result of group instruction in pitch accuracy and head voice on song material they had learned prior to the instruction.
2. It was not concluded that group instruction in head voice and pitch accuracy would extend practical range significantly more than other types of training and maturation (62).

The investigator recommended that it would be beneficial to replicate the study over a longer period of time to determine more conclusively whether instruction in head voice and pitch accuracy could expand practical singing range, and pitch accuracy on songs already learned (64).

It was also recommended that elementary music teachers do more formal research with children's voices. He said that not only do they have first-hand knowledge of the children, but that they have flexibility to work within the system that an outsider does not have such as making up missed classes, and working on singing instruction with "other activities" (64).

PITCH DISCRIMINATION TRAINING

Porter (1977) investigated the effect of multiple discrimination training on pitch-matching behaviors of uncertain singers. The purpose of the research was to determine the effect of multiple discrimination training on aural pitch discrimination, vocal pitch matching, and instrumental pitch matching, as well as any transfer between those variables (69).

The study was designed to: (1) See if multiple discrimination procedures increase pitch accuracy more than successive approximation alone; (2) Determine what the correlation between vocal and instrumental pitch-matching abilities and aural discrimination before and after training was; (3) Determine if uncertain singers perceive differently than they sing; and (4) Determine the relationship between vocal and instrumental pitch-matching accuracy and time spent and trials to criterion (70).

To obtain a sample of eighty uncertain singers, 263 fourth-grade and fifth-grade students were screened on two of the three pretest measures. The students attended a public school in the New York City area and were classified as being of a low socio-economic status (70).

A five-group, pretest-posttest, experimental control design was used in the study. After an initial screening, students were assigned randomly to the following groups:

X1—multiple discrimination training and vocal pitch matching with successive approximation (MD-VPM)
X2—multiple discrimination training and instrumental pitch matching with successive approximation (MD-IPM)
X3—vocal pitch matching with successive approximation only (VPM)
X4—instrumental pitch-matching with successive approximation only (IPM)
X5—no treatment (70).

The pretest and posttest consisted of three measures. Pitch Test 1 of the standardized *Music Achievement Test 1* (MAT) was used to test for aural pitch discrimination ability. A screening procedure was used for the identification of uncertain singers. The screening procedure also served as the pretest for uncertain singers retained after the screening. This vocal pitch-matching task (VPM) was constructed as follows: the student was asked to sing or hum one tone that he or she felt comfortable with. Then the experimenter determined by stroboscope the pitch of the note of the closest scaled tone to the sound emitted. The remainder of the pretest segment was composed of the following five pitches played consecutively on the *Johnson Intonation Trainer*: the pitch just emitted by the student, a pitch one step above, one two steps above, one a step below, and one two steps below. After each pitch was played, the student was asked to match the pitch vocally. If the student emitted a tone at least one half-step sharp or flat (100 cents deviation), he was penalized with one uncertain pitch. Of the five sung, three uncertain pitches were necessary to meet the operational definition of an uncertain singer (71).

The third measure of the pretest used an instrumental pitch-matching task (IPM). The student was asked to match the same five stimulus tones as used in the VPM task by moving the variable-pitch control dial on the *Intonation Trainer*. Dials moved were randomly set at one hundred cents above or below the pitch to be matched (71).

The specific multiple discrimination training consisted of presenting concept instances (five individual tones) and not-instances while varying the irrelevant characteristics. Since there were a variety of tones as well as irrelevant characteristics, the term multiple discrimination (MD), rather than double discrimination, was considered appropriate (71).

The VPM task, required the student to listen to the tone presented and, as soon as the tone stopped, to sing the tone he or she had just heard. The tone for each student was the same tone he had originally chosen and had emitted on the vocal section of the pretest (screening procedure). The student was informed that the experimenter would give verbal feedback for good performance. The items for the task consisted of the original response, as specified in the pretest measure. The cent deviation for the criterion tone was plus or minus five cents (71).

Each uncertain singer in the treatment groups had a possible total of 128 trials available on each of the five pitches in which he was trained. For the multiple discrimination training groups, X1 and X2, the timbre, duration, and intensity were varied according to a predetermined plan. The timbres used were flute, clarinet, oboe, and violin. The four decibel levels employed were "fifty," "sixty," "seventy," and "eighty." The number of seconds the student could consume during each trial in attaining the correct response was "ten," "fifteen," "twenty," or "twenty-five." Initially, sessions were held in which a student could practice beginning the task promptly when a light flashed on, and stopping when the light went off. The student was also instructed that he might stop before the light was extinguished if he felt the correct response had been achieved. For treatment groups X3 and X4, the timbre was clarinet and the volume was sixty decibels. Students in those two groups were instructed to work until they felt they had matched correctly, i.e., no time limit was imposed for those trials. The procedure for shaping remained constant across all four treatment groups. Each student was reinforced for performances that successfully approximated the correct pitch (72).

For the instrumental pitch matching task (IPM), each student was asked to move a designated dial on the *Intonation Trainer* to make the tone sound the same as the *Intonation* Trainer tone being sounded. The four indicated timbres on the *Intonation Trainer* were varied in order, as well as in the length and intensity of the stimulus tone, exactly as had been done for the MD-VPM task. Each student was trained in five individual tones. The student received verbal praise for correct or approximation of correct responses, with the experimenter following the same shaping procedures that were used for the VPM task (72).

The investigator reported the following data:

1. Pretest and posttest data consisted of scores on a scale of zero to twenty-five for Part I of the MAT, as well as individual vocal pitch-matching (VPM) and instrumental pitch-matching (IPM) scores measured in cent deviation.
2. VPM and IPM scores represented error scores, or the amount of cent deviation from the correct pitch. An analysis of variance between pretest scores on each of the three measures, MAT, VPM, and IPM revealed that the means of the five groups were not significantly different. This suggested to the investigator that randomization procedures used were successful.
3. A correlated *t*-test was used to assess pretest-posttest changes for MAT scores. Results showed a significant difference in scores for Group II (MD-IPM) beyond the $p < .05$ level of significance. No other significant pretest-postest results were found with the MAT measure.
4. An examination of the variance for VPM and IPM posttest scores, as shown by the magnitudes of the standard deviations indicated that a correlated sample *t*-test was inappropriate (72).
5. An analysis of variance was run on the MAT posttest scores to test for a difference between treatment means. No significant differences between treatments were found for this variable.
6. The investigator reported that although there were large differences for the trained variable and the control group, and the trained variable and the untrained variable between VPM posttest means, no large differences were found for MD training versus SA only, trained or untrained on the VPM variable. On IVM there were large differences for the trained variable and the control group. As on the VPM variable, no large differences were found on IPM for MD training versus SA only, trained or untrained (73).

7. Following the analysis of VPM, and then IPM, comparisons were made to determine main treatment effects of MD versus SA only. Although no large differences were found when MD training and SA only were compared for VPM and IPM separately, grouping by training procedures regardless of instrumental or vocal training illustrated large differences between MD and SA: untrained multiple discrimination versus untrained successive approximation, and multiple discrimination versus successive approximation, trained and untrained. Those differences revealed to the investigator that MD was more of a factor in increasing pitch accuracy than SA alone (74).
8. The means for the trained MD versus trained SA groups were 121 and 96.78, respectively, indicating no difference for this comparison.
9. Comparisons were made to observe differences between the trained VPM and IPM variables. The means for the untrained VPM and untrained IPM were M = 888.25 and M = 847.16, respectively, showing a lack of difference between those variables when compared. VPM versus IPM trained indicated large differences.
10. One of the major questions of the study was whether or not uncertain singers perceive differently than they sing. Those results seemed to confirm that perception of pitch was not a major problem since students performed better on the instrumental task than the vocal.
11. An analysis of variance used to test the significance of the differences between time spent, in minutes, during treatment for each experimental group showed a significant diffrence at the .01 level. The use of Scheffe Contrasts indicated significant comparisons ($p < .05$) for five out of the seven groups for time spent during treatment.
12. An analysis of variance was used to test the significance of the differences between the total number of trials to criterion for each of the four treatment groups. There was a significant difference at the $p < .01$ level. Scheffe Contrasts indicated significant comparisons ($p < .05$) between three groups for total number of trials to criterion (76).
13. The Scheffe Contrasts for time spent and trials to criterion during treatment were similarly significant or non-significant for all contrasts with the exception of Groups III and IV. For this contrast, time spent was non-significant, while the total number of trials to criterion was significant (77).

The investigator summarized the following results:

1. Students who received multiple discrimination (MD) training performed better on pitch-matching tasks than did those students in the successive approximation (SA) only groups.
2. Students who received MD training performed better on the untrined variables than did those students in the SA only groups. The results were said to support the hypothesis that MD training not only insured learning a

task, but increased the probability of transfer of learning to other related tasks.

3. Students who were trained on the instrumental task (IPM) were better at IPM than students trained on vocal pitch matching (VPM) were at their task. It was hypothesized that since there were differences found on time spent for the IPM tasks, it may have been that the groups trained on IPM were significantly better due to the increased amount of time spent on IPM or that students learned to tune unisons free of beats on the IPM task.
4. No evidence was found that demonstrated that faulty singing was the result of inaccurate pitch perception. Superior performance on the instrumental task was said to support the hypothesis that faulty perception of specific pitches is not the major problem of the uncertain singer.
5. Training on VPM and IPM had no significant effect in improving MAT scores.
6. The mean initial pitch response of students qualifying as uncertain singers was slightly lower than that of the more accurate singers and pitch behavior was normally distributed (80).

Apfelstadt (1984) investigated the effects of melodic perception instruction on the auditory discrimination of pitch and vocal accuracy of kindergarten children. Sixty-one subjects were assigned to three different instructional settings for eleven weeks. E1 had vocal instruction designed to promote melodic perception through visual aid and kinesthetic reinforcement; E2 had vocal instruction consisting primarily of imitation alone; and C had a traditional, nonconceptual approach. Subjects were pretested and posttested on the *Gordon Primary Measures of Music Audiation* (Tonal Test), the *Boardman Test of Vocal Accuracy*, and a rote-singing test (17).

Subjects were kindergarten children (N = 61) in three intact classes in Madison, Wisconsin, public schools. The two experimental groups came from one school; the control group came from another of a comparable socio-economic level (17).

The experimental classes, taught by the investigator, were randomly assigned to Treatment 1 (E1: vocal instruction reinforced with attention to melodic perception through visual and kinesthetic reinforcement), and Treatment 2 (E2: instruction

consisting primarily of imitation alone). Treatments were conceptually based in that they sought to develop students' awareness of musical elements. The control group (C), taught by another music specialist, was chosen because of its traditional, non-conceptual orientation. All songs were taught by rote without overt reference to musical elements (17).

A total of sixty-one students participated in the study, twenty-four in E1 (eight males, sixteen females), twenty-two in E2 (nine males, thirteen females), and fifteen in C (six males, nine females). To obtain information on the subjects' musical background, the investigator distributed a questionnaire to the parents. Questions pertained to three areas of musical involvement: (1) parent/sibling involvement with music in and out of the home, (2) child's involvement with music in and out of the home, and (3) type of musical equipment (including instruments) in the home. Fifty-six questionnaires were returned for a response rate of 91.8% (17).

The study consisted of five phases: orientation, pretesting, instruction, posttesting, and data analysis. The orientation period allowed the children to become familiar with the investigator before pretesting to reduce test anxiety. With the experimental groups, this period lasted three-and-a-half weeks, and consisted of seven half-hour classes with each group. So as not to bias results of the main study, instruction during orientation period involved no reinforced pitch training. All songs were learned solely by aural imitation (17).

During the orientation phase, the control group met with its regular music specialist. To maintain consistency, the investigator also tested the control group and visited with the children in their classroom several days before the testing began (18).

A second purpose of this first phase was to allow the children time to develop some degree of vocal control. Because two portions of the pretest involved individual singing, the investigator deemed it important that the children felt comfortable using their singing, as opposed to speaking voices (18).

The pretest consisted of three parts, the tonal portion of the *Primary Measures of Music Audiation* (PMMA) by Gordon, a test of vocal accuracy by Boardman, and a rote-song test. The PMMA was administered to each group of children following a brief preparatory activity designed by the investigator to familiarize the students with the concept of "same" and "different" in a melodic pattern context (18).

Several days after the PMMA was given, the investigator began individual vocal testing. Each child attempted to reproduce the twenty melody patterns in the Boardman test after hearing each sung on tape three times, and also sang a familiar song of his or her own choice. Most children sang the chorus of *Jingle Bells*, and the investigator gave the starting pitch (F#4) on a pitch pipe. The test took approximately nine minutes. Subjects were tested in random order in a quiet room with only the investigator and the child present. All vocal tests were tape recorded to be transcribed later. Verbal instructions were written down and repeated exactly for each child. During the process, the investigator kept her comments to a minimum, speaking only to encourage the children and to maintain their concentration (18).

The eleven-week instructional segment consisted of two thirty-minute classes per week with the experimental groups and two twenty-minute sessions plus a thirty-minute recreational singing period with the control group. Vocal instruction in E1 was reinforced kinesthetically, visually, instrumentally, and to some extent, verbally. The E2

group learned all songs without reinforcement of pitch contour or direction; only duration of sounds within a melodic pattern was emphasized. The investigator was careful to spend the same amount of time on corresponding materials so that each group had comparable repetition of a song. All songs were taught without harmonic accompaniment with the intent by the investigator that students could focus on the melodic contour without distraction, and were pitched within a comfortable range from B3 (below middle C4) to C5 (18).

Approximately one-third of each class period (ten to twelve minutes) was devoted to song learning and development of pitch (E1) or duration (E2) concepts. During the remaining time, other concept areas such as timbre and dynamics were taught in identical fashion to both experimental groups. The investigator's consistency was periodically evaluated by the school's regular music specialist as well as by a graduate student in music education, both of whom completed observation forms designed to reveal any inconsistencies in the teaching. Interjudge agreement was $r = 1.00$ (18).

The control group's instruction was activity oriented rather than concept oriented, and the program participatory in nature. Development of specific auditory perception skills or musical concepts was not emphasized. The investigator observed the control group's music specialist three times during the eleven-week instructional phase, and at no time heard her refer to specific musical elements such as pitch or duration or reinforce them in any way. Songs were taught by rote and frequently with piano accompaniment (19).

Posttesting began during the week after the final music class. The PMMA was re-administered first, followed several days later by the *Boardman* and rote-singing tests. The rote-song measure was altered to include two criterion songs taught to all three

groups. Starting pitches were given on a pitch pipe, using the same keys in which the songs had been learned. At no time did the investigator sing to prompt the children. As in the pretests, subjects were heard individually and in random order (19).

Data analysis was accomplished by the investigator, who transcribed all vocal tests onto scoring sheets. The *Boardman* test was graded for pitch accuracy according to criteria established by Boardman (1964), and the rote songs evaluated for accuracy of melodic contour, melodic interval, and maintenance of tonal center using Ramsey's (1981) criteria. Melodic rhythm was not judged in the pretest because the investigator could not determine what version of the song(s) the child had learned. In the posttest, however, accuracy of melodic rhythm was assessed (19).

Two independent raters checked a random sample of twenty percent of both pretests and posttests for accuracy of notation and scoring procedures. Inter-rater correlation ranged from $r = .99$ to $r = 1.00$ on pretests, and $r = .99$ to $r = 1.0$ on posttests (19).

The investigator reported that there were: (1) no differences among groups in auditory discrimination, (2) significant differences on vocal pitch-pattern accuracy between E1 and C and E2 and C, and (3) significant differences in rote-singing accuracy between E2 and C (21).

RANGE TRAINING

Rooks (1987) investigated the effects of remedial vocal training on inaccurate singers. The subjects for the study were students in a Christian grade school in Michigan. The class sizes in the school ranged from seventeen to twenty-seven students per class. The subjects included in the study were drawn from kindergarten, first-grade, second-grade,

and third-grade. Initially, a total of fifty-five subjects were informally identified by the music teacher as inaccurate singers (32).

The subjects were individually given a pretest to assess their vocal deficiency. As a result of the test, forty-two subjects were categorized as inaccurate singers with a complete range, eleven as inaccurate singers with a restricted range, and two as accurate singers. The accurate singers were not used for the study. All eleven of the restricted range singers were included in the study. Out of the group of forty-two complete range singers, thirty-four were randomly selected according to grade level (32).

From the resultant group of forty-five inaccurate singers, three equivalent groups of fifteen randomly-assigned subjects were formed. Each group included five kindergarteners, four or five first-graders, four or five second-graders, and one third-grader. Each group was also balanced with eleven or twelve inaccurate singers with complete ranges and three or four inaccurate singers with restricted ranges (33).

A test of singing accuracy was constructed by the investigator and used as a pretest and a posttest. The first section of both the pretest and posttest was comprised of eight one-note, two-note, and three-note pitch-matching tests and four seven-note matching tests. Half of the test pitches were in the subject's low range, while the other half were in the subject's high range using a retrograde sequence of the low range test. The second part of the test included the singing of a song in both the low and high registers. There was one difference between the pretest and posttest; the pretest song was the familiar *Mary Had A Little Lamb*, while the posttest song was a newly taught song learned during the training sessions (34).

The vocal range of the tests was designed to suit what was believed to be the average range for the different grade levels. The kindergarten and first-grade test range was a ninth with the low range extending from middle C4 up to A4 and the high range from A4 up to D5. The second-grade and third-grade test range was a tenth with the low range extending from middle C4 up to A4 and the high range from A4 up to E5. The pretest and posttest sessions were tape recorded to assist scoring. Three adjudicators were selected to score the tapes independently. The adjudicators had earned at least a BA degree in music education. They scored the tapes by giving a number rating according to the accuracy of the subject's sung pitches (34).

The scoring of the pitch tests required the subject's sung pitch to match the test pitch for a minimum of one second to be considered valid. Each measure of the test was given a numerical score according to pitch accuracy. Four points were given if the sung pitch(es) matched the single pitch test or the multiple pitch test accurately. Three points were awarded if all sung pitches were within a whole step of the test pitch(es). Two points were given if the sung pitch of the single pitch test matched within a major third of the test pitch or all pitches of the multiple pitch test moved in the direction of the test pitch, but one or more pitches were greater than a whole step from the test pitch. One point was granted if the sung pitch of the single pitch test was within a perfect fifth of the test pitch or if the pitches of the multiple pitch test moved from a static pitch level, but did not resemble the melodic contour of the test pitches. No points were awarded if the pitch matching the single pitch test was an interval of a perfect fifth or greater or if the pitches matching the multiple test did not move from a single, static pitch level (35).

The singer's low range and high range scores were computed separately. The singer who achieved a score from seventy-one to eighty-eight was considered an accurate singer in that voice range. The subject who scored from zero to thirty-seven was considered a non-functioning singer in that voice range (35).

The totaling of low range and high range points allowed for classification of subjects as inaccurate complete range singers or inaccurate restricted range singers. The "Inaccurate Complete Range Singer" achieved a score between "zero" to "seventy" in the low range and "thirty-eight" to "seventy" in the high range. The "Inaccurate Restricted Range Singer" earned a score between "zero" to "seventy" in the low range and "zero" to "thirty-seven" in the high range. The total amount of points possible on the combined low and high range scores was "176," while the minimum possible was "zero" (35).

The subjects participated in music class with a music specialist twice a week for one half-hour, except the kindergarten students who had music only once a week for one half-hour. During each music period the students sang for approximately ten to fifteen minutes each period. In addition to singing that occurred in their music class, the students also sang occasionally in their regular academic classrooms. The classroom teachers varied in the amount of time they sang with their students: kindergarten students generally sang fifteen minutes a week outside of the music time; first-grade students sang thirty to forty minutes a week; second-grade students sang about seventy-five minutes a week, and third-grade students sang ten minutes a week (36).

The forty-five subjects, randomly divided into three groups, were given remediation by means of a *Systematic Vocal Training* (SVT) method. The types of SVT used were:

1. SVT-Low--training was given only in the low vocal register (middle C4 up to A4).
2. SVT-High--training was given only in the high register (A4 above middle C to E5).
3. SVT-High/Low-- training was given in the complete range (middle C4 up to E5) (36).

For each training session, the SVT concentrated in three areas: (1) "phonation readiness skills" which included posture and breath control; (2) "vocal control skills" which included speech activities, initial pitch control, single and multiple pitch matching, unison singing, kinesthetic reinforcement, melody matching, and new song learning; and, (3) "pitch and tonal memory skills" which included high/low pitch discrimination, singing accuracy discrimination, silent tone imagery, tone bell usage, and tape recorder assistance in discrimination of accuracy (36).

The SVT training sessions took place at a consistent time outside of the regular music time. The twenty-minute training sessions took place once a week for eight weeks. During the testing, only the experimenter and the subject were in the room. During the training, the groups met in the music room with the fifteen subjects (37).

To determine the interjudge reliability, the judges' scores for the pretest and posttest were intercorrelated. All correlations were $r = .90$ or greater (38).

The investigator summarized the following effects on the study's original null hypotheses:

1. There was no significant difference between the means of the overall pretest to posttest gain scores on a vocal accuracy test due to the training system being either in the low range, high range, or high/low range.
2. There was a significant difference between the means of the overall pretest to posttest gain scores on a vocal accuracy test due to the subjects' pretest vocal range.
3. There was no significant difference between the means of the overall pretest to posttest gain scores on a vocal accuracy test due to the interaction of the vocal training and the subjects' initial range.
4. There was no significant difference between the means of the low range section of the pretest to posttest gain scores on a vocal accuracy tests due to the training system being in either the low range, high range, or high/low range.
5. There was a significant difference between the means of the low range section of the pretest to posttest gain scores on a vocal accuracy test due to the subjects' pretest vocal range.
6. There was a significant difference between the means of the low range section of the pretest to posttest gain scores on a vocal accuracy test due to the interaction of the vocal training and the subjects' initial range.
7. There was no significant difference between the means of the high range section of the pretest to posttest gain scores on a vocal accuracy test due to the training system being in either the low range, high range, or high/low range.
8. There was a significant difference between the means of the high range section of the pretest to posttest gain scores on a vocal accuracy test due to the subjects' pretest vocal range.
9. There was no significant difference between the means of the high range section of the pretest to posttest gain scores on a vocal accuracy test due to the interaction of the vocal training and the subjects' initial range (47).

The following implications were made based on the results:

1. The results did not establish that the posttest gains specifically resulted from the training. This may have been due to the fact that when the three treatments were implemented, a control group was not established for comparison with the treatment group.
2. At least some of the gains in scores from pretest to posttest in the study were due to the training method, and not merely maturation (48).
3. The treatments appeared to be equally effective in promoting vocal accuracy.

4. Twenty minutes per week for eight weeks may have been too brief a time to allow significant differences in the three types of treatment to emerge.
5. The size of the training groups may have been too large for a remedial method which required each child to sing and to analyze pitches individually. It was recommended that a smaller group would have allowed for more individualized attention (49).

In the area of investigation pertaining to the effect of the subjects' initial range on vocal accuracy in the low range, the high range, and the overall vocal range, the investigator reported that the comparison of gain scores between restricted range singers and complete range singers indicated a significant difference with the restricted range singers demonstrating a larger gain in their vocal accuracy scores than the complete range singers. The investigator hypothesized that one factor which may have influenced that restricted range singers' gain may have been regression, the tendency of extreme scores to move toward a median score upon retest. It was also mentioned that another factor may have been that the basic singing skill of utilizing both the high and low ranges is a more easily learned skill than full-range accuracy. It was stated that the implications of the hypotheses for the field of music education was that restricted range singers will respond quickly to basic remedial training and must not be considered beyond vocal assistance (50).

The final area of investigation in the study examined the interaction effect of subjects' initial range and the treatment type on vocal accuracy. The investigator reported that analysis showed a significant interaction between the three treatment types and the subjects' pretest vocal range only for the low range section of the test, and not for either the high range section of the test or the complete test. The investigator therefore, used

only data from the low range section of the test to be submitted to the next levels of examination (50).

Comparison tests revealed that only the gain scores of the restricted range singers showed significant differences due to treatment. In contrast, the gain scores of the complete range singers were not significant when given the three different types of treatment. Looking more closely at the restricted range singers, it was found from the comparison tests that the singers trained in high/low ranges gained significantly more accuracy than the restricted range singers trained in only the high range. The investigator reported that the other comparisons were not significant for singers of a restricted range (51).

The investigator reported that those results may have implications for the training situation and the music classroom. It was mentioned that since complete range singers did not show significant differences on the comparison tests, it could be again suggested that either the study was too short and the group size too large to produce significant differences or that the training concepts could be implemented with comparable effectiveness regardless of range. The investigator reported that the implication for music education was that training is beneficial for complete range inaccurate singers. The investigator reported that the study seemed to suggest that since complete range singers could function in both the high and low ranges, songs used for singing could be in either or both ranges as long as proper singing is emphasized (51).

The investigator concluded that since the analysis revealed that restricted range singers trained in both the high and low range gained significantly more accuracy than those trained in only the high range, a discussion of implications concerning restricted range

singers was possible. It was hypothesized that it may have been that restricted range singers needed to control both their low and high range in order to function accurately in their low range. The investigator recommended that to promote accurate low range singing, training exclusively in the high range is less effective than full range training. This implied that for restricted range singers, accuracy in the lower range could be better gained by singing songs pitched in both their high and low ranges than exclusively in their high range (51).

The following recommendations were made by the investigator:

1. Future studies explore further use of range training with the inaccurate singers of different vocal deficiencies.
2. A larger group of subjects be included in future studies to promote more reliable and more revealing comparisons.
3. A longer duration of training time to make the newly acquired skills more habitual (53).

SEQUENTIAL GAMES

Mathias (1997) investigated the use of a teaching technique aimed at aiding the development of vocal accuracy for inaccurate singers in grades one, three, and five. The technique utilized sequential games consisting of a card game, sound jar game, and voice-matching game. The problem in the study was to determine if the teaching technique helped vocally inaccurate singers in grades one, three, and five become more vocally accurate and to determine if singers could correctly assess their singing performance (5).

Subjects for the study (N = 78) were first-grade, third-grade, and fifth-grade students from five urban public schools in Columbus, Ohio. Elementary music teachers were asked to select children in grades one, three, and five whom they considered to be

inaccurate singers. One hundred children returned consent forms and entered the study; twenty of them received a perfect score on the pretest and thus were not eligible to continue. Two more students left the study for other reasons (37).

Subjects were randomly placed into two groups. Group I consisted of students who played all three games. Group II consisted of students who played only the voice-matching game. In grade one, fourteen students were in Group I and thirteen in Group II. In grade three, twelve students were in Group I and thirteen in Group II. In grade five, fourteen students were in Group I and twelve in Group II (38).

Students were instructed that they would see the investigator singing short phrases on a video tape and that they would be asked to sing what they heard into a tape recorder. The student listened to the first example. The investigator paused the video tape but kept the audio recorder going while the student sang. Then investigator then asked the student if it was a match, and the student responded. The investigator reinforced the student if the response was correct. If the response was incorrect, the investigator told the student it was not a match and asked the student if the response was above or below the example on the video tape. If the student could not give any response, the investigator told the student the answer. The investigator then paused the audio tape. This process was followed for the remaining two examples. Those students who matched their voices correctly with each example were thanked, invited to choose a sticker, and returned to the classroom. The investigator made note of perfect scores on the response sheet. Students who received a perfect score on the pretest were excused from the remaining part of the study at that point. Those students who did not sing all responses correctly remained for the teaching aspect of the study (40).

Subjects were randomly selected to be in Group I or Group II. The participants in Group I were shown the pictures of animals on laminated cards and asked to make the sound of each animal. The students were instructed how to play the card-matching game where cards were turned over and mixed up and matches found by turning them over one at a time. The investigator turned the cards face down and instructed the student to turn one card over and make the sound of the animal on the card. The student then turned a second card over and made the sound of the animal on that card. The investigator asked the student if the cards were "a match." The student answered. The investigator reinforced the child's response if it was correct and instructed the student to turn the cards face down if incorrect and continue playing. If the cards were "a match," they were removed from the playing field and the game continued. The game continued until all matches were found (40).

Next, the student played the sound jar game. The investigator told the child that, because the jars were covered, matches would be found by sound alone. Two jars were filled with pennies, two with rice, and two with popcorn kernels. The investigator mixed up the order of the jars and set them in a line. The child was then asked to select one jar, shake it, and set it out in front of the others. The child was asked to shake another jar and decide if it was a match with the first one. If it was a match, the investigator reinforced the student's answer, and both jars were removed from the playing field. If it was not a match, the student put both jars back in line. This game continued until all matches were found (41).

Next, the student played the voice-matching game. The student watched as the video asked if they could match the hand movement and the singing they heard on the video.

The hand movements were the Curwen hand signs that correlated to the solfeggio syllables in the pattern that was sung. The patterns sung contained the same set of pitches as the pretest, but the words and the rhythms were different. Again, the video was paused after each example for the student to give a self-assessment regarding the ability to match what was heard and seen on the video. The investigator reinforced a correct response. The student was informed if an answer was incorrect, and was asked if the response was above or below the model given (43).

The Group II participants did not play the card or jar game. They played the voice matching game immediately following the coin toss. Group II did the same posttest following the voice matching game and completed the repeated posttest one week later (43).

Posttest phrases were the same as the Pretest, and Repeated Posttest phrases were the same as the Pretest. The patterns were said by the investigator to be based on studies by Davidson (1985), Flowers and Dunne-Sousa (1990), Goetze et. al. (1990), Heaton (1992), Small (1983), Tatem (1990), and Vaughan (1980) (45).

Three sets of the audio response tapes were made and distributed to three judges, along with three response forms for each student—one for the pretest, one for the posttest, and one for the repeated posttest. The judges were all said to be experienced music teachers (46).

The judges listened to each student's pretest, posttest, and repeated posttest and rated each response using the Boardman (1964) scale:

7 = Accurate matching of all tones in the pattern, without hesitation.
6 = The child "slid" into one or more of the pitches in the pattern, but eventually sang all accurately.
5 = An exact transposition of the pattern.

4 = A child maintained the general contour of the pattern, but sang incorrect intervals.
3 = The child maintained the general direction of the pattern but not the exact contour.
2 = Responses which ignored the contour of pitches.
1 = The child spoke rather than sang a response or did not respond at all (36).

After testing for interjudge agreement, judges' scores were averaged for each task on each test. The averages for each task for each test were added together to create a score for vocal pitch matching for each test. Seven points were possible for each task, making a possible perfect score of twenty-one. Data was also collected on correctness of self-assessment for subsequent analysis (47).

After all judges' responses were completed, the investigator compared their responses for interjudge agreement. The investigator reported interjudge agreement in terms of percentages. The percentage of judgments in total agreement or within one difference was tallied. There were forty-two percent in perfect agreement, and seventy-seven percent that fell within the criterion of one difference on the Initial Pretest (IPR). The Initial Posttest (IPO) showed forty-three percent were in perfect agreement, and eighty percent fell within the criterion of one difference. Those percentages of agreement averaged to forty-eight percent perfect agreement among the judges and eighty percent agreement among the judges within the criterion of one difference (49).

It was noted by the investigator that, even though the groups had been randomly assigned, preliminary examination of data on the Initial Pretest indicated unequal performance among groups. For that reason, change from test to test was examined within each group to determine if there was a significant difference. The *Wilcoxin Matched Pairs Signed-Ranks Test* was chosen for the analysis of changes in singing

accuracy from the IPR (Initial Pretest) to the IPO (Initial Posttest), the IPR to the RPO (Repeated Posttest), and IPO to the RPO (50).

It was hypothesized by the investigator that the games may have heightened a child's interest in the task, caused a higher level of focus on the task, or aided in a child's ability to transfer knowledge and learning from one area to another, yielding more immediate vocal accuracy (63).

The following conclusions were made by the investigator from the results of the study:

1. Children's vocal accuracy improved when they were led to apply the aural use of the term "match" and after completing a sequential series of matching games.
2. Vocal accuracy of students in grades three and five showed immediate significant improvement when they completed only the voice-matching game.
3. Most children in grades one, three, and five could correctly assess their own singing performance.
4. Most of the children maintained their improvement after instruction on vocal accuracy one week later (66).

The following recommendations were made by the investigator from the results of the study:

1. Replicate the study in an elementary music class over a period of time at regular intervals to determine if the teaching technique is effective in teaching vocal accuracy in a group setting.
2. Replicate the study with inaccurate singers from different age levels beyond elementary school to determine if the teaching technique is effective with older students.
3. Replicate the study using group comparisons to try to determine if the teaching technique had any effect on any one group or if improvement in vocal accuracy was due to repeated testing.
4. Replicate the study in different cultural contexts to determine if the teaching technique has universal affect with inaccurate singers.
5. Initiate a study that compares the teaching technique of the study with other pitch-matching teaching techniques over a long period of time.

6. Replicate the study and analyze data gathered on the number of tries in each matching game and compare that number with test scores.
7. Replicate the study and analyze self-assessment responses regarding melodic contour discussion and vocabulary used in self-assessments in grades one, three, and five (67).

VERBAL VERSUS VISUAL FEEDBACK

A study by Welch, Howard, and Rush (1989) explored the relationship of visual feedback to pitch accuracy in singing. The authors used a microcomputer-based system to provide real-time visual feedback of vocal pitch production with a class of seven-year-old subjects (N = 32), which was divided into three matched groups based on pretest singing accuracy. During the treatment period of seven weekly sessions, the first two groups used an author-designed software program, *SINGAD*. Group one received singing lessons in pairs of three with one of the authors, while group two, similarly in pairs of threes, used the software without any adult help. The control group undertook group singing activities of a traditional nature, and received verbal feedback as to pitch accuracy (1).

The results of the study by Welch and colleagues (1989) found that on a posttest of pitch accuracy, there was a significant difference between experimental group one and the control group, but no significant difference between experimental group two and the control group, nor between the two experimental groups. A further reassessment of group one after six months revealed that the improvement was generally sustained. The authors noted that all groups showed improvement over the treatment period, but concluded that a microcomputer-based system promoted individual development in singing, in that "verbal feedback on its own appears to be less powerful in promoting learning than real-time, meaningful visual feedback" (156).

This study was said by the investigators to be of major interest to music teachers as to one application of microcomputers in the music classroom. The computer was said to provide the opportunity for students who were inaccurate singers to receive remedial help and feedback on an individual basis. The results of the study indicated that the program was effective with and without teacher assistance. However, the findings were said to be interpreted with caution for at least two reasons. First, students were required to match single pitches only; the investigators noted that there was no evidence to suggest that matching single pitches results in general singing accuracy of song phrases. Second, it was possible for students to learn to match pitch to a visual image using improper vocal technique, in other words, straining at the laryngeal level. The investigators said that although the use of microcomputers in the music class had obvious advantages, it was recommended that care must be taken to monitor for both psychological and motor responses in the psychomotor process (156).

VERTICALLY ARRANGED KEYBOARD

Jones (1971) investigated the use of a vertically-arranged keyboard instrument by the inaccurate singer as a means of improving vocal performance. A test instrument developed for the study included aural and vocal tasks to be performed by the subject. The test was used in a pretest-posttest design (185).

Jones defined the uncertain singer as "the individual who lacks aural-vocal skills." Aural vocal skill was defined as "the ability to perceive pitches in a spatial-temporal relationship and to reproduce pitches vocally with accuracy" (7). The subjects in the Jones (1971) study were from second-grade, third-grade, and fourth-grade classes in four

elementary schools in Tallahassee, Florida. Music teachers in the participating schools identified students who had singing problems. Those students were given a screening test for vocal accuracy in pitch-matching. Only students who matched fewer than seven of the twenty pitches were considered for the study. From that population of uncertain singers, three subjects were randomly selected from each of the three grade levels in the four participating schools, providing a total of thirty-six subjects. Twelve subjects were randomly assigned to each of the three instructional groups with an equal representation of second-grade, third-grade, and fourth-grade students. Each subject received a total of three hours of individual instruction that was divided into daily sessions of fifteen minutes each, covering a period of twelve school days (186).

The test instrument that was developed for the study included a series of aural and vocal tasks to be performed by the subject. Vocal tasks were designed to test the child's ability to match a single tone and to match two or three tones in a series or pattern. The aural tasks required the child to (1) determine how many sounds were played; (2) indicate whether two tones were the same or different; and (3) determine which of two tones was higher. The test of aural-vocal skills was administered at the beginning of the study and again at the end of the study. The *Bentley Measures of Musical Ability* (1966) were also used to test pitch discrimination and tonal memory for each subject (186).

The keyboard instrument selected for use in this study was an electric reed organ. The instrument had a keyboard range of three octaves, beginning on C3 an octave below middle C4. The instrument was modified in the following ways:

1. Those pitches lying within the practical range of the child voice were labeled with letter names so that the instrument could be manipulated by a child with no previous keyboard experience.
2. A panel of red lights was installed beneath the keyboard. The light was visible to the child through electrical contact manipulated by the instructor. It served as a reinforcement for correct response.
3. The instrument was used in a vertical position (186-187).

The instruction in the two keyboard groups was structured around six sequential steps for singing improvement modeled after those of Myers (1965). Those steps were: (1) discrimination between high and low pitches of extreme range; (2) discrimination between high and low pitches within the octave; (3) match a single pitch; (4) match two or three successive pitches; (5) sing a short phrase; and (6) sing two short phrases or one long phrase. Each subject's level of singing development was determined by his pretest performance, and each subject began work toward the next sequential level. Each practice session included discrimination between high and low pitches, pitch-matching, and singing of patterns or melodic fragments (187).

Each subject began vocal practice within a personal range as indicated on the pretest. The sequence that was used presented simple songs of limited range first, and then progressed towards materials requiring a more extended range. Test instructional sequence began with songs using the "five-three" and "three-two-one" scale degree patterns and progressed to patterns involving skips such as "one-three-five" and "five-three-one." Both keyboard procedures required the child to play each pitch or pattern before singing it. The child first watched as the instructor played pitches in sequence; he was then asked to play the same pitches or patterns (188).

The instructional procedure for the vertical-keyboard group began with an introduction to the instrument in vertical position. In order to develop a concept of "high" and "low"

in relation to pitch, the instructor established the relationship between low sounds and low position on the keyboard by having the child play and identify the sound as "low." The procedure was repeated for "high" sounds. The spatial significance on the keyboard was emphasized by asking the child where he would reach to play the sound (188).

Pitch discrimination began with sounds that were more than an octave apart. The range was gradually diminished until the child could discriminate between tones a half-step apart. In order to develop pitch-matching skills, each child began at the pitch he sang most often on the pretest. The child was then asked to play and then sing individual pitches. Keyboard lighting was used at that point to reinforce correct responses. Skills in pattern-singing began with the minor third. The child was required to play the pattern and listen, then play and sing the pattern simultaneously. Finally, the child played the pattern and then sang it, thus demonstrating the ability to remember and reproduce vocally what was played (188).

The instructional procedure for the horizontal keyboard group included the same developmental approach to vocal coordination as the vertical keyboard procedure. The difference between the two procedures lay in the treatment of the concepts of "high" and "low" and pitch direction. Instead of presenting those concepts in terms of the spatial relationship on the vertical keyboard, the horizontal keyboard procedure used "high" and "low" and "up" and "down" only as labels for describing sounds and the relationship between different sounds. The subject had no visual or physical representation of the spatial relationship of pitch when he played the keyboard. The practice of playing all pitches and patterns before singing was maintained (188).

The conventional procedure included the same basic concepts, skills, and song sequence as the other two groups. The same diagnostic procedure was also used; however, the conventional group was not restricted to the sequencing of skills. The keyboard instrument was used as a pitch-giving device with stimulus pitches and patterns presented vocally by the investigator. This procedure used physical action and visual representation of pitch direction, but the child was not involved in playing pitches and patterns as was the practice in the other two groups. In all groups primary emphasis was given to careful listening and accuracy in singing (189).

Results indicated that the vertical-keyboard procedure produced more effective performance results than did the horizontal keyboard or the conventional procedure. Vocal problems related to concept of pitch direction, lack of attention, and low speaking voice were said to be most efficiently handled in the vertical keyboard group (193).

The vertical keyboard procedure and the test instrument used in the study were developed and tested in a pilot study (25). The test instrument included the following. Sections I and III involved pitch-matching with Section I requiring only a limited range of A3 below middle C4 to G4 above middle C4, while Section III extended the range to C5 above middle C4. Intervals and patterns were presented in Sections V and VII. Included in those two sections were the intervals of a major second, major third, minor third, fourth, fifth and octave. Scale degree patterns included in those two sections were "one-two-three," "three-two-one," "one-three-five," and "five-three-one." The pitch range for Sections V and VII stayed within the octave from middle C4 to the C5 above (26).

Aural tasks included in the test required the child (1) to determine how many sounds were played; (2) to indicate whether two tones were the same or different; and (3) to determine which of two tones was higher. Those tasks were presented in Sections II, IV and VI of the test instrument. The test of aural-vocal skills was administered individually to each of the subjects at the beginning of the study. The test and each subject's responses were tape-recorded to insure uniformity of testing procedures (27).

The test of aural-vocal skills was administered again at the end of the study and the posttest was compared with the pretest to determine what improvement, if any, had occurred. In addition to the pretest and posttest, recordings were made of each child's singing voice before and after participation in the study. The recording made at the end of the study demonstrated whether or not the child had reached the fifth and sixth levels of singing ability which required the singing of short and long phrases (28).

In addition to the test of aural-vocal skills designed for the study, the *Bentley Measures of Musical Abilit*y were used to measure pitch discrimination and tonal memory for each subject. The pitch discrimination test required the child to listen to pairs of sounds and to determine whether the second sound moved up, down, or stayed the same. Differences in the paired pitches ranged from twenty-six to three Hertz (28).

The tonal memory test required the child to determine which note, if any, had been changed in paired five-note melodies. The ten test items contained five changes of a semitone and five changes of a whole tone. Although each item contained one changed tone, subjects who did not hear a difference could state that those items were the same (28).

Each subject's performance on pitch discrimination and on tonal memory was compared with Bentley's mean and range of scores for the appropriate age group. Tests were run to determine the correlation between subjects' achievement on the test of aural-vocal skills and performance on the Bentley measures. Bentley claimed a reliability of $r = .84$ on retest. The researcher reported split-half reliabilities of $r = .91$ on the aural-vocal skills test used in the pilot study (29).

Revisions in the aural-vocal skills test included an adjustment in the range of pitches included in the vocal portions of the test. The original range included in the test was from A3 to D5. The adjusted range extended from A3 to C5. The lower extremity of the range, A3, was retained due to the frequent appearance of the pitch in the responses of uncertain singers in the pilot study. Other revisions dealt with the rearranging of some items within the separate sections of the test so as to avoid sequential ordering of similar patterns (29).

Pretest-posttest gain scores were recorded for each of the three instructional groups and three by three analysis of variance was computed using those gain scores. There was a significant difference in the performance of students in the three instructional groups. From those the investigator concluded that the vertical keyboard method produced more effective results in terms of performance on the aural vocal skills test than did either of the other methods. The following results were reported:

1. Statistical analyses showed no significant differences due to grade level or the interaction between grade level and instruction.
2. Second-grade, third-grade and fourth-grade subjects performed equally well in each instructional group, and the instructional sequence appeared to be equally effective with all three age groups (82).
3. Pretest performance indicated that most subjects were able to determine the number of sounds played and whether two sounds were the "same"

or "different," but several subjects were unable to discriminate between high and low pitches.

4. Vocal performance on the pretest indicated that most subjects were deficient in pitch matching.
5. Some subjects were reportedly more accurate in the singing of patterns.
6. The vocal range of responses on the pretest was said to be limited and that most responses fell in the range from A3 below middle C4 to F4 (83).

Analysis of posttest performance revealed the following:

1. Only two subjects who were unable to discriminate between "high" and "low" pitches.
2. Most subjects showed improvement in pitch matching and the singing of patterns.
3. For the majority of subjects, range was increased to include the octave above middle C4.
4. The frequency of patterns matched on the pretest and posttest indicated a preference for descending patterns including the minor third and the three-two-one pattern.
5. The next group of patterns most frequently matched included ascending patterns such as "one-two," "one-two-three," "one-four," and "one-five."
6. The patterns "one-three-five" and "five-three-one" were not often matched on the pretest but appeared as a third group in the frequency count on the posttest (83).

The investigator reported that recordings made at the beginning of the study revealed three basic types of vocal performance:

1. The droning singer who sang on one tone with little variation up or down.
2. The random singer who was more flexible in that he sang both above and below the melodic line, but vocal performance only approximated the contour of the melody.
3. The transposing singer appeared to change key whenever the melody moved out of a comfortable vocal range (84).

Evaluation of the instructional procedure was based on progress noted in daily records. A study of individual progress revealed that most subjects matched patterns before they could match pitches. The researcher noted that discrimination between high and low pitches and pitch-matching did not appear to be prerequisites for the singing of patterns

and short phrases. Individual records from the study showed those skills developed simultaneously (84).

Early in the study the investigator identified a lack of attention as a major problem in singing. Echo clapping was added to the daily practice routine in order to develop listening skills. This addition was said to be beneficial to all subjects, some of whom were able to reproduce extended patterns by the end of the instruction period (85).

The investigator observed the following pitch matching tendencies:

1. Subjects were more accurate in matching single pitches when those pitches were more than a step apart.
2. Accuracy increased when stimulus pitches were at least a minor third apart.
3. Wider skips were found to be useful in helping subjects to gain the use of the upper range or the head voice (85).

Four stages of vocal development were observed by the investigator during the period of instruction. In the first stage, the subject learned to match simple patterns such as "five-three" and "three-two-one," in his own range. The second stage of development included the transposition of familiar patterns for the purpose of extending the vocal range and the acquisition of new patterns such as "one-three-five," "one-five," and "five-three-one," which required the subject to use the newly acquired range. Unison singing was experienced in stages one and two, and the subject was able to sing accurately when singing with the instructor or with the instrument. Stage three was a transition period in which the subject moved from unison singing to independent singing. At stage four the subject was able to maintain a melodic line throughout an entire song without the help of the instructor or the instrument. Most subjects reportedly achieved stage three during the study with quite a few subjects advancing as far as stage four. This skill was

demonstrated in recordings made at the completion of the twelve-day period of instruction (86).

Second-grade, third-grade and fourth-grade subjects were found to be similar in vocal development; however, the investigator recognized a problem in working with fourth-grade boys. Boys of that age were reported as reluctant to sing in the soprano range. In some cases, boys forced their voices down rather than "sing like girls." The investigator said that fourth-grade subjects, both boys and girls, appeared less interested in singing than the younger subjects (86).

Cross-group analysis made it possible to identify common problems among the uncertain singers who participated in the study. Eighteen of the thirty-six subjects were below average in pitch discrimination and tonal memory according to the *Bentley Measures*. Sixteen subjects scored average or above in pitch discrimination, but only four subjects scored average or above in tonal memory. Most second-grade subjects exhibited extreme confusion in responding to directions given in the *Bentley Measures*. For this reason, the investigator was inclined to question the validity of the pitch discrimination and tonal memory tests for children of that age (87).

Several problem areas were identified in which groups of subjects shared the same problem. Ten subjects showed no understanding of the concept of pitch direction. Several subjects sang below pitch due to lack of development of the upper range, while other subjects appeared to be out of tune due to lack of attention. Lack of experience in singing and low speaking voice were also identified as problem areas for some subjects. Individual problems such as extreme shyness emphasized the need for careful diagnosis of singing problems. Although the three instructional methods were equally effective in

dealing with problems related to range development and limited experience, the vertical keyboard procedure was most effective in dealing with problems related to concept of pitch direction, lack of attention, and low speaking voice (87).

Results of group comparisons and cross-group analyses seemed to indicate to the investigator that the vertical keyboard procedure was effective in bringing about improvement in aural-vocal skills. The vertical keyboard procedure produced a higher mean score for improvement than did either of the other two procedures. Second-grade, third-grade and fourth-grade subjects responded equally well to instructional techniques used in the vertical keyboard procedure (88).

Modifications in the instructional sequence were suggested by the investigator:

1. Remedial instruction should begin with the singing of simple patterns since this skill appeared to develop before pitch-matching skills.
2. Pitch-matching and aural discrimination are skills which develop along with singing skills and should not be considered prerequisites for singing.
3. Practice in pitch matching should proceed up or down by intervals of at least a minor third as these pitches are more easily matched than neighboring tones.
4. Exercises designed for range extension should include wide skips such as intervals of the fifth and the octave.
5. Echo-clapping should be included in remedial instruction in singing as a means of improving listening skills (88).

The investigator made the following recommendations for further research:

1. Findings of this study should serve as guidelines for the preparation of an auto-tutorial program for the uncertain singer.
2. Additional research will be necessary to determine the effectiveness of programmed instruction in alleviating the problems of the uncertain singer.
3. Consideration should be given to the designing of a vertical keyboard instrument which could function as a teaching machine.
4. This instrument would provide the uncertain singer an opportunity for individual instruction tailored to his needs.
5. Appropriate measures must be found for assessing the music skills of the child under eight years of age.

6. Song literature and teaching techniques used in fourth grade should be thoroughly evaluated so that singing experiences should not be embarrassing experiences for boys at this age.
7. Teachers should be trained to deal with the vocal development of young boys in terms of their feelings about music as well as their skills in music (89).

VOCAL AND INSTRUMENTAL INSTRUCTION

Buckton (1977) investigated the effects of vocal and instrumental instruction on the development of melodic and vocal abilities in young children. In addition to the vocal and instrumental instruction programs, a control program was included in the experimental design so that changes in performance arising from sources such as fatigue and familiarity with the tests, maturation, and the class music program which the children continue to receive during the time of the experiment were taken into account (37).

Vocal range was regarded as a performance objective within the research and a measure was included in the design. The objectives of the investigation were to investigate pitch discrimination, tonal memory, and vocal range (38).

A second-year infants class with an age range from R = 6.5 to 8.1 years was selected for the experiment. The school was in New Zealand. The pitch subtest of the *Bentley Battery* was individually administered to each student. The split half reliability coefficient computed by the *Spearman Brown Prophecy Formula* yielded a coefficient of $r = 0.68$ (40).

The *Seashore (1939) Tonal Memory Test* was administered. This test used ten three-note, ten four-note and ten five-note tunes and it was found by the investigator that the

use of the three-note and four-note tunes provided a better distribution of scores. The split half reliability coefficient was $r = .71$ (40).

The test of vocal range determined whether each child could use his vocal chords "in such a way as to be recognized as singing, distinct from speaking" (42). If this was possible, the vocal range of notes that each child could sing was ascertained. No attention was paid to accuracy of response, purely their pitches. Children were encouraged to sing their highest and lowest notes by encouragement from a female tester using voice and a glockenspiel. It was not assumed that the obtained range was the absolute limit that each child was capable of achieving. However, vocal range scores seemed to the investigators generally indicative of the typical range that a child used when asked to sing (40). It was reported that only thirty-nine percent of children could sing from middle C4 to C#5 and fifty percent of the children could sing from B3 to C5 (42).

The vocal accuracy groupings were established by employing a test devised by Boardman (1964). The test consisted of twenty melodic fragments, typical of children's song literature of the time. A tape recording of a tonal pattern sung three times by a female was heard. After the third time, the tonic chord which had preceded each of the three renditions was heard again and the child echoed back the song fragment. Each child's response was later scored on a seven-point scale for pitch accuracy. Split-half reliability coefficient randomly selected, adjusted for length through the *Spearman Brown Prophecy Formula* resulted in a coefficient of $r = .97$ (42).

The instruction programs were planned to improve pitch discrimination and tonal memory as efficiently as possible. Taking into account obvious differences in method

due to characteristics of the two media, both procedures were designed to be as similar as possible in terms of activities (43).

The glockenspiel was selected for instrumental instruction. Letter names of the notes were used instead of solfegge syllables. Handsigns, as advocated by Kodaly, were used in both programs, however, in the instrumental program the child applied them to the fixed pitch letter names as if "do" was C, "re" was D, etc. (43).

The same teacher was used for the experimental programs. To examine the effects of the mere difference of media on interaction as an extraneous variable, a tape recording of two sample lessons was made and an interaction analysis was made of those lessons. No marked difference in the type of interaction of the two programs was noted from the analysis (43).

The control group was taught by another teacher and consisted of a music program with predominantly rhythm objectives. Each group received twenty half-hour lessons over a period of eight weeks. Pretests and posttests of pitch discrimination, tonal memory and vocal range were administered before and after the instruction period (43).

The investigator reported the following results:

1. Analysis of the adjusted means scores revealed that the instrumental instruction had been more effective than the vocal instruction. Results indicated that the vocal program did significantly improve the vocal accuracy scores of the vocal group while the instrumental and control groups' scores slightly deteriorated on the posttest.
2. Both instructional programs resulted in a significant improvement in pitch discrimination and tonal memory.
3. The research hypothesis that the vocal program would be more effective than the instrumental or control group in improving vocal range was not sustained.
4. Analysis of interaction for a vocal and an instrumental lesson revealed that there were no marked differences in the type of interaction of the two programs and there was no particular favor on any one group (43).

Using multiple *r* ratios to examine means adjusted for the effects of prior learning, the following results were reported:

1. Both programs improved pitch discrimination and tonal memory.
2. There was no evidence to suggest the vocal program was superior to the instrumental program in improving melodic abilities.
3. The vocal program improved the vocal accuracy of the children.
4. There was no evidence to suggest that vocal range was increased by either of the experimental programs.
5. Children with high vocal accuracy levels and children with low vocal accuracy levels showed more improvement in melodic and vocal abilities than children with medium vocal accuracy (44).

The investigator made the following recommendations for further study:

1. Young children, regardless of initial singing ability, may benefit from either vocal or instrumental instruction.
2. A "playing by ear" program using keyboard percussion may be a valuable aid in improving melodic abilities in young children.
3. The instructional programs, as designed and taught in twenty half-hour periods, may be effective in improving melodic abilities with other junior classes (46).
4. There is a need to design vocal programs to cater for the wide variation which may be expected in the vocal range of children (47).

VOCAL RANGE EXTENSION

Roberts and Davies (1976) investigated a method of extending the vocal range of monotone schoolchildren. A survey of poor pitch singing was carried out among the 18,902 school children attending state schools within five miles of Chester, England. Music teachers were asked to rate each child they taught as being either a normal singer or a poor pitch singer and to classify the nature of any pitch production defect observed. Returns were received for 14,301 children in fifty-one schools. The survey discovered 745 boys and 226 girls, who were rated as monotones or droners and whose singing was described as "Always completely untuneful with little variation in pitch" (31). Among the children sampled within the age range of six-plus to eight-plus years of age, 290 boys

and eighty-seven girls were rated as monotones by their teachers. From those, ninety children at five schools were selected and allocated at random to one of three groups of thirty children. Group one, the remedial group, received the remedial training. Group two was the traditional training group and received extra instruction in the form of singing lessons, and group three, the control group, received only the singing lessons provided as a normal part of the schools' curricula. Of the thirty children in the remedial group eighteen were boys. Of the reported thirty children in the traditional group, eighteen were also boys (31).

The effectiveness of the remdial training of musical recognition and production were determined through tests. The tests were given before and after—which included single-note production and exploring intervals, melodies and rhythms. Those were tape recorded and scored by a panel of three experienced music teachers (31).

The training program for all three groups extended over an eight-week period for two thirty-minute sessions a week. Each child in the remedial program received fifteen minutes of individual instruction at each training session, and spent the remaining fifteen minutes in group work. Individual remedial training started from the pitch skills the child already possessed and then aimed to transfer those skills to singing (31).

The remedial program consisted of the following training steps outlined below:

1. Extension of range of spoken pitch. Utilizing voice inflections in speech and nursery rhymes.
2. Finding a personal note. The child starts to sing any note of his choice on "lah." It involved finding the note on the piano and then doing various exercises on the personal note.
3. Interval production. The child was taught to produce musical intervals on animal sounds. The notes were then reproduced by the child on the chime bells. This occurred during week two.
4. The introduction of further intervals and short tunes. This was introduced during weeks three and four. It involved the interval of a downward

fourth sung to the syllable "moo." Typical excercises were G4 to D4; A4 to E4 and Bb4 to F4. Rhythm work was continued also. Various exercises of this type were employed.

5. Feedback exercises. Those were employed the weeks three to seven. The aim of the feedback exercises was to encourage the child to listen attentively to tones presented and to get the child to monitor the tones as they were self-produced (38).

The purpose of the traditional training program was to give the children extra practice in singing in small groups and individually using songs as the medium of tuition. The main difference between groups was the content of the extra singing lessons. The traditional group learned thirteen songs with either guitar or piano accompaniment (39).

The pretraining and posttraining assessments were scored blind by a panel of three experienced music teachers and the scores examined by a split plot analysis of variance for each test separately. The factors of the analysis were treatment group, and occasion of test which was either pretest or posttest (39).

The investigator reported the following results:

1. All groups of monotones did better on the second occasion of test in seven out of the nine tests.
2. Only on "single note" recognition and "rhythm recognition" was there no overall significant improvement.
3. The remedial group showed greater improvements on two of the nine tests; "single note production" and "interval production."
4. On the other tests there were no greater improvements for the remedial group over the other two groups (39).
5. A significant extension of vocal range for all the monotone groups over the eight weeks of the study was reported, but this extension was greater for the children who had been exposed to the pitch training drills.
6. The same was reported as true for the interval production. The remedial group improved significantly more than the other two groups.
7. The remedial training was concluded a success by the investigators (40).

TEACHER-SUGGESTED METHODS

Gordon (1985) concluded from a survey of the literature that "most strategies have only a few studies to support their effectiveness" (14) and "further study is needed concerning the effectiveness" (16). Gordon recommended that a large body of knowledge in this area needs to be built for practitioners to draw upon during teaching practice.

Gordon sent out a questionnaire to one-hundred randomly selected teachers and requested information from them on techniques they had used in working with pitch defective singers. Responses were received from fifty teachers. Gordon's survey was designed to identify ways in which school music teachers in Pennsylvania worked with inaccurate singers. The frequency of use was the variable to indicate the perceived effectiveness for each technique by the teacher. Gordon hypothesized that since most respondents were experienced teachers it was unlikely that they would frequently use any approach if they felt it to be ineffective. All strategies were then compared with those suggested in the literature (14).

The questionnaire revealed that all respondents indicated that they worked in a remedial manner with inaccurate singers. He found that many teachers found time in their schedules to work in small groups or individually with such students. He hypothesized that such an effort would not be made if pitch matching was of minor importance to music educators (12).

Gordon collected data on materials available to the teachers for the improvement of inaccurate singing. Responses indicated that there was no text specifically designed for aiding the inaccurate singer. He stated that the study revealed a "desperate need for such a collection of tested and proven techniques" (12).

The questionnaire asked the teachers if the materials available were adequate for the improvement of inaccurate singers. The teachers indicated that there were no materials available with pitch matching as the primary concern (12).

Gordon collected data on strategies available for the improvement of inaccurate singing. Teachers responded with many strategies. To obtain the average response for each strategy, Gordon used the following method. Each of the five numbers on the Likert scale was treated as a "quality point" to be multiplied by the number of persons which indicated that level of use. The total number of points was then divided by the total number of responses to that item, thereby producing an average use figure (13) (see Table 1).

Table 1. Strategies Used to Aid the Inaccurate Singer in Gordon's Survey (Gordon, 1985; 13).

Strategy	Mean Frequency of Use
Use of actual song material rather than contrived drills	4.31
Student immediately echoes back pitch or melodic fragment	4.25
Have students think of going up or down stairs (or similar visual cues)	4.23
Use of "call and response" songs	4.13
Begin work with tones within child's apparent vocal range	4.08
Use of entire body motions as an aid for higher and lower Pitch concepts	3.73
Increase amount of time spent on singing	3.61
Play pitches on piano while student simultaneously sings them	3.56
Emphasis of melodic rhythm of given phrase	3.44
Match isolated pitches, then tonal patterns	3.44
Use of hand signals (Kodaly)	3.38
Place inaccurate child between two stronger (in pitch) singers	3.29
Use of contrived drills or songs which concentrate on specific tonal patterns	2.90
Use of listening examples where certain melodic fragments are Used (actual and not contrived music)	2.82
Glissando up or down until pitch is correctly matched	2.81
Begin work with wide, then narrow intervals	2.63
Use of natural speech inflection in speaking exercises	2.54
Use of a vertical keyboard where notes get higher spatially as they get higher in pitch (also bell set or organ)	2.40
Students use humming sound when trying to match a given pitch	2.25
Student sings and plays with melody bells simultaneously	2.19
Use of modes other than major and minor	1.94
Student hears a simple three or four note phrase played on melody bells and is asked to play it back without having seen it being played (student given only those bells needed)	1.92
Child matches sound of a tone bell to the same pitch on piano	1.79
Use of atonal music	1.60
Ask student to think of toothpaste tube; squeeze to go higher	1.27
Pump student's arm to raise pitch	1.25
Pull student's earlobe to lower pitch	1.13

Teachers were asked if the strategies they named were adequate for the improvement of inaccurate singing. The teachers responded that the strategies were "possibly" adequate and it was also indicated that there could never be "too many approaches to the solution of such a highly individualized and complex problem" (14).

Teachers in the survey were also asked if the strategies used were supported by research. Survey answers indiciated that a few were perceived as being supported by research but that most had only one study to support implementation. Teachers admitted that many aids were outgrowths of personal teaching experience (14).

Gordon concluded that teachers need to be made aware of successful strategies. He discovered that tested and somewhat proven approaches "remain relatively unused." He noted the same paradox existed with other techniques which "receive little or no support from research yet are used with great frequency." Gordon noticed that music teachers often simply increased the amount of time spent on classroom singing with no remedial intent. He viewed this to be a sign of frustration with the remedial strategies they were aware of or had used (14).

Gordon found that among those strategies frequently used, the use of actual song material was used more than any other technique. He noted that no research supported the value of using actual songs over contrived drills. Gordon said the immediate repetition of a short musical phrase was supported by several research studies. He said receiving a similar degree of support from research was the concept of using visual cues. Gordon also noted that several studies began with the child's apparent vocal range (14).

Of the moderate use category, playing the piano with the inaccurate singer was used by many teachers. He found only one study to support this approach. He also said the

technique of having the student match an isolated pitch, then a tonal pattern gained support from several studies. Gordon mentioned that the use of the Kodaly hand signals, though wide spread needed to be tested. Gordon said that the placement of students weak in pitch between two good singers had not yet been a focus of scientific study (15).

Gordon cited the use of the vocal glissando as recommended by several studies as well as the use of natural speech inflection as a preparatory experience for calling attention to fluctuations in the voice (15).

Gordon recommended that while the vertical keyboard concept was not yet used by many teachers, that its extensive support by researchers should make it a key aid for teachers. Gordon also said that while many diagnostic devices for determining inaccurate singers existed, very few teachers used them in determining the types of deficiencies an inaccurate singer might have. He said, rather, the situation existed where the student was bombarded with "cures" before the probable causes of the condition were adequately determined (15).

Gordon concluded that while few music teachers would deny children the joy and aesthetic value of group singing, continual group singing without remedial assistance to the inaccurate singer blatantly ignored the problem. Gordon suggested that such an approach "cruelly relegates the problem singer to a lifetime of inaccurate singing" (14).

VOCAL TRAINING OF THE CRICOTHYROID MUSCLE

Yuba (1998-2003) developed a vocal training method that targeted the cricothyroid muscle in singing. His background, as a singer follows. Toru Yuba was a singer who studied with Professor Cornelius L. Reid. He was a leading soloist in such operas as

Pinkerton in *Madam Butterfly*, Rodolfo in *La Boheme* and Don Jose in *Carmen* as well as concerts with many professional orchestras. He was invited by the Chinese Government as a guest of the *Visiting Artist Guild* to perform concerts with Chinese Orchestras in Peking and Shanghai. He performed solo recitals and concerts in Milan and New York. In addition, Yuba performed a leading role in an opera at Columbia University as a part of a New York festival in September, 1997 (Yuba 2001, 3).

Toru Yuba graduated from Tokyo University of Fine Arts and Music in 1981, and was a full member of *The Japan Society of Logopedics and Phoniatrics*, and the *Japan Association for Research in Singing* (Yuba 2001, 2). He began his quest for helping the *tone-deaf* in order to assist students who were pursuing an education degree but could not sing accurately. He explained that the ability to sing was a prerequisite in order to become a teacher in Japan (Yuba 2001, 2).

Yuba originally labeled those who could not carry a tune "tone-deaf." It was his hypothesis that even though the psychological aspects of pitch discrimination and voice production are interrelated, the physiological tendencies outweigh the psychological tendencies in singing a simple song in tune. He noted cases where even professional musicians and composers who had absolute pitch or relative pitch could not sing a simple song in tune. Yuba also noted conversely, that cases were documented where tone-deaf singers could not recognize whether their pitches were correct or not but could sing in tune throughout a song after being "cured" (Yuba 2000a, 1).

Yuba's definition of "tone-deaf" was "unable to perceive differences of musical pitch accurately." Yuba refered to "amusia" as "a form of aphasia characterized by an inability to produce or recognize music." He classified those into four types: (1) instrumental

amusia: loss of ability to play a musical instrument; (2) motor amusia: inability to produce music; (3) sensory amusia: inability to interpret or appreciate musical sounds; and (4) vocal amusia: the inability to sing, although speech is intact (Yuba 2000a, 1).

Yuba explained that the Stedman dictionary defined tone deafness as a type of sensory amusia. Yuba postulated that tone deafness really does not have any relation to "deafness" as the word is used in everyday language. Yuba elaborated that tone deafness can be defined as an inability to carry a tune while being able to hear and perceive musical pitches. Yuba added therefore, that tone-deafness is a physiological concern (Yuba 2000a, 1).

Yuba explained that the phrase "tone deafness" had been used throughout the twentieth century and that this term had given rise to discrimination. Yuba therefore preferred to use the more accurate term "off-key singing" or the politically correct term "vocally challenged" (Yuba 2000a, 4).

Yuba (2002a) identified several causes of off-key singing. He cited environmental factors in infancy, congentifal brain defects, and hereditary, or physiological defects. He hypothesized that a lack of correct vocal coordination is the biggest cause (1).

Research by Yuba revealed that it was possible to acquire a considerable understanding of the vocal mechanism from a physiological aspect. He discovered that there was some data on the human brain processes for speaking, but that information was not reliable even with science and technology being at the forefront. Yuba recommended that more research needs to be conducted in the area of the brain during the process of singing because of limited information on that topic. He believed that singing is like an

exaggerated form of speaking and coined the term "musical data processing" to describe research in that area. (Yuba 2000a, 2).

Yuba (2003a) explained the historical background of off-key singing research. He said that during the twentieth-century, the cause of off-key singing had been widely debated between the dichotomous causes of the psychological hearing ability and the physiological vocal mechanism ability. He said *The First International Symposium on Poor Pitch Singers* was held in 1992 and focused on off-key singing. Yuba said world-renowned scholars discussed the technically varied approaches to the problem. According to Yuba, no concrete solutions relating to the cure of off-key singing were suggested. Yuba said that in 1998, *A Success in Developing a Method of Curing Off-Key Singing* written by Haruki Abe of New York City, was given nation wide coverage in Japan through the *Asahi Shinbun*, the second largest Japanese newspaper. He said that since then, research results on the topic have been frequently reported by the Japanese media. Yuba maintained that the rise in interest is reflective of the growing popularity of Karaoke, or recreational singing in Japan. Yuba estimated approximately forty-nine million people participate in Karaoke singing in Japan. Yuba further cited that in 1999, off-key singing was selected and redefined as a topic at the *Forty-Fourth Annual Conference of the Japan Society of Logopedics and Phoniatrics* by the suggestion of President Sotaro Komiyama of Kyushu University. Yuba added that in 2001, a clinical seminar titled *Curing Off-Key Singing* was held at the *102nd Annual Conference of the Oto-Rhino-Laryngological Society of Japan* and a big step was made towards a concrete solution. He said this attempt in the medical field for the development of a curative method was unique and progressive from an international perspective and leading in

global research. Yuba said that practical and clinical practices in medical institutions and the setting up of a curative system are presently in demand (1).

The following premises gave Yuba the impetus and rationale for his research:

1. Most tone-deaf people have been medically and educationally ignored or not treated properly, because most doctors and teachers strongly believe that their problem is related to aural perception.
2. The existing treatments have been at the level of "old wives-tale" remedies.
3. This has given rise to serious and social problems because there are situations where people usually sing in public such as national anthems, music education and churches everywhere and karaoke especially in Asia.
4. Inaccurate singing makes some people maladjusted to their social environments.
5. In the worst cases, this becomes a serious problem for those who wish to choose a vocation involving music and may lead to psychosomatic problems (Yuba 2000a, 2).

Yuba developed a method called the *Yuba Method*, to cure off-key singing and to develop singing ability based on his research on the vocal mechanism. He claims he successfully cured an entire group of 300 off-key singers on October 30, 2000 (Yuba 2000a, 1). He added that to date, he has cured over 900 people (____, 2003).

The Yuba Method involves the following pedagogical concerns. According to Yuba, pitch, intensity and tone quality are almost always determined by the sound source which is the vibration on the vocal folds. He explained that once a vibration error occurs, the sound cannot be controlled. Yuba claimed that the major vocal training methods of the time addressed breath control and bringing the sound to the resonance position but did not focus on controlling the sound source. The Yuba Method focuses on the physiological aspect of singing. He said that it is possible to quickly train the cricothyroid muscle used for singing and also possible to quickly elevate the air pressure

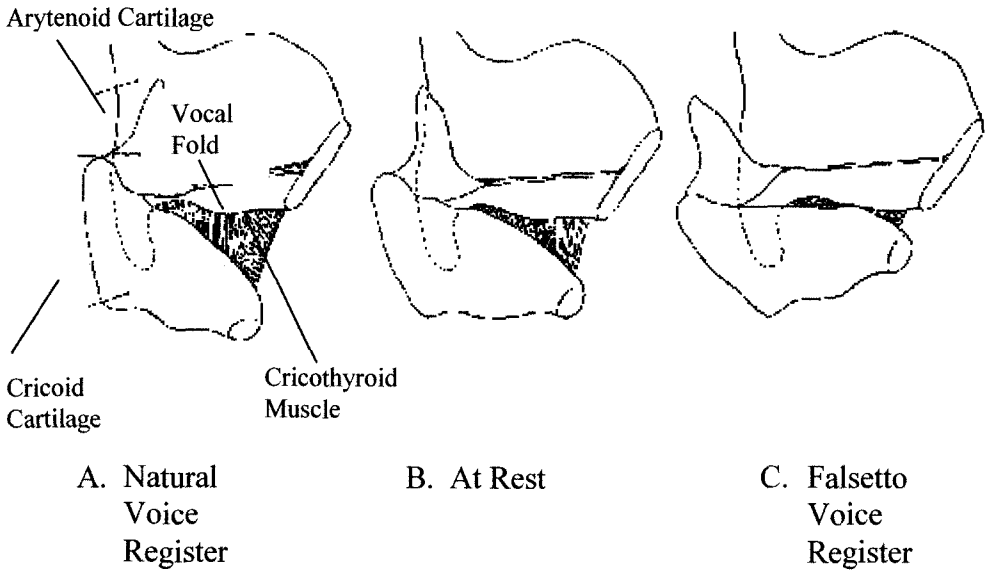
by using this method. Yuba said the method rapidly developed the following voice skills: (1) Expansion of the tonal range; and (2) Increased volume and improvement of tone quality. In addition to those vocal skills, Yuba claimed that the method opened the door to curing off-key singing—which most people strongly believed to be incurable (Yuba 2000, 2).

Yuba explained that mechanically, the cricothyroid muscle determines the pitch like a guitar reel or spool. He said its main function is to act as tensors, tilting the thyroid cartilage forward and downward, lengthening the vocal folds and making them thinner, resulting in raised pitch. Yuba said that conversely, its relaxation lowers the pitch. Yuba elaborated that the preponderance of the cricothyroid muscle against the closing muscle group, or arytenoid muscles, makes a falsetto register made up of breathy sounds because it cannot close the glottis completely. Yuba added that on the other hand, the preponderance of the closing muscle group, or arytenoid muscles, against the cricothyroid muscle makes a natural voice register, or non-breathy sounds. It was explained that the two muscle groups compete with each other to determine the vibrating conditions on the vocal folds. He maintained that subtle control of this vibration is possible only when the vocal folds are being stretched properly. Yuba explained that the cricothyroid cartilage moves in two distinct ways: (1) Forward and downward; and (2) Backward and downward (Yuba 2000a, 4).

Yuba (2002) elaborated that intrinsical laryngeal muscles cannot be directly controlled, but can be controlled indirectly by imitating certain sounds, utilizing a combination of pitch, intensity, vowel, length, and sound quality. He noted that for example, to produce a voice combining those five elements suitable for the objective of

producing the falsetto voice, he has the subject imitate certain sounds to make the movements of intrinsic laryngeal muscles voluntary. He said the contraction of the cricothyroid muscle stretches the use of the vocal folds which mainly elevate the pitch. He added that its voluntary control is difficult to achieve, but that the high-expiration and high-pitched falsetto voice leads to the development of this muscle's desired performance. He said most cases of off-key singing can be cured by balancing the intrinsic muscular movements, or movements of the cricothyroid muscle and closing muscle group, or arytenoid muscles (5). Figure 14 shows the movement of the cricoid cartilage and vocal folds in response to contraction of the cricothyroid muscles.

Figure 14. Movement of the Cricoid Cartilage and Vocal Folds in Response to Contraction of the Cricothyroid Muscles (Phillips, 1996; 229).



Yuba explained that the quality of vibration of the vocal folds is determined by the balance of three factors: (1) the action of the cricothyroid muscle stretching the vocal folds which mainly elevates the pitch; (2) that of the closing muscle group or lateral cricoarytenoid muscle, transverse arytenoids muscle, and oblique arytenoid muscle, which closes the glottis along with the pressure of expiration; and (3) the physical movement of articulation (Yuba 2000a, 2).

According to Yuba, of the above-mentioned three types, the first is the key to curing off-key singing. He explained that intrinsic laryngeal muscles are skeletal structure muscles, typically classified as a group of voluntary muscles, although they cannot be directly controlled at will. Yuba said this is demonstrated in the typical phenomenon of the vocal register break. He elaborated that the intrinsic larygeal muscles instead, react to the environment: pitch, intensity, vowel, length and sound quality. It was further explained that the preponderance of cricothyroid muscle movement produces a falsetto voice and that the closing muscle group, or arytenoid muscles movement makes a natural, or chest voice. Thus, Yuba identified the need to train the cricothyroid muscles by producing a falsetto voice. He also identified the need to train the closing muscle group by producing a natural voice. Yuba hypothesized that harnessing this reaction would be particularly effective in curing off-key singing. Yuba also believed that it is crucial to correctly control the sound source level in order to improve the singer's voice condition (Yuba 2002, 4). Yuba devised a solution to the problem by correctly balancing the following conditions and by doing so, potentially improve the singing abilities of inaccurate singers:

1. High pitch: stretching the vocal folds strongly.
2. Low pitch: stretching the vocal folds weakly (2000a; 4).

Yuba devised a series of musical exercises based on his physiological and mechanical research. Following are the Yuba Method basic steps in curing off-key singing: (1) Distinguish between the falsetto voice and the natural voice; (2) Sing some very simple songs in the falsetto voice and the natural voice; (3) Sing from the falsetto voice to the natural voice and then sing from the natural voice to the falsetto voice (Yuba 2002, 4).

Yuba defined the natural or chest voice as a voice chiefly found in one's speech or low keys and especially notable while producing the vowel sound "ah." He said it is the voice found in the case of the preponderance of the closing muscle group, or arytenoid muscles, over the cricothyroid muscle, resulting in the vocal folds being completely closed, few breathy sounds, and all parts of them vibrating (Yuba 2003b; 3). Yuba elaborated that the sound range of this voice cannot be specified because it shifts up and down according to the changes of its environment from person to person and even in one person (Yuba 2003b; 1).

Yuba defined the falsetto voice as the voice chiefly found in one's high keys and especially notable while producing the vowel sound "oo." He said it is the voice found in the case of the preponderance of the cricothyroid muscle over the closing muscle, or arytenoids muscle group, which result in the vocal folds being incompletely closed, producing breathy sounds, and only their periphery vibrating (Yuba 2003b; 2). He classified the falsetto voice into two kinds: (1) pure falsetto, with no closing muscle group working and resulting in an inability to produce songs, and (2) coordinated falsetto, with some of the closing muscles group working and an abililty to produce some songs.

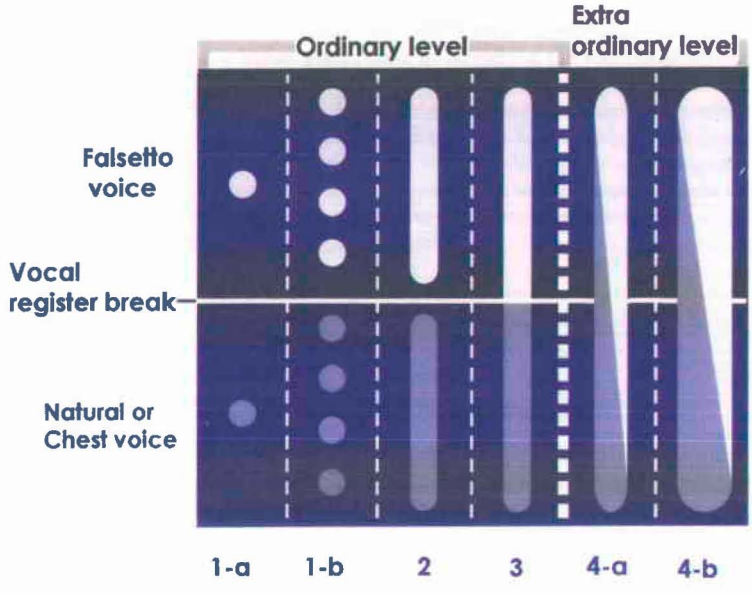
(Yuba 2003b; 2). Yuba said that in the case of pure falsetto, a breathy sound is produced and the closing of vocal folds is partially incomplete at the arytenoids cartilage. He added that the sound range of this voice cannot be specified because it shifts up and down according to the changes of its environment from person to person and even in one person (Yuba 2003b; 1).

It was explained by Yuba that the overlap of the natural and falsetto voices produces a sound range that both overlap and coexist. He said that in this range a voice is produced with the quality of the characteristics of both voices weakened and blended (Yuba 2003b; 1).

The register break was defined by Yuba as a boundary between the natural and falsetto voices. He considered it to be a phenomenon caused by the quick changeover of the functions between the cricothyroid muscle and the closing muscle group, or arytenoid muscles. He concluded that therefore, an unstable voice is often found around the register break. Yuba explained that the position of the register break shifts through a difference in the strength of a voice or a difference in its upturn, or the shift from the natural to the falsetto voice, and its downturn, the shift from the falsetto voice to the natural voice. He said that in the case of its upturn, the position of the register break is usually higher than in the case of its downturn. He added that the position could also shift according to a difference in vowels. Yuba observed cases of no register break either in its upturn or its downturn and also a rare case of no register break both in its upturn and its downturn because of both register mechanisms working together. Yuba said that the position of the register break is usually higher in women than in men. He added that the singing ranges of the natural voice and the falsetto voice overlap each other.

Yuba found that singers could often move across the register break in both registers (Yuba 2003b; 2).

Figure 15. Vocal Registers and Curative Procedures of Off-Key Singing. (Yuba, 2002; 4).



Yuba (2002) illustrated and explained the concepts of the vocal registers and his curative procedures of off-key singing by the Yuba Method (see Figure 15):

1-a = Illustrates the separate voice of the register.
1-b = Shows random voicing in a register separated state.
2 = Illustrates singing a simple song in each register.
3 = Portrays singing with a distinct register break but freely going higher and lower without losing the pitch.
4-a = Illustrates an inconspicuous register break voicing by coordinating the two registers in good balance.
4-b = Shows singing with both registers more strongly and well coordinated (4).

Yuba explained that in phases 1-b and 2, the subject should take care not to make the low-pitch falsetto part of the natural voice. In phase 4-a he instructs the subject to try to

coordinate focusing on the falsetto voice. Finally, in phases 1-a and 4-b he has subjects practice all processes with sample singing and imitative singing. He says that if a subject cannot follow the specified key, he recommends guiding them to it by first adjusting to their personal, or most comfortable key (4).

Yuba classified off-key singing as the following eight types:

1. Gaining correct pitch in the process of singing.
2. Losing the pitch in the process of singing;
3. Repeating off-key and on-key.
4. Singing completely out of tune but able to go higher and lower;
5. Singing parallel to the correct key in higher than normal pitch is a form of distonation, but this kind of off-key singer sings one octave lower, which may make singing partners uncomfortable;
6. Singing in a monotone range, like speaking, losing the pitch or singing in a lower than normal pitch is a form of distonation;
7. Losing the pitch around the gap between the falsetto and natural voice;
8. A composite of various problems from types one to seven (Yuba 2002, 2).

Yuba discovered that it was possible to cure all eight of the off-key singing types by the following three basic approaches. In the case of a singer:

1. Producing a pitch that was too low, it was necessary to produce a breathy falsetto voice by increasing the amount of breathing to raise the pitch.
2. Producing a pitch that was too high, it was necessary to produce a natural voice by decreasing the amount of breathing to lower the pitch.
3. Being out of tune at the register break, it was necessary to coordinate the two registers, the natural voice and the falsetto voice (Yuba 2002, 3).

Yuba said that a special feature of his method was its drastic and simultaneous elevations of cricothyroid muscle tension and of air pressure under the glottis while maintaining the stability of articulation. Yuba said this was possible by imitating a dog howling—an imitation which produced a high-pitched, breathy, falsetto voice (Yuba 2000a, 4).

Yuba said that training must be maintained to ensure sustained singing ability. Yuba did not provide an exact time frame for this training. There were six targets for the Yuba Method vocal training:

1. To sing smoothly in a wide tonal range with a register break. Once the subject learned how to blend and work the two voicing mechanisms for the falsetto voice and for the natural voices respectively, the register break could be eliminated. The subject was then able to sing songs meeting all the musical requirements.
2. To sing with flexible control of volume from an extremely minute sound (*ppp*) to a tremendously powerful one (*fff*).
3. To sing in a perfect pitch suitable for the musical expression.
4. To sing for a very long time in one breath (transforming a breath into voice quite effectively).
5. To sing with a clear pronunciation, or articulation, with good resonance.
6. To sing in every shade of tone color (Yuba 2001, 1).

Yuba (1998) also used a rubric rating system for his self-study method book for subjects to check on their progress. Yuba designed two check charts—A and B for his off-key song correction program. In Chart A he outlined five basic steps for the subject to follow: (1) Hidden or falsetto/natural voice; (2) Establishment of falsetto/natural voice separation; (3) Falsetto strengthening and stabilization; (4) Transference from falsetto to natural voice; and (5) Transference from natural voice strengthening to falsetto. Yuba then had subjects rate themselves with the following rubric numbering system:

1. Incapable
2. Somewhat capable
3. Capable
4. Able to do with confindence
5. Consistantly able to do with confidence (56)

Yuba specified that once the subject had "confidence in the correctional program" to continue on to Check Chart B. Check Chart B consisted of eight basic steps for the subject to follow: (1) Hidden or falsetto/natural voice; (2) Establishment of falsetto/natural voice separation; (3) Falsetto strengthening and stabilization;

(4) Transference from falsetto to natural voice; (5) Transference from natural voice strengthening to falsetto; (6) Falsetto and natural voice fusion one; (7) Falsetto and natural voice fusion two; and (8) vowel and consonant stabilization. He then had subjects rate themselves with the same five-point rubric system stated above (57).

Yuba (2000b) explained how he kept count of off-key singers during a December 2000 treatment session. He used a rubric system consisting of symbols:

X = The singing is extremely off-key.
△ = The singing is partially off-key.
□ = The singing is slightly off-key.
● = The singing is in a permissible range (1).

The 2000 treatment session consisted of twenty subjects. Out of the twenty subjects, thirteen were reported as being off-key singers and seven as on-key singers. Out of the thirteen off-key singers, nine were reportedly classified as extremely off-key singers, two as partially, off-key singers, and two as slightly off-key singers. Yuba worked with the subjects for a total of ninety minutes utilizing the Yuba Method treatment exercises. After the treatment session, all twenty subjects were reportedly classified as singing in a permissible range (1).

Yuba (1998) developed a Yuba Method book *Dream Voice Training* designed to be used as a self-study for inaccurate singers. The instructional book came with illustrations, instructions, and an instructional CD with synthesized instrumental accompaniments and vocal examples to echo. Male and a treble/female versions of the exercises were included in the method book. The male version used a male vocal model and the treble/female version used an adult female model. The male version of the exercises were pitched a perfect fifth below the treble/female version. A total of

approximately forty-five to sixty minutes were required to complete all of the exercises. Yuba said to repeat the exercises to maintain vocal control and training.

A summary of the treble/female exercises follows (see Appendix VIII for the complete exercises):

1. Mouth configuration and proper standing position.
2. Vocal modeling of the falsetto and natural voices.
3. Practice of falsetto voice sounds: Echoing of sounds "Yahoo!" "Hoo, hoo," "woo, woo."
4. Practice of natural voice sounds: Echoing of sounds "Ah, ah, ah."
5. Falsetto/natural voice separation and establishment: Ten echo exercises alternating between the natural and falsetto voices. Utilizes the pitch C4 for the natural voice and the pitch C5 for the falsetto voice.
6. Falsetto strengthening and stabilization: Nineteen echo exercises that ascend and descend from A4 to F5 and increase in length from a single pitch to seven pitches.
7. Transference from the falsetto to the natural voice: Ten echo exercises that descend from D5 to D4 with increasing length from a single pitch to four pitches.
8. Transference from strengthening the natural voice to the falsetto voice: Thirty-two echo exercises. The first 24 exercises use durations of whole notes tied to dotted half note rhythms that begin on D4, descend to G3, ascend to B4, and descend back down to G3. The next eight exercises ascend scale-wise, in the G major scale from G3 to B4 (across the register break) and back down from B4 to G3.
9. Increased singing ability, Falsetto and natural voice fusion 1: Twelve echo exercises that either descend and ascend scale-wise, in the key of D major, from D5 to D4, increasing in length from one measure to three measures, and also increase in the size of the intervals.
10. Falsetto and natural voice fusion 2: Eight echo exercises of patterns in the key of D major that ascend from D4 to D5. The exercises begin with ascending scale patterns and then progress in intervallic size up to the octave. Falsetto voice syllables ("hu") and natural voice syllables ("ah") are alternated throughout the exercises.
11. Vowel and consonant stabilization: Six echo exercises that focus on the vowel sounds of "a," "e," "i," "o," "u" and "la." The exercises begin on repeated tones of A4 using the vowel sounds and descend to A3. The final exercises use intervals of the octave and then descending and ascending scales from D4 to D5.
12. Singing muscles final check: Thirteen echo exercises in range from D4 to G5. These begin from short and separated motives which descend from G5 to G4 and progress to ascending/descending scale degree patterns from D4 to D5 increasing in interval size to the octave (42-55).

Yuba claims that to this date, he has cured over 900 off-key singers. This has been documented in a report to the Japanese government (_____, 2003) *The Report of a Candidate for Distinguished Researcher Medal* in which Yuba received an achievement award *The Development of Correcting Off-Key Singing Based on the Controlling Technique of Muscular Activities of the Cricothyroid Muscle*. This award is annually awarded to only six or seven researchers working at universities out of approximately 600 universities in Japan (1).

Need for the Study

In light of the review of literature in regards to inaccurate singing—causes and remedial strategies, it is apparent that there needs to be further investigation into remedial strategies that can help the inaccurate singer. As Gordon (1985) mentioned, most strategies have only one or two studies to corroborate their effectiveness or noneffectiveness. Furthermore, those results were sometimes conflicting. While there are many causes for inaccurate singing, no remedial strategy has proven to be adequate for the vast needs of inaccurate singers and their teachers to assist them. This review of literature finds that there is no known research study based on the physiological training of the cricothyroid muscle as suggested by Toru Yuba. Review of literature on the physiology of singing indicates that the cricothyroid muscle controls the pitch in singing. Yuba's claim of a success rate of curing over 900 tone deaf singers has been unsurpassed by any known remedial technique. Previous remedial strategies have found some success—many in matching single pitches or intervals which may not correspond to free song singing--but none has proven to be as effective as the Yuba Method claims and in

such short a duration of training. With the limited amount of time music educators have with their students, it seems advantageous that a method that best utilizes the relatively small amount of time with students be investigated. Empirically testing the Yuba Method would provide more accurate information concerning its effectiveness with children.

CHAPTER III

METHODOLOGY

NULL HYPOTHESES

The Yuba Method vocal training program will have no effect on the vocal pitch accuracy of inaccurate fourth-grade, fifth-grade, and sixth-grade students.

PILOT STUDIES

An initial pilot study was conducted in the spring of 2002, and a second pilot study was conducted in the fall of 2002 with Toru Yuba present, prior to the implementation of the present study. In both pilot studies, adjustments were made to the singing stimuli and the Yuba Method procedure. The method was verified by Toru Yuba and members of the graduate faculty at the University of Hawai'i at Manoa.

SELECTION OF SUBJECTS

Sample Population Demographics

The subjects were fourth, fifth, and sixth grade students (N = 320) in one public urban elementary school in Honolulu, Hawaii. This comprised the entire enrollment of grade four, five and six students at the school. The total number of students enrolled in the school was 791. The percentage of students enrolled for the entire school year was 94.6 percent. The percentage of students receiving free or reduced lunch cost during the school year was 45.6 percent. The percentage of students in special education programs during the school year was 2.5 percent or twenty students. None of the students in the population had any hearing anomalies. This was verified through discussion with the

school's special education teacher. The percentage of students with limited English proficiency during the year was eleven percent (N = 87). The ethnic make up of the school population consisted primarily of students of Asian and Pacific Islander heritage.

Letters to the administration of the school were sent by the researcher, notifying the administration that the music teacher would be provided and asked for written permission to conduct the study. Permission was obtained. Parental permission forms were also sent home to the prospective subjects and permission was obtained. The *Use of Human Subjects in Research* form for the University of Hawaii at Manoa was completed and submitted. Permission was granted (see Appendix VII).

The population of fourth graders included all subjects between the ages of 8.5 to 9.4 years of age by September 1, 2002. The population of fifth graders included all subjects between the ages of 9.5 to 10.4 years of age by September 1, 2002. The population of sixth graders included all subjects between the ages of 10.5 to 11.4 years of age by September 1, 2002.

Selection Process

PRETEST SINGING STIMULUS

The 320 fourth-grade, fifth-grade, and sixth-grade students were given the *Pretest Singing Stimulus* in November 2002. The *Pretest Singing Stimulus* was designed by the investigator to classify subjects either as "accurate" or "inaccurate" singers. The *Pretest Singing Stimulus* consisted of the first phrase of the Israeli folk song *Shalom Chaverim* in the key of D minor. This song was chosen by the investigator because it was a song that was intended to be part of the regular music curriculum (see Appendix I). Actual song material rather than single pitches was chosen by the investigator based on prior research

by Goetze (1985). Goetze concluded that "the pitch material used for the singing task to measure accuracy should be a melodic pattern rather than a single tone" (66). A script was prepared to insure uniformity of administration of the *Pretest Singing Stimulus* (see Appendix III). The range of the *Pretest Singing Stimulus* was from A3 to D5. Collins (1981) noted that varying ranges were reported for the different age groups. Collins reported that for grades four through five, reported ranges varied from A#3-D5 to A3 to F5 (10). Goetze (1985) reported similar ranges for ages eight through twelve as C4 to F5 to A3 to F5. The selected criterion pitches for this study were therefore within the recommended ranges for the age group selected. Students were given the first two starting pitches of the phrase on the Yamaha PSR-540 Electronic Keyboard and they were required to sing the entire phrase "a capella." It was decided to use "a capella" singing rather than echo singing after a vocal model because gender or pitch accuracy could contaminate the singing of the subjects. Selected criterion pitches were used to calculate singing accuracy rather than using a subject's deviation on all of the notes in the melody. Goetze (1985) noted that "when such a subject's deviation on all of the notes in the unidirectional melodies were averaged, the score was found to be less descriptive of the student's accuracy than an average of selected notes" (75). No standardized set of criterion pitches has yet been determined and have varied from study to study. Aaron (1992) used the criterion pitches of C4, E4, F4, G4, A4, and C5 for his study. Smale (1987) used the criterion pitches of C4, D4, E4, and G4. Goetze (1986) used the criterion pitches of D4 and A4 (67). Clayton (1986) used the criterion pitches of B3, D4, F#4, and A4. Cooper (1992) used the criterion pitches of C#4, E4, and F#4. The selected criterion pitches to be analyzed in the present study were D4, F4, C5 and D5. It was desirable for

the purposes of this study to encompass both the chest and head registers. Past research indicated that the vocal register break was a possible cause of singing inaccuracy. Cooper (1992) identified that children who had not yet learned to use the head voice register had difficulty matching pitches above the voice break (36). Cooper (1992) identified the break between the chest and the head voice to occur around G4 or A4 (36). The Yuba Method used in this study, specifically addresses the vocal register break, further emphasizing the need for criterion pitches to be within A3 to D5 as chosen.

CALCULATION OF THE VOCAL PITCH ACCURACY SCORE

A vocal pitch accuracy score (VPA) was calculated for each student using the *Sona-Speech Model 3600 Software Program.* The *Sona-Speech* software program was used to calculate in Hertz, the frequency of each selected criterion pitch of the *Pretest Singing Stimulus* and the *Posttest Singing Stimulus*. The Sona Speech is the software-only component of the *Visi-Pitch*--a clinical package of speech training and analysis programs which was used in previous research by Brown, 1988; Clayton, 1986; Goetze, 1985; Goetze & Horii, 1989; Moore, 1991; and Smale, 1987).

The *Sona-Speech* sampled the recorded voice and displayed the frequency curve of the criterion patterns on the computer monitor. The investigator then moved cursors to outline the segments of the curve representing the pitch to be analyzed, and the *Sona-Speech* automatically calculated the frequency, in Hertz of the pitch area between cursors. The entire sampled frequency curve could be played back in real time as a cursor showed the segments being played back. This helped the investigator to find the selected criterion pitches on the curve. The outlined segments of the curve to be analyzed could

further be confirmed through playback of the selected areas between the cursors by selecting "Speak Selection." The *Sona-Speech* then played back the selected area to confirm that it was the exact placement of the desired criterion pitch. The selected parameters were thus outlined by the cursors visually and aurally by the investigator. Because frequencies in Hertz are not equal-interval data, logarithms were used to calculate the interval or deviation in cents, where 100 equal cents equaled one semitone, between each response pitch and its corresponding stimulus. The total deviation in cents between each response pitch and its corresponding stimulus was calculated. Figure 16 presents *Sona-Speech* cursors outlining a sample criterion pitch. Figure 17 presents the *Sona-Speech* display showing the analysis of a selected sample criterion pitch.

Figure 16. *Sona-Speech Model 3600* Software Program *Pretest Singing Stimulus* Vocal Pitch Display with Cursors Outlining a Sample Criterion Pitch of F4.

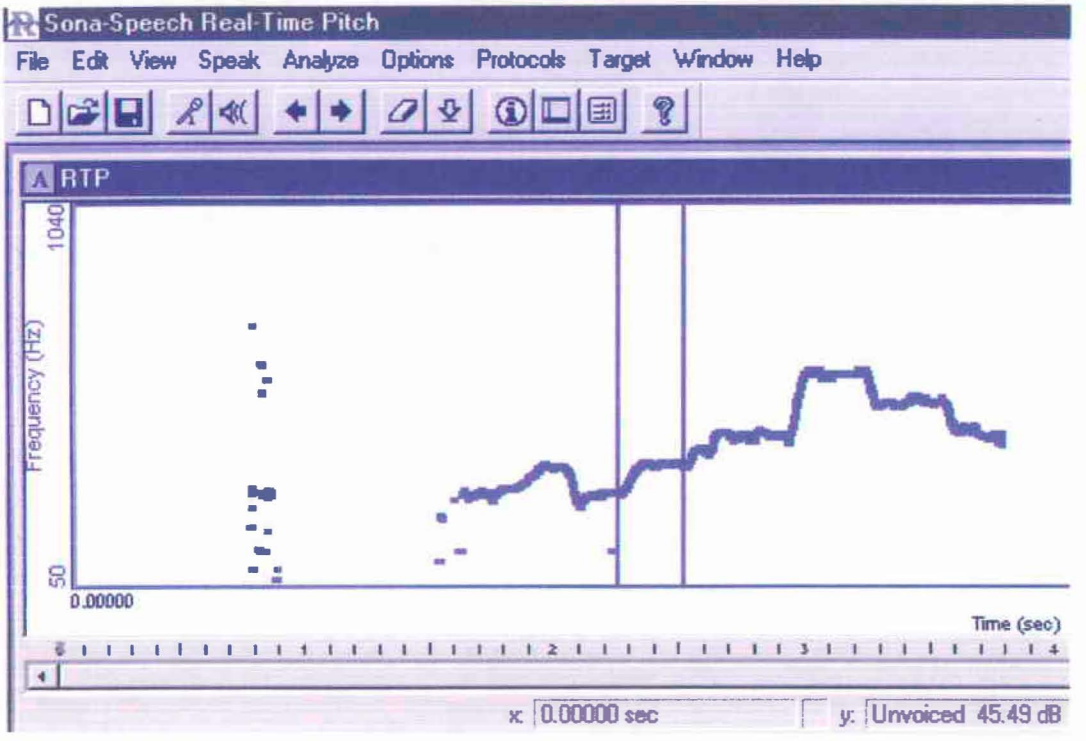


Figure 17. *Sona-Speech Model 3600* Display Showing the Analysis of a Selected Criterion Pitch on a Sample Singing Stimulus Result.

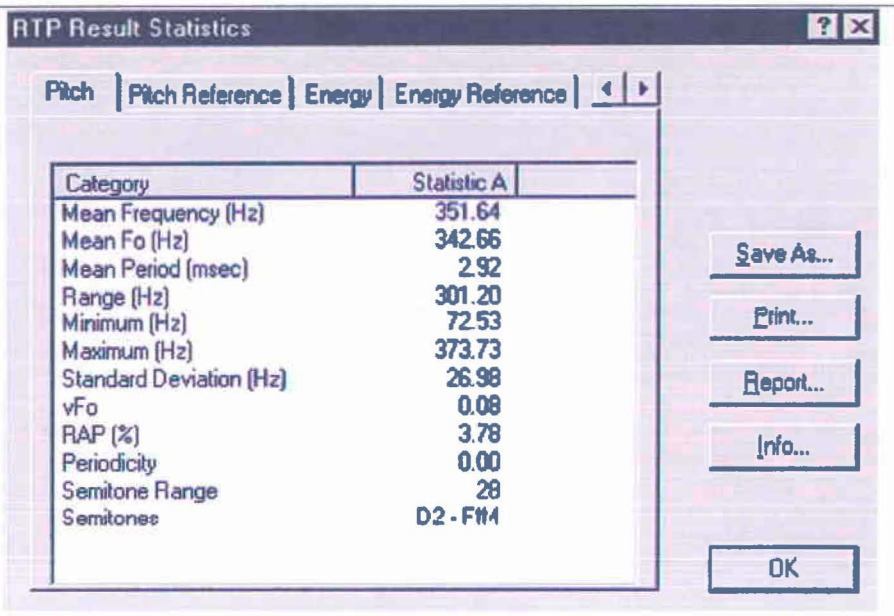
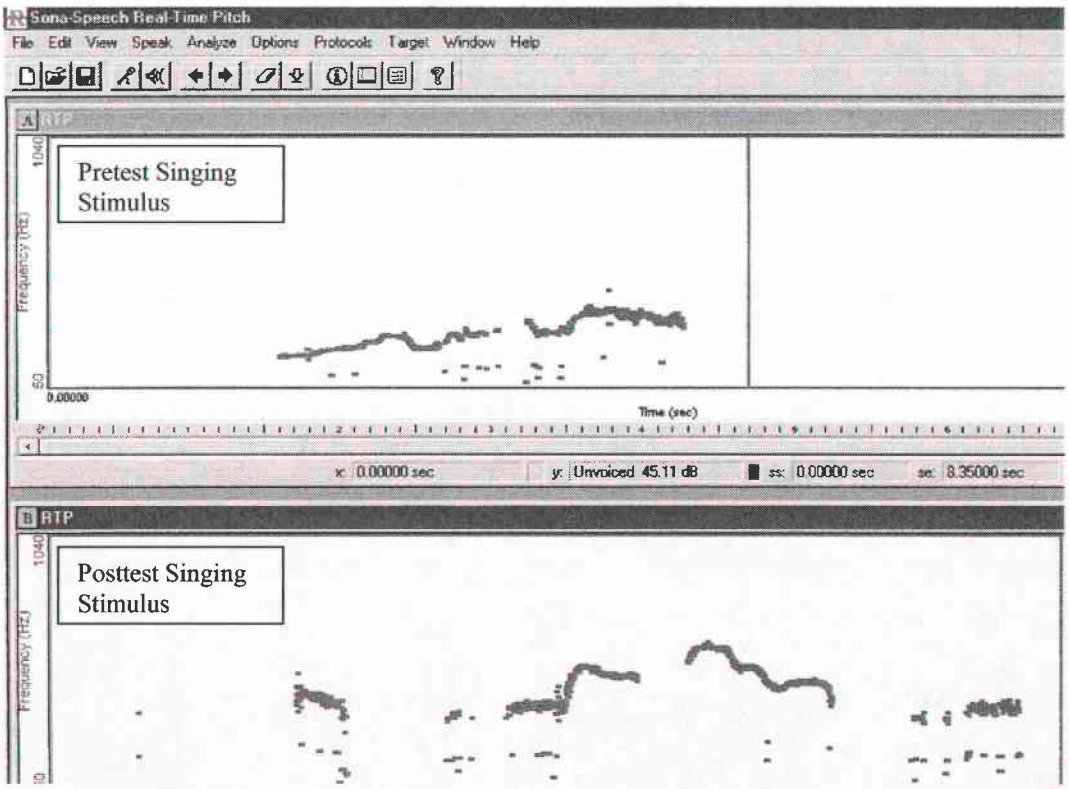


Figure 18 presents an example of the display of the *Sona-Speech Model 3600* "high" treatment subgroup subject number 61310 with a *Pretest Singing Stimulus* VPA score of 755.47 followed by the same subject's *Posttest Singing Stimulus* display with a VPA score of 25.32. Differences in the melodic contours of the stimuli can be visually on the *Sona-Speech* display.

Figure 18. *Sona-Speech Model 3600* Display of High Treatment Subgroup Subject Number 61310 *Pretest Singing Stimulus* (VPA 755.47) and *Posttest Singing Stimulus* (VPA 25.32) Recordings.



Absolute values were used in these calculations to avoid the possibility of both positive and negative cent deviations. For example, sharp and flat responses, respectively, on different pitches within the pattern might cancel each other out. Therefore, although VPA scores represented overall deviation from the model or actual pitches, they did not provide an indication of the direction of deviation or contour of the response. Because VPA scores represented divergence from the model, lower scores indicated more accurate performance and higher scores indicated more inaccurate performance.

Calculation of the size of pitch intervals followed the Campbell & Greated (1987; 77-79) procedure. The following formula was used:

Pitch interval, in cents = 1200log2(f2/f1)

In this formula "f2" represented frequency "two" (the larger of the two frequencies being compared), and "f1" represented frequency "one" (the smaller of the two frequencies being compared). The *Musical Note to Frequency Chart* (Kay Elemetrics; 234) was used to calculate the absolute note value to frequency in Hertz (see Appendix XI).

A reliability coefficient of $r = .98$ for the Visi-Pitch was obtained by Goetze (1985) using three raters. Aaron (1991) estimated a Visi-Pitch reliabililty coefficient of $r = .97$ as a result of rescoring sixteen individual pitches of a pilot tape.

This study similarly obtained a *Sona-Speech* reliability coefficient of $r = .99$ after the investigator rescored sixteen individual pitches of randomly selected *Prestest Singing Stimulus* recordings. The rescoring of the sixteen individual pitches helped to confirm the consistency of scoring by the investigator.

SUBJECT GROUP PLACEMENT

The population of 320 students was classified into two groups, accurate (N = 152) and inaccurate singers (N = 168), based on the *Pretest Singing Stimulus* vocal pitch accuracy score. Subjects with a VPA score of 100 or greater were identified as inaccurate singers. Subjects with a VPA score below 100 were identified as accurate singers. The accurate singers were not used in the remainder of the study. The criteria of using the VPA score of 100 or greater to determine the inaccurate singer was used by Goetze (1986), Smale (1987), Brown (1988), Goetze & Horii (1989), and Cooper (1995).

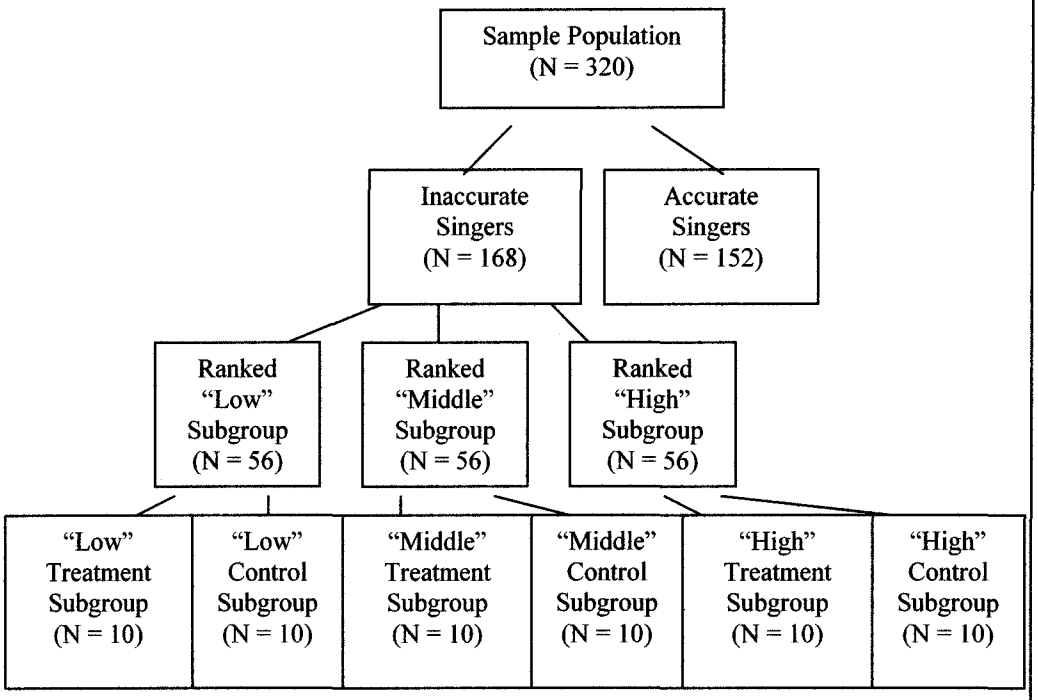
All of the inaccurate singer scores (N = 168) were listed in ascending order from the lowest to the highest scores. The inaccurate singers were divided into three subgroups of equal size (N = 56) to constitute "low," "middle," and "high" VPA scores. Dividing and classifying the inaccurate singers into three subgroups assured that subjects would be randomly selected at all degrees of VPA scores.

The *Komolgorov-Smirnov* test for comparing two populations was calculated between the three subgroups of "low," "middle," and "high" to determine if in fact the populations were different. Results of the *Komolgorov-Smirnov* test for comparing two populations between the "low" and "middle" groups indicated at the $p < .05$ confidence level that the two subgroups were from different populations. Results for the test between the "middle" and "high" groups also indicated at the $p < .05$ confidence level that the two subgroups were from different populations. The three classification subgroupings were therefore deemed an appropriate design for the experiment.

A total of sixty subjects were randomly selected, from the three subgroups of "low," "middle," and "high" VPA inaccurate singers, to be either in the treatment group (Yuba

Method) or control group, using a table of random numbers. The following procedure was used to randomly select subjects from the three subgroups. From the "low" subgroup (N = 56), ten subjects were randomly and alternately selected, to be in the "low" treatment subgroup (N = 10) and the "low" control subgroup (N = 10). From the "middle" subgroup (N = 56), ten subjects were randomly and alternately selected, to be in the "middle" treatment subgroup (N = 10) and the "middle" control subgroup (N = 10). From the "high" subgroup (N = 56), ten subjects were randomly and alternately selected to be in the "high" treatment subgroup (N = 10) and the "high" control subgroup (N = 10). Subsequently, there were a total of thirty students that were trained by the Yuba Method, in the treatment group, and thirty students that were not given the Yuba Method training, in the control group (see Figure 19).

Figure 19. Diagrammatic Representation of the Experimental Design for Assessing the Effects of the Yuba Method Training on Inaccurate Elementary Singers.



Note: "Low," "Middle," and "High," refer to VPA scores. High values denote the most inaccurate singers.

The range of VPA scores for the various subgroups is summarized in Table 2.

Table 2. Range of VPA Pretest Singing Stimulus Scores in Subgroups.

Subgroup	Range of VPA Pretest Singing Stimulus Scores
Low Treatment	115.12 – 152.44
Low Control	106.42 – 132.15
Middle Treatment	161.64 – 242.75
Middle Control	162.83 – 224.77
High Treatment	276.37 – 755.47
High Control	282.12 – 518.71

No distinction was made between subjects based on prior musical training, musical aptitude, hereditary musical background, or factors other than VPA scores. All testing and treatments occurred during the hours of the regular school day.

Only subjects who completed the entire Yuba Method training session for the treatment group and were present for a minimum of seventy five percent of the instructional song stimulus instruction period were included in the study. No subjects were eliminated on the basis of missing the minimum requirements of instructional time. Two subjects in the treatment group were replaced with pre-randomly-selected alternates when they became ill during the treatment session. The total number of subjects for the control group was

N =30 and the total number of subjects for the treatment group was N =30. There were no direct costs for the subjects and no direct cost for the music teacher since the music teacher was also the researcher.

Each subject in the study was assigned a five-digit subject number which indicated on the following: (a) Digit one represented the subject's grade level; (b) Digit two represented the group assignment. The group receiving the Yuba Method treatment was assigned the number "one." The control group was assigned number "two"; (c) Digit three represented the "low" (1), "middle" (2), or "high" (3) groupings within the experimental or control groups; (d) Digits four and five represented the subject number within the treatment or control groups. For example, subject number 62101 was a sixth grader, the first subject in the control "low" group.

EQUIPMENT

All equipment was provided by the researcher. The equipment consisted of a Gateway laptop PC computer, an Electro Voice Microphone, a Yamaha PSR-540 Electronic Keyboard, a two-foot by four-foot mirror, a Sony CFD-V17 CD Player, a Sony Hi-8 Video Recorder. The software program used to analyze the criterion pitches in the *Pretest Singing Stimulus* and the *Posttest Singing Stimulus* was the *Sona-Speech Model 3600* software program by Kay Elemetrics.

The facility was an air-conditioned, relatively sound-proof room. The testing room was adjacent to the music room.

IMPLEMENTATION

Duration of the Study

The total duration of the testing and treatment portions of the study was twelve weeks from the *Pretest Singing Stimulus* administration to the *Posttest Singing Stimulus* administration. The time period to complete the experimental treatment on all thirty subjects lasted no longer than three weeks. Each subject in the treatment group received one forty-five-minute treatment session using the Yuba Method. It took two weeks to administer the *Pretest Singing Stimulus* and another two weeks to administer the *Posttest Singing Stimulus*.

All subjects in grades four, five, and six in the public elementary school were taught to sing the *Pretest Singing Stimulus* and *Posttest Singing Stimulus* one month prior to the testing. This was done as a whole class in their regular fifty-minute, once-a-week general music class. This ensured that all students had fairly equal exposure to the test stimuli. Subjects were taught to sing the stimuli rather than using "familiar songs" as in previous research. Teaching of the stimuli reduced the possibility of some subjects singing inaccurately due to unfamiliarity of the song material. The *Pretest Singing Stimulus* and the *Posttest Singing Stimulus* were taught in the typical rote song teaching approach. The whole to part song teaching method was also utilized where the instructor sang the whole song to the subjects and then had the subjects echo back phrase by phrase until they learned the song and then sang the song as a whole song. The *Pretest Singing Stimulus* and *Posttest Singing Stimulus* were reviewed at each subsequent class meeting prior to the testing times for a period of four weeks.

Regarding the Singing Stimuli, the researcher chose two different songs--*Shalom Chaverim*, first phrase, and *The Star-Spangled Banner*, first phrase. Two different singing stimuli were chosen due to past research findings that mistakes were often carried over in the same song regardless of training (Goetze 1985). Goetze (1985) noted that it was more discriminating to evaluate selected pitches rather than every note in the song or phrase. The criterion pitches selected for this study were D4, F4, C5, and D5 for both test stimuli. The criterion pitches were selected by the investigator to span the tones across the vocal register break (D4-D5), a tone below the register break (D4), and a tone around the register break (F4). A song phrase was used rather than utilizing the matching of single tones because past research indicated that the matching of single pitches had no correlation to singing a song in tune (Flowers and Dunne-Sousa, 1990; 111).

The range of the song passages was selected for the study based on the following. The vocal register break had been identified in past research as a possible cause of singing inaccuracy (Cooper, 1992; 36). Registration was found to affect range (Brown, 1988), in that first, third, and fifth graders who were aware of, and could use their head voices, had significantly wider ranges and higher mean pitches within their ranges. The break had been identified in past research (Joyner, 1971; Wassum, 1979) as usually occurring between G4 and A4 in students of this age. Singing up to D5 in this study ensured that students were able to sing above their vocal register break and were thus utilizing both their chest and head voices. The range of the *Pretest Singing Stimulus* and *Posttest Singing Stimulus* in this study was from A3 to D5.

The pretest and posttest stimuli for the present study were sung on the neutral syllable "loo." Previous research (Goetze, 1985) indicated that students sang more accurately on

a neutral syllable. Gould (1969) reported that "initial progress seemed to occur when the words were omitted" (15). Edwin Gordon (1984) recommended that "students must echo in solo with a neutral syllable" (30). Gordon added that "the use of words of a song actually inhibits the learning of tonal syntax" (143). The neutral syllable "loo" was used for the singing stimulus by Goetze, (1985), Smale (1987), and Cooper (1995).

The administration of the *Pretest Singing Stimulus* and the *Posttest Singing Stimulus* included the playing of the first two pitches of the stimulus followed by the a capella singing by the subject. Two pitches were selected to be played by the investigator based on personal judgment from the instructional period of teaching the *Pretest Singing Stimulus* that students were able to sing more accurately when given the first two starting pitches verses one, or more pitches which might overload tonal memory. A script was read for the administering of the *Pretest Singing Stimulus* and the *Posttest Singing Stimulus* (see Appendix III).

All students in grades four, five, and six in the public elementary school (N = 320) who had at least seventy-five percent of the instruction of the *Pretest Singing Stimulus* were administered the *Pretest Singing Stimulus*. No students were excluded on this basis from this study. Cooper (1992) found that vocal pitch accuracy scores did not reflect improvement in singing accuracy at each grade level. This included grades three, four, and five. The researcher, therefore, hypothesized that there would be no significant differences in pitch accuracy between the three age groups so subsequently, they were treated as one combined population.

The *Pretest Singing Stimulus* test was recorded using the *Sona-Speech* software program installed on a Gateway laptop PC computer for analysis. Students were

recorded using an Electro Voice Microphone held approximately one inch away from the subject's mouth. The volume level on the Gateway laptop PC computer was preset at approximately seventy-five percent of the input capacity. At this volume setting, there were no problems with subjects being inaudible due to inadequate or overloaded amplitude levels.

The method administered was from *Dream Voice Training* (Yuba, 1998). The method outlined in the book was designed to be a self-study for inaccurate singers (see Appendix IX). The accompanying instructional CD was used during the experimental training process. These instructions consisted of synthesized instrumental accompaniments with vocal examples to echo. The recording was recorded and produced by *Female Voice* Saito Masahiro Musical Arrangement & Production Corporation, Victor Yamanaka-ko Studio, 1998. The Yuba Method was translated into English by Shawn Yacavonne from the *Center for Interpretation and Translation Studies* at the University of Hawai'i at Manoa. This translation in turn was adapted and simplified for use by elementary students by the investigator. Although the Yuba Method vocal exercises remained intact, the instructions prior to each exercise were simplified for age appropriate terminology, and condensed to accommodate for the relatively shorter attention spans of the subjects in the study as compared to adults.

All subjects in the treatment group were video taped during the treatment session for possible verification of the method at a later time and record keeping purposes. The video camera used was a *Sony Hi-8 Video Recorder*.

Subjects in the control group took the *Posttest Singing Stimulus* at the end of the three week treatment period. The subjects in the treatment group took the *Posttest Singing Stimulus* immediately following the forty-five minute vocal training session.

Statistical Design and Analysis Procedure

To determine whether or not the modified Tora Yuba training improved singing ability of treatment subjects, a statistical design was required that would provide for Analysis of Variance (ANOVA). The statistical design and tests employed assessed the effect of the various sources of variability for the design. The structure of the analysis of variance used is shown in Figure 20. This structure shows the analytical procedure as computed by the SAS statistical software program. The design consisted of a two-way classification, with the sources of variation being (1) the effect of the Yuba Method (TY) training, and (2) the pretest ranking (R) of inaccurate subjects into the "low," "middle," and "high" subgroups. The analysis also provides an assessment of the variation contributed by the interaction of "main effects" (1) and (2) defined above, as indicated by the "TY x R" source of variation in Figure 19.The labeling system in for ANOVA of *Pretest Singing Stimulus* VPA scores, as used in Figure 20, is as follows.

1. "X_{ijk}" represents the measured result for the individual student.
2. "i = 1" represents the Yuba Method treatment group.
3. "i = 2" represents the control group.
4. "j = 1, 2, 3" represents the *Pretest Singing Stimulus* levels of "low," "middle," and "high."

5. "k = 01, 02, … 10" represents the repetitions for each subgroup ("low" treatment, "low" control, "middle" treatment, "middle" control, "high" treatment, and "high" control).

6. The system of notation adhered to in Figure 20 is the following as described by Ostle (1954), with "X" here substituted for "Y" in Ostle:

A sum of terms, such as $X_1 + X_2 + X_3 + X_4$, is designated by:

$$\sum_{i=1}^{4} X_i. \tag{19}$$

When used in this way, "Σ," the Greek capital letter "sigma," is called the "summation sign" and the subscript "i" is called the "index of summation." The "i = 1" below "Σ" indicates that the first term in the sum is found by giving "i" the value of "1." The "4" above "Σ" indicates that the last term in the sum is found by giving "i" the value of "4." The other terms are found by giving "it" the integral values between "1" and "4." In general:

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + \ldots + X_n.$$

It is often convenient to designate observations by symbols such as "X_{ijk}." For example, observations on the yield of a certain crop may be classified according to three different criteria: (i) variety, (ii) place, and (iii) fertilizer used. The letter "i" could then be associated with the varieties, "j" with the locations, and "k" with the fertilizers. If in an experiment four locations, seven varieties, and three fertilizers were involved, then "i = 1, … , 7; j = 1, … , 4; k = 1, 2, 3." In this case, "X_{512}" would represent the yield of the fifth variety at the first location when the second fertilizer had been applied. The period (.) in subscript denotes a summation over the appropriate number of experimental units. The "Σ" notation is a great saver of time and space, not to mention effort (20).

Figure 20. Analysis of Variance for the Experimental Design.

Source of Variation	*df*	Sums of Squares
Total	59	$\Sigma X^2_{ijk} - C$ $i=1,2$ $j=1,2,3$ $k=1,2,\ldots,10$
TY Impact (TY)	1	$[\Sigma_{i=1,2} X_{i..}]^2/30 - C$
Pretest Ranking (R)	2	$[\Sigma_{j=1,2,3} X_{.j.}]^2/20 - C$
TY x R	2	$[\Sigma_{i=1,2;\ j=1,2,3} X_{ij.}]^2/10 - C$
Error	54	Total – TY – R – (TY x R)

Note: $C = [\Sigma X_{ijk}]^2/N$, where N = 60

*The source of variation "Total" with *df* "59" was the idealized description but it was changed during the course of the study due to three extreme values being dropped. The saAS Proc Univariate Program accommodates the loss of a few "missing values" in ANOVA without invalidating the statistical tests.

CHAPTER IV

RESULTS

This study investigated the effects of the Yuba Method as compared with a control group and between the subgroups of "low," "middle," and "high" VPA singing accuracy based on the *Pretest Singing Stimulus* vocal pitch accuracy scores (VPA) for students classified as inaccurate singers as defined in Chapter Three. The instruction given to the treatment group consisted of the Yuba Method exercises administered to individual subjects at one forty-five-minute treatment session. The control group received no remedial training other than what was included in their regular once a week, fifty-minute general music class.

THE PRETEST SINGING STIMULUS

Results of the *Pretest Singing Stimulus* are summarized in Table 3 by grade level and gender.

Table 3. Percentage and Number of Accurate and Inaccurate Singers by Grade Level and Gender Based on *Pretest Singing Stimulus* VPA Scores.

Grade (N), Gender (N)	Accurate Singers % (N)	Inaccurate Singers % (N)
Grade 4 (104)	40.39 (42)	59.61 (62)
Boys (51)	39.22 (20)	60.78 (31)
Girls (53)	41.51 (22)	58.49 (31)
Grade 5 (113)	47.79 (54)	52.21 (59)
Boys (64)	45.31 (29)	54.68 (35)
Girls (49)	51.02 (25)	48.97 (24)
Grade 6 (103)	54.37 (56)	45.63 (47)
Boys (50)	71.00 (23)	24.27 (27)
Girls (53)	62.26 (33)	19.41 (20)
All (320)	47.50 (152)	52.50 (168)
Boys (165)	43.64 (72)	56.36 (93)
Girls (155)	51.61 (80)	48.38 (75)

Figure 21 shows the *Pretest Singing Stimulus* frequency distribution of VPA scores for the sample population (N = 320).

Figure 21. *Pretest Singing Stimulus* Frequency Distribution of VPA Scores for the Sample Population.

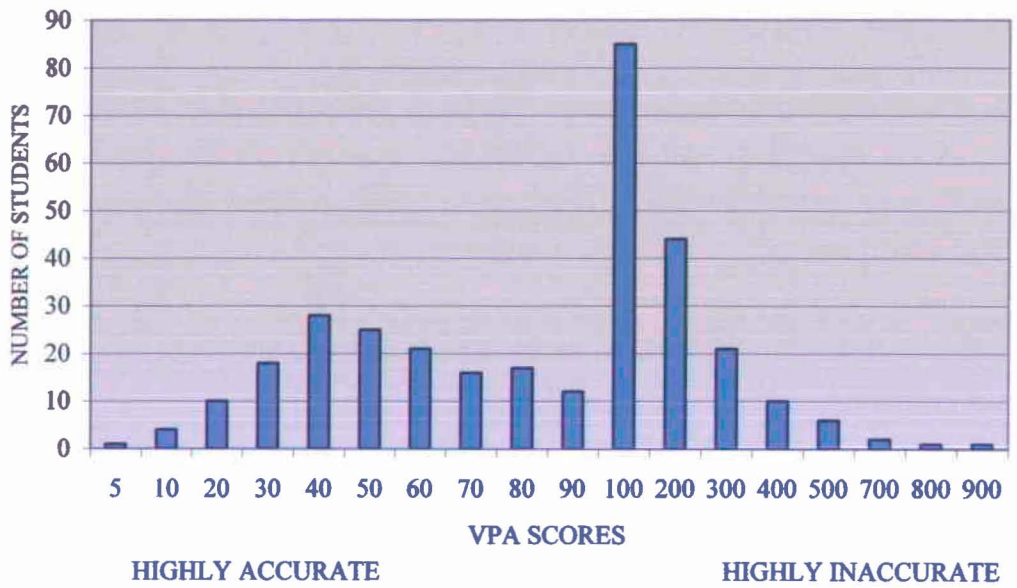


Figure 22 shows *the Pretest Singing Stimulus* frequency distribution of VPA scores for all inaccurate singers (N = 168).

Figure 22. *Pretest Singing Stimulus* Frequency Distribution of VPA Scores for all Inaccurate Singers.

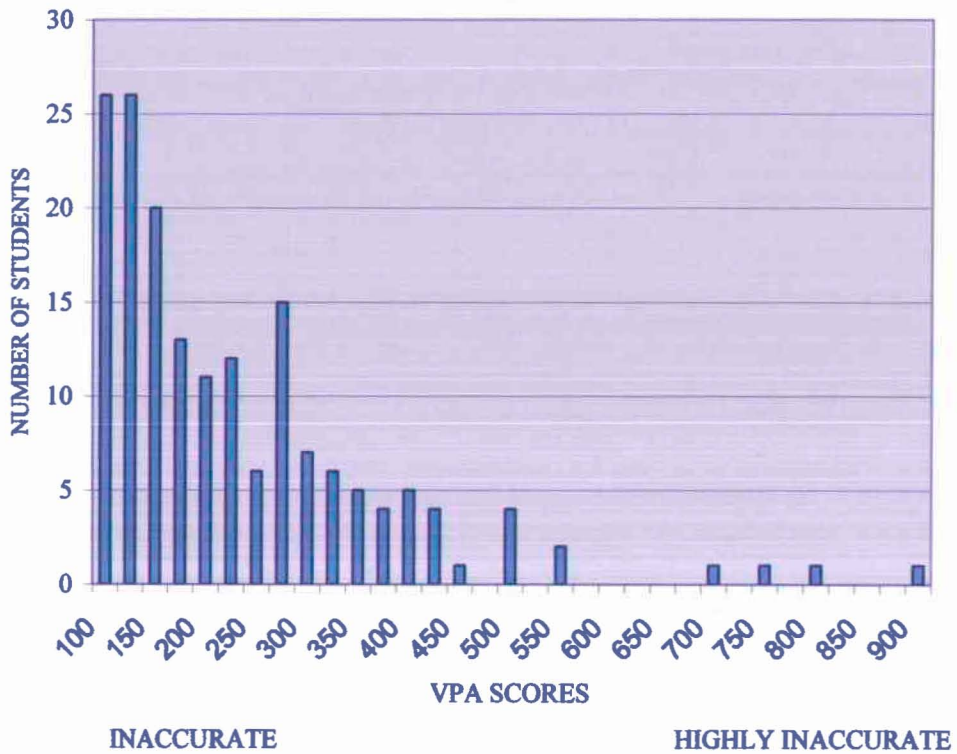
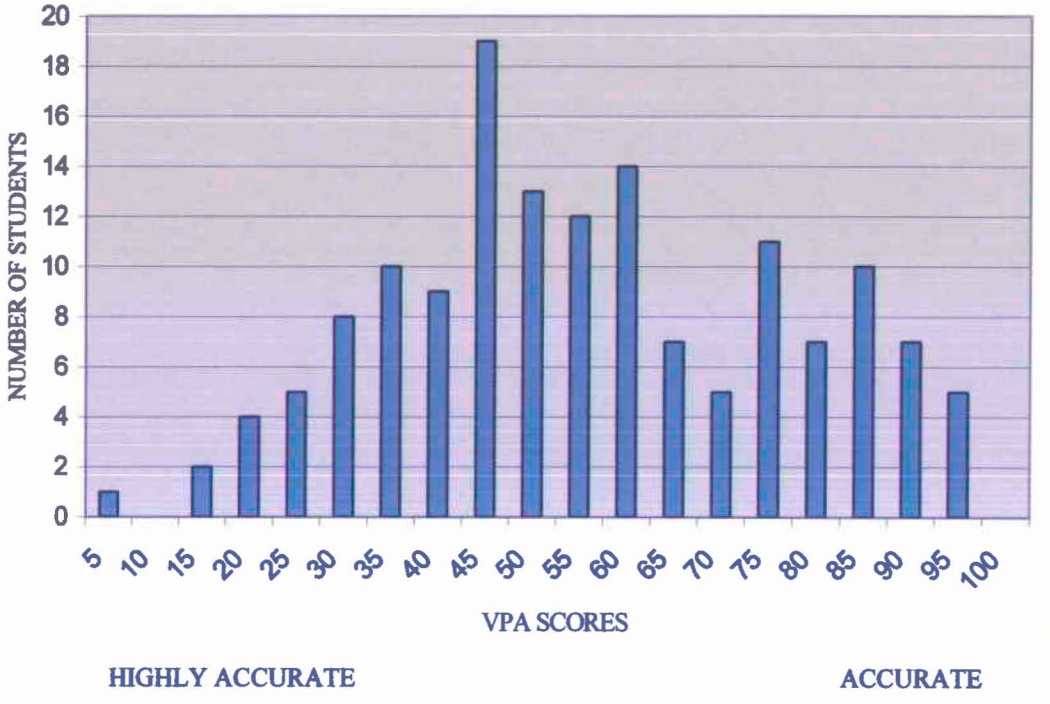


Figure 23 shows the *Pretest Singing Stimulus* frequency distribution of all accurate singers (N = 152).

Figure 23. *Pretest Singing Stimulus* Frequency Distribution of all Accurate Singers.



The range of the sample population for the *Pretest Singing Stimulus* vocal pitch accuracy scores was 5.26 to 957.04 cents. Data for the *Pretest Singing Stimulus* VPA score means and standard deviation measured in cents are summarized in Table 4.

Table 4. *Pretest Singing Stimulus* VPA Score Means and Standard Deviation Measured in Cents.

	SD	Mean
Sample Population	139.07	154.40
Accurate Singers	21.93	57.33
Inaccurate Singers	142.03	242.23

THE POSTTEST SINGING STIMULUS

The researcher-designed *Posttest Singing Stimulus* results yielded the data that represents the mean number of cents subjects deviated from all four of the selected criterion pitches. High scores indicate highly inaccurate singing, and low scores indicate more accurate singing. The scores for each subgroup are given in Table 5. The mean VPA score for the "low" treatment subgroup was 76.24. The mean VPA score for the "low" control subgroup was 117.96. The mean VPA score for the "middle" treatment subgroup was 109.39. The mean VPA score for the "middle" control subgroup was 184.86. The mean VPA score for the "high" treatment subgroup was 84.90. The mean

VPA score for the "high" control subgroup was 301.74. Most subjects, in both groups, treatment and control, improved VPA scores on the *Posttest Singing Stimulus*.

Table 5. Mean *Posttest Singing Stimulus* VPA Scores between Groups and Subgroups.

	Low	Middle	High
Treatment	76.24	109.39	84.90
Control	117.96	184.86	301.74

DATA ANALYSIS

In order to determine the effects of the Yuba Method subjects' *Pretest Singing Stimulus* VPA scores were compared with their corresponding *Posttest Singing Stimulus* VPA scores. Gain scores were computed by subtracting a subject's *Posttest Singing Stimulus* VPA score from *the Pretest Singing Stimulus* VPA score. Positive gain scores represented an increase in singing accuracy. Negative gain scores represented a decrease in singing accuracy. Pretest-posttest gain scores were computed for each of the two groups and a double classification analysis of variance was computed using the mean gain scores to determine if there was a significant difference in the performance of subjects in the treatment group versus the control group ($p < .05$).

Two subjects in the treatment group received negative gain scores. Ten subjects in the control group received negative gain scores. The mean gain score for the treatment group

was M = +148.90. The mean gain score for the control group was M = +24.63. The mean gain score for the treatment low group was M = +51.27. The mean gain score for the treatment middle group was M = +82.91. The mean gain score for the treatment high group was M = +312.52. The mean gain score for the control low group was M = +2.30. The mean gain score for the control middle group was M = +7.32. The mean gain score for the control high group was M = +64.26 (see Table 6).

Table 6. Mean *Posttest Singing Stimulus* VPA Gain Scores between Groups and Subgroups.

	Low	Middle	High
Treatment	+51.27	+82.91	+312.52
Control	+ 2.30	+ 7.32	+ 64.26

In analyzing the gain scores, some of the VPA gain scores turned out to be negative numbers due to a decrease in singing accuracy on the *Posttest Singing Stimulus*. To compensate for this, 300 cents were added to all of the VPA gain scores for the calculations. Moore and McCabe (2003) explained that converting numerical descriptions of a distribution from one unit of measurement to another is a linear transformation of the measurements (51). They explained that linear transformations do not change the shape of a distribution (53). After adding 300 cents to all of the VPA gain scores, a negative number was still present in subject 52102. This was a fifth grade subject in the "low" control group. The subject went from a VPA score of 109.53 on the

Pretest Singing Stimulus to a VPA score of 432.45 on the *Posttest Singing Stimulus*. It was not desirable to add more than 300 cents to all of the VPA gain scores as then it would tend to overshadow the original scores. Because you cannot take the log of a negative number for the log transformation for normality, it was decided to drop the score of subject 52102.

In order to determine whether or not to employ parametric or nonparametric statistical procedures, tests for normality of the sampled population were calculated. Rainbow and Froehlich (1987) stated that parametric statistics are more powerful than nonparametric tests. They defined "powerful" in statistical terms to mean that a test discriminates between two sets of data in such a way that the null hypothesis may be rejected even if the differences in scores are relatively small. They further explained that because researchers are concerned about minimizing the probability that a null hypothesis is maintained when it is in fact false, the more powerful statistic should be given preference. Rainbow and Froehlich concluded that when presented with the choice of using parametric versus nonparametric tests in a research situation, parametric tests should be employed (256).

According to Rainbow and Froelich (1987) the normality in regards to tests for normality of population is based on the normal curve first introduced by Francis Galton in the late 1860s. Galton was the first to urge the application of the logical principles of mathematics to the observation of human behavior. The normal curve serves as a mathematical model that describes how repeated independent measurements of an infinite number of characteristics of objects may be distributed over a continuum (165-166). Moore and McCabe (2003) defined the normal distribution curve as a mathematical

model for the distribution which is an idealized description. Ostle (1954) stated that the normal population is an example of a theoretical population which has a symmetrical distribution in which the mean, median, and mode are all the same (23). The tests of normality were important because otherwise the statements about the probability were not likely to be true (63). They explained that the decision to describe a distribution by a normal model determines the later steps in the analysis of the data (78).

The scores were tested for normality by the *SAS Statistical Software Program*. Both the *Shapiro-Wilk* and the *Kolmogorov-Smirnov* tests of normality indicated that the raw scores, VPA, were non-normally distributed at $p < .0001$ and $p = .01$, respectively. Figure 24 shows the non-normality of the 59 observations of the raw VPA gain score data as shown by stem leaf and box plots from the SAS Proc Univariate program. This is an example of skewed data—with a positive skew towards extremely large values. The extreme non-normality justified efforts and concerns to find a transformation that would give a normal distribution of the data. Snedecor (1956) explained that the technique of transforming data to achieve a normal, Gaussian, distribution is the recommended procedure in experimental statistics when standard F or t tests are to be applied to determine the significance of differences between treatments in experiments designed for analysis of variance. Snedecor stated that logarithms are used to transform data if effects are known to be proportional instead of additive (320).

Figure 24. Non-normality of the 59 Observations of VPA Gain Score Data as Shown by Stem Leaf and Box Plots from SAS PROC UNIVARIATE.

```
Stem Leaf                     #             Boxplot
   7 6                        1                0
   7
   6
   6
   5
   5 12                       2                |
   4                                           |
   4 012                      3                |
   3 567778                   6                |
   3 0123                     4             +-----+
   2 8889                     4             |     |
   2 11122224                 8             |  +  |
   1 5666777788999           13             *-----*
   1 1111122222233333344     19             +-----+
     ----+----+----+----+
 Multiply Stem.Leaf by 10**+2
```

Moore and McCabe (2003) explained that a stemplot (also called a stem-and-leaf plot) gives a quick picture of the shape of a distribution while including the actual numerical values in the graph. They added that the stem-leaf plot is a short-hand way of looking at a histogram of the data (10). Moore and McCabe stated that a stemplot can reveal distinctly nonnormal features of a distribution, such as outliers. They added that outliers appear as points that are far away from the overall pattern of the plot (79).

Stemplots can be interpreted as follows using the VPA data in Figure 24 as an example:

1. The first number of the stem is "7" in Figure 24. This number is multiplied by "100" as indicated in the note at the bottom of the figure by "10**+2."
2. The resulting number is "700." The leaf next to stem "7" in Figure 24 is "6." This represents the VPA score of "760."
3. Stemplot numbers in this case are rounded to two digits.
4. Using this same procedure, the next scores in the stemplot in Figure 24 are "510" and "520," respectively (10).

Boxplots can be interpreted as follows using the same VPA data in Figure 24 as an example:

1. The boxplot in Figure 24 shows virtually the same thing as the stemplot in a different format.
2. The "+" sign in the boxplot indicates the "mean" of the VPA, while the "*-----*" line represents the median of the VPA.
3. The "+-----+" above and below both the mean and median indicate "hinges."
4. "Hinges" relate to the median and indicate twenty-five percent or an interquartal range of the data (10).

The gain scores in the present study were then "log e" transformed. Box and stem-leaf plots indicated that even when VPA scores were log-transformed they were non-normal. Figure 25 shows the vastly improved distribution of the data after the log-transformation, though still non-normal, and also points out the two troublesome observations at the bottom of the figure. Those two observations determined that the log transformation was insufficient to normalize the data.

Figure 25. Non-normality of the 59 Observations of Log-Transformed VPA Gain Score Data as Shown by Stem Leaf and Box Plots from SAS PROC UNIVARIATE.

```
Stem Leaf                     #             Boxplot
   5 135                      3                0
   4 49                       2                0
   3 5                        1                |
   2 0456                     4                |
   1 0122234557              10             +-----+
   0 1222333455566899        16             *--+--*
  -0 64333110                8             |     |
  -1 83320                   5             +-----+
  -2 5110                    4                |
  -3 0                       1                |
  -4 9                       1                0
  -5 60                      2                0
  -6
  -7
  -8
  -9 1                       1                *
 -10
 -11
 -12
 -13 4                       1                *
     ----+----+----+----+
 Multiply Stem.Leaf by 10**-1
```

The residual, or error term for testing for significance of the treatment included the six means for the log-transformed gain scores, for each group versus the treatment, the three overall means for group differences, and the two means for the treatment versus control. The non-normality was traced to two outlier data points belonging to subject 41210 (a fourth grade subject in the "middle" treatment group) and 62308 (a sixth grade subject in the "high" control group).

An outlier is defined by Moore and McCabe (2003) as an important kind of deviation which value falls outside the overall pattern (12). They explained that the handling of outliers is a matter for judgment and that they can point to the unusual nature of some observations. Moore and McCabe added that outliers in research studies can result from equipment failure, or errors in recording the data. They said that the outlier may be evidence of an extraordinary occurrence or just a mistake that went unnoticed. It was finally recommended that if equipment failure or some other abnormal condition caused the outlier, it can simply be deleted (17). In the present study, the researcher noted that the two outlier data points belonged to subjects whose characteristics deviated from the average. The researcher noted that one subject had a poor attention span. This behavior was verified by the regular classroom teacher--that the student took a long time to assimilate information and was not a relatively quick learner. The second subject eliminated also was noted by the researcher to have a poor attention span and this was also verified by the classroom teacher, who reported that the student also had behavior problems in the regular classroom. The investigator chose to remove the two outlier data points of subjects 41210 and 62308. After those points were removed, the residuals from the *Analysis of Variance* of transformed data were normal. That is, both the *Shapiro-Wilk* and the *Kolmogorov-Smirnov* tests were non-significant at the $p = .08$ and $p > 0.15$ levels, respectively. Based on the aforementioned data, parametric statistics were employed for the analysis. The ANOVA was calculated and results are reported in Tables 7-10.

Table 7 presents the ANOVA *General Linear Model Procedure* of the log-transformed scores with 300 cents added to each score and the two abnormal scores eliminated.

Table 7. The General Linear Model ANOVA Results for Log-Transformed VPA Scores with 300 Cents Added (N = 57)

Source	*df*	SS	Mean Square	*F*	*p*
Treatment	1	1.14	1.14	21.14	<.0001
Group	2	1.29	0.64	11.90	<.0001
Group x Treatment	2	0.43	0.22	4.01	0.0241

Table 8 presents the VPA log-transformed gain score by level of treatment with fifty-seven observations.

Table 8. VPA Log-Transformed Gain Scores by Group.

Level of Treatment	N	Mean	SD
Treatment (1)	29	6.09	0.27
Control (2)	28	5.81	0.30

Table 9 presents the data of the log-transformed VPA gain mean scores by level of group with fifty-seven observations.

Table 9. Log-Transformed VPA Gain Mean Scores by Group Level.

Level of Group	N	Mean	SD
High	19	6.17	0.38
Middle	19	5.84	0.28
Low	19	5.84	0.11

Table 10 presents the data of the log-transformed general linear model procedure by level of group and level of treatment with 300 cents added to each individual gain score. The equal values of the "middle" subgroup mean and the "low" subgroup mean produced unequal values for the back-transformed VPA gainscores in Table 13 due to the unequal SD values.

Table 10. General Linear Model Procedure on Log-Transformed Scores with 300 Cents Added by Group.

Group	N	Mean	SD
High Treatment	10	6.39	0.23
High Control	9	5.92	0.38
Middle Treatment	9	6.01	0.09
Middle Control	10	5.69	0.31
Low Treatment	10	5.86	0.10
Low Control	9	5.82	0.13

The "log e" transformed differences were then back transformed to make the data meaningful in terms of comparisons in cents. The following back transformation formula (Haan, 1977; 107) was used for the mean log-transformed VPA scores:

$$e^{(\mu y + \sigma y^2/2)} - 300$$

Where "μ" equals the mean of the transformed data, and "σy^2" equals the variance of the transformed data. Tables 12 through 14 provide the back transformed mean scores for the General Linear Model log-transformed scores with 300 cents added to each score.

Table 11. Back-Transformed VPA Gain Mean Scores by Level of Treatment Measured in Cents.

Level of Treatment	N	Mean
Treatment	29	157.24
Control	28	46.82

Table 12. Back-Transformed VPA Gain Mean Scores by Level of Group Measured in Cents.

Level of Group	N	Mean
High	19	212.65
Middle	19	57.95
Low	19	45.61

Table 13. Back-Transformed VPA Gain Mean Scores by Group Measured in Cents.

Group	N	Mean
High Treatment	10	309.66
High Control	9	100.45
Middle Treatment	9	109.12
Middle Control	10	10.16
Low Treatment	10	52.31
Low Control	9	38.36

Table 14 presents the VPA gain scores by group, log-transformed and back-transformed as a means for comparison.

Table 14. VPA Gain Scores by Group, Log-Transformed and Back-Transformed.

Level of Treatment	N	Log-Transformed Mean	Back-Transformed Mean
Treatment (1)	29	6.09	157.24
Control (2)	28	5.81	46.82

Table 15 presents the VPA gain mean scores by group level, log-transformed and back-transformed as a means for comparison.

Table 15. VPA Gain Mean Scores by Group Level, Log-Transformed and Back-Transformed.

Level of Group	N	Log-Transformed Mean	Back-Transformed Mean
High	19	6.17	212.65
Middle	19	5.84	57.95
Low	19	5.84	45.61

Table 16 presents the general linear model procedure on log-transformed scores and back-transformed scores as a means for comparison.

Table 16. General Linear Model Procedure, Log-Transformed and Back-Transformed Scores.

Group	N	Log-Transformed Mean	Back-Transformed Mean
High Treatment	10	6.39	309.66
High Control	9	5.92	100.45
Middle Treatment	9	6.01	109.12
Middle Control	10	5.69	10.16
Low Treatment	10	5.86	52.31
Low Control	9	5.82	38.36

MAIN EFFECTS AND INTERACTIONS

Based on the ANOVA results, the null hypothesis that there will be no significant differences between the treatment and control groups was rejected at the $p < .0001$ significance level. The null hypothesis that there will be no significant differences between the subgroups of "low," "middle" and "high" was also rejected at the $p = .0024$ significance level. The main effect of treatment was found to be highly significant at $p < .0001$ significance level for all groups.

In addition, some sub-groups benefited from the treatment more than others. The treatment "high" group benefited the most, followed by the treatment "middle" group, and lastly the treatment "low" group.

Based on *Posttest Singing Stimulus* VPA individual score data (see Tables 17 to 22), some of the subjects became accurate singers, in other words, their VPA scores were under 100. Eight out of the ten of the treatment "low" subgroup subjects became accurate singers. Six out of the nine of the subjects in the "low" control subgroup became accurate singers. Six out of the nine subjects in the "middle" treatment subgroup became accurate singers. None of the ten subjects in the "middle" control subgroup became accurate singers. Six of the ten subjects in the "high" treatment subgroup became accurate singers. One of the nine subjects in the "high" control subgroup became an accurate singer.

Table 17. Low Treatment Subgroup Pretest and Posttest VPA Scores.

Subject Number	Pretest VPA Score	Posttest VPA Score	Difference
61101	115.12	45.93	+69.19
61102	117.48	34.98	+82.50
51103	118.21	77.28	+40.93
61104	119.16	38.97	+80.19
41105	123.54	58.79	+64.75
51106	130.33	123.09	+7.24
61107	130.71	72.65	+58.06
51108	137.35	64.44	+65.06
61109	138.54	155.45	-16.91
51110	152.44	90.78	+61.66

Table 18. Low Control Subgroup Pretest and Posttest VPA Scores.

Subject Number	Pretest VPA Score	Posttest VPA Score	Difference
62101	106.42	91.94	+14.48
52103	109.62	31.72	+77.90
52104	111.70	33.35	+78.35
42105	114.21	153.78	-35.57
62106	123.05	78.77	+44.28
52107	130.03	63.41	+66.62
62108	130.75	51.65	+79.10
62109	131.12	134.97	-3.85
62110	132.15	107.59	+24.56

Table 19. Middle Treatment Subgroup Pretest and Posttest VPA Scores.

Subject Number	Pretest VPA Score	Posttest VPA Score	Difference
61201	161.64	81.11	+80.53
51202	162.62	49.41	+113.21
41203	172.98	50.32	+122.66
51204	173.25	54.17	+119.08
61205	181.51	32.62	+148.89
51206	189.45	117.70	+71.75
51207	210.56	135.66	+74.90
61208	211.88	135.73	+76.15
41209	216.35	42.56	+173.79

Table 20. Middle Control Subgroup Pretest and Posttest VPA Scores.

Subject Number	Pretest VPA Score	Posttest VPA Score	Difference
42201	162.83	175.74	-12.91
52202	163.63	123.78	+39.85
52203	171.57	121.28	+50.29
42204	177.04	297.84	-120.80
52205	185.08	304.34	-119.26
42206	189.68	247.76	-58.08
42207	210.13	124.97	+85.16
42208	218.93	141.33	+77.60
52209	219.22	208.78	+9.44
52210	224.77	102.82	+121.95

Table 21. High Treatment Subgroup Pretest and Posttest VPA Scores.

Subject Number	Pretest Score	Posttest Score	Difference
51301	276.37	42.00	+234.37
51302	276.84	149.83	+127.01
61303	289.13	26.11	+263.02
51304	315.85	26.72	+289.13
61305	347.65	54.39	+293.26
51306	369.30	36.32	+332.98
41307	382.18	186.40	+195.78
51308	450.45	178.24	+272.21
51309	510.89	123.65	+387.24
61310	755.47	25.32	+730.15

Table 22. High Control Subgroup Pretest and Posttest VPA Scores.

Subject Number	Pretest Score	Posttest Score	Difference
42301	282.12	85.76	+196.36
42302	282.47	228.14	+ 54.33
42303	310.62	272.92	+ 37.70
62304	333.33	107.69	+225.64
42305	355.86	108.12	+247.74
62306	368.52	135.07	+233.45
42307	374.50	413.23	- 38.73
62309	420.61	537.38	-116.77
42310	518.71	499.92	+ 18.79

CHAPTER V

SUMMARY, DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

The main effect of treatment was found to be highly significant at the $p < .0001$ significance level. In addition, some subgroups benefited from the treatment more than others. The treatment "high" subgroup (the most inaccurate singers) benefited the most, followed by the treatment "middle" subgroup, and lastly the treatment "low" subgroup.

The null hypothesis that there will be no significant differences between the treatment and control groups was rejected at the $p < .0001$ significance level. The null hypothesis that there will be no significant differences between the subgroups of "low," "middle" and "high" was also rejected at the $p = .002$ significance level.

The results of this study indicated that the Yuba Method, which targeted training of the cricothyroid muscle, was highly effective in improving the vocal pitch accuracy of inaccurate elementary singers. The treatment was also most effective with highly inaccurate singers. All of the treatment group subjects improved except for two, who became more inaccurate. The investigator noted that those two subjects had poor attention spans and also classroom behavior problems as confirmed by the classroom teachers. It could be that those two subjects' singing inaccuracy was due to problems other than physiological reasons.

The impact of this study is that subjects improved in vocal pitch accuracy through the Yuba Method, which focused on the physical training of the cricothyroid muscle. The training was also completed relatively quickly, in one forty-five minute training session.

The arbitrary VPA of 100 cents or greater to define the inaccurate singer, as selected by Goetze (1985) and utilized in this study, needs to be confirmed through further empirical research. Cooper (1995) recommended that a study comparing subjective ratings of perceived accuracy with objective electronic accuracy evaluations of the same sample be conducted (230). Pilot test results prior to this study indicated that seven subjects in the sample population who were selected to be in the *American Choral Directors Association Western Division Honor Choir* had VPA scores ranging from 18.55 to 125.00. This seems to indicate that a higher VPA might be utilized to indicate the inaccurate singer.

Although different singing stimuli have been used in previous research, comparing the accuracy of subjects in this study with findings reported in previous research using the VPA of 100 cents or greater to define the inaccurate singer is noteworthy. The following can be reported. A total forty-eight percent of subjects in this study were designated accurate singers. Smale (1987) reported eighteen percent of subjects to be accurate singers. Goetze found thirty percent to be accurate.

In the present study, there was a tendency for singing accuracy to increase as students got older at each successive grade level. A total of forty-one percent of fourth graders were accurate singers. A total of forty-eight percent of fifth graders were accurate singers. A total of fifty-four percent of sixth graders were accurate singers. This contrasts with the findings of of several studies. Cooper (1995) found "no trend toward improvement in VPA at successive grade levels" (229). Wilson (1970) found that children's voices developed at different rates with marked individual differences at grade levels one through six. Sims, Moore, and Kuhn (1982) found no significant differences

between five-year-olds and six-year-olds in singing ability. Smale (1987) found that four-year-olds and five-year-olds were similar in vocal pitch accuracy. Aaron (1991) did not find a general trend or an increase in singing accuracy in grades four through six. Levinowitz, Barnes, Guerrini, Clement, D'April, and Morey (1998) found no significant difference in children's singing voices in grades one through six.

The findings of this study agree with several others. Petzold (1966) found that improvement in singing accuracy took place at a constant rate over a four-year period. Petzold, however, found that only two-year intervals produced significant differences. The results of the Goetze study (1985) indicated a general trend towards increased singing accuracy with age for subjects in kindergarten, first grade, and third grade. Green (1994) found that singing improved with each successive grade level. The present study did not test for significant improvement in singing accuracy from grade level to grade level, however, the general trend of improvement at each successive grade level suggests that students were continually increasing in singing accuracy each year. This may have been due to classroom music instruction or, perhaps, maturation.

Most subjects, in both the control and treatment groups performed better on the *Posttest Singing Stimulus*. This may be due to test anxiety, and being more familiar with the testing situation. This could also be that th*e Posttest Singing Stimulu*s was a more familiar song—*The Star-Spangled Banner*, which students frequently sang after the morning opening exercises. Further research could use a different stimulus or test between the two singing stimuli to see, if in fact, they are of equal difficulty. The two singing stimuli used for this study utilized the same criterion pitches for analysis,

however, the levels of singing ability necessary may have been different between the two. For example, the *Pretest Singing Stimulus* had a slightly greater pitch range (A3 to D5) than did the *Posttest Singing Stimulus* (Bb3 to D5). The *Pretest Singing Stimulus* also tended to move more by step and with smaller intervals, the largest being a perfect fourth, than did the *Posttest Singing Stimulus*, where the largest interval was a minor sixth and interval movement were mostly minor third skips. Based on the aforementioned, the *Posttest Singing Stimulus* was probably the more difficult of the singing stimuli to sing.

Similar to the other research studies, the *Visi-Pitch* and the *Sona-Speech* were effective, but time consuming methods. Similar observations were noted in studies which used the Visi-Pitch (Goetze, 1985; Smale, 1987; Goetze & Horii, 1989; Cooper, 1995; Clayton, 1986; Brown, 1988; Aaron, 1991; and Collins, 2000). One of the benefits of using the *Sona-Speech* was that there was no need to use a tape recorder and then feed the taped recording into a computer to patch into the *Visi-Pitch* hardware system. It was a direct recording.

Extant research on remedial methods has had some significant results in terms of improving singing accuracy. Roberts and Davies (1975) investigated monotones and remedial training. They found that it was possible "to effect some improvement in pitch production among children rated by their teachers as monotones" (236). Roberts and Davies utilized speech devices to extend the speaking range and exercises in finding a personal note. Unlike the Yuba Method, their study did not take into account vocal registration, but rather, tried to extend the chest voice register as much as possible. Tone-matching exercises were also utilized by Roberts and Davies to suggest that monotonism

was a result of poor pitch perception. This is in contrast to the Yuba Method, which presumes that inaccurate singing is primarily a result of a poorly functioning cricothyroid muscle. The test of free song used by Roberts and Davies also permitted the subject to sing any chosen song on a personally selected note. This suggests that the test of free song did not necessarily utilize and recognize the head voice register in the singing of songs—a critical component in the Yuba Method to singing simple songs in tune. It was also recognized by Guerrini (2002) that "once students are able to sing one song accurately, using notes above the register break, they appear to be able to sing other songs accurately" (56).

Richner (1976) conducted a study comparing three common methods of elementary music instruction. He found that remedial voice instruction had a significant positive effect on the ability of inaccurate singers to reproduce pitches. The treatment group in which singing was emphasized also had significant, positive effects (56). The treatment employed by Richner differed from the Yuba Method in this study in that it did not specifically target the training of the cricothyroid muscle and recognize the different vocal registers. Richner first utilized finding pitches the student could sing on the piano. Subjects were instructed to sing "more loudly" (38) to attain higher pitches. The exercises concentrated on range expansion and then the singing of song melodies within those given pitches. Once again, the head voice register may not have been utilized in the Richner study, differing from the Yuba Method, which advocates using both the natural and head voice registers.

The physiological treatment of the use of breath control management training was investigated in several studies with mostly positive results. Phillips (1983) investigated

the effectiveness of group breath control training on physical and vocal singing measures. He found that breath control management significantly improved vocal range, vocal intensity, vocal duration, and pitch accuracy (181). The Phillips method differed from the Yuba Method in that it did not target the specific training of the cricothyroid muscle, but instead that of the act of respiration in singing. The Phillips study presumed that there were problems in singing due to poor respiration or breath management. Phillips' treatment was effective and suggests that poor breath management may be a factor of inaccurate singing.

Gackle (1987) examined the effects of selected vocalizes which employed breath management techniques. She found the exercises significantly improved pitch perturbation (116). Gackle however, found no differences in singing range as the result of the use of selected vocal techniques. This differed from the Yuba Method, which targeted increased singing range in both the chest and head registers.

Aaron (1991) employed a treatment that emphasized posture and breathing exercises. He found that vocal coordination instruction was more effective in improving vocal pitch accuracy for boys than for girls and that highly inaccurate boy singers benefited the most from vocal coordination instruction. The Aaron study differed from the Yuba Method. Yuba did not claim that there would be differences between the genders in training results. Aaron had subjects singing on the neutral syllable "ah" to on the *Vocal Range Test*. The Yuba Method specifically states that the syllable "ah" produces the chest voice sound and the syllable "oo" produces the head voice sound.

Phillips and Aitchison (1995) found that highest and lowest pitch of vocal range was improved through breath control management instruction. Subsidiary findings were that

pitch accuracy was greater for girls than for boys and that male pitch accuracy may not be helped with instruction when they are reluctant to sing in the treble range (195). It was not found in this present study that boys were reluctant to sing in the treble range. It seemed to the investigator in this present study that the Yuba Method instructions presented singing in the head voice in a constructive way that did not intimidate boys in using their head voices. The Yuba Method explained to the subjects that they had "two different voices"—both of which were necessary in singing.

Collins (2000) concluded that breath-management may be so interdependent with students' abilities to coordinate the vocal mechanism that it alone may not produce significant results in vocal performance (78). The Collins study suggests that breath management alone may not be sufficient to correct some inaccurate singers. It lends support to the Yuba Method which focuses on coordinating the vocal mechanism—the cricothyroid muscle. While the Yuba Method does employ and utilize concepts of breath control management in the singing exercises, the main focus is on the training of the cricothyroid muscle. Collins' suggestion to combine breath control management technique exercises with that of vocal mechanism coordination may seem optimum, however, this investigator does not recommend combining the two techniques as specifically employed in the Yuba Method and in the Phillips (1983), Gackle (1987), Aaron (1990), Phillips and Aichison (1995), and Collins (2000) studies. The two methods might not successfully be employed in combination for the simple reason that since it took a whole semester to employ the breath management exercises to yield the stated results, it does not seem feasible that two complete methods might be combined in such a way that elementary students could attenuate to all of the details and instructions

necessary. It may be more effectual to utilize the two types of treatments on a separate basis to keep students focused on the necessary tasks.

Kramer (1985)--found that imagery training improved the ability of inaccurate singers to match pitches vocally (87). The Kramer imagery training differed from the Yuba Method in its major premise—that singing inaccuracy was caused by psychological rather than physiological reasons. The Kramer treatment employed physical, visual, and vocal imitation of sounds and objects such as a rainbow, and ghosts flying. The use of bodily movment, imagination games, charts, and pictures were also incorporated. The Kramer study did utilize the use of the head voice as vocal ranges were from C4 to D5. Similar to the Yuba Method, the Kramer method employed the use of the neutral vowel sounds "oo" or "loo" until the melodic material was mastered. The use of text was introduced last as in the Yuba Method. The Kramer study did not focus or recognize the two distinct vocal registers as the Yuba Method did, nor did it focus on the training of the cricothyroid muscle. The Yuba Method did not utilize imagery training in the exercises specifically, but did use the vocal imitation of sounds as Kramer did.

Jarjisian (1981) investigated pentatonic and diatonic pitch pattern instruction. Results showed that young children's rote singing achievement was benefited by pitch pattern instruction which included both diatonic and pentatonic patterns (46). The Jarjisian study suggested that singing inaccuracy was due in part to pitch perception, that students were unable to sing certain patterns due to the fact that they were not previously introduced to them. This differs from the Yuba Method, which does not assume pitch perception to be a major factor in inaccurate singing. The Yuba Method focused on the physiological cricothyroid muscle training.

Apfelstadt (1984) investigated the effects of melodic perception instruction. She found significant differences on vocal pitch-pattern accuracy and in rote singing accuracy (21). The Apfelstadt study again differs from the Yuba Method for similar reasons to that of Jarjisian (1981). The Yuba Method does not deny the factor of melodic perception in singing, but acknowledges that inaccurate singing is due mostly to physiological reasons—mainly the inadequate functioning of the cricothyroid muscle.

Rooks (1987) investigated the effects of remedial vocal training on inaccurate singers. Findings were that restricted range singers trained in both the high and low range gained significantly more accuracy than those trained in only the high range (51). The Rooks study lends support to the Yuba Method in that she concluded that vocal training in both the high range and the low range are necessary for improving the inaccurate singer. The Rooks remedial method differed from the Yuba Method in that it utilized posture and breath control, pitch matching, and tonal memory exercises. The Yuba Method utilizes only training of the cricothyroid muscle exercises and work on the two different vocal registers. Rooks did not address the vocal registers in training, although she addressed two different vocal ranges—low (C4 to A4) and high (A4 to D5), which seemed to correspond to two different vocal registers as in the Yuba Method, there was no attempt to separate, stabilize, and blend the registers as in the Yuba Method.

Matthias (1997) investigated the use of sequential games. Vocal accuracy was said to have improved after completing a sequential series of matching games (66). Treatment exercises employed the use of voice matching games. Curwen hand signs were used to aid the subject responses. Subjects in the Matthias study were reinforced for their correct responses. The Yuba Method similarly used hand motions to indicate lower and higher

pitches during the exercises, but these did not correspond to Curwen hand signs for each pitch of the diatonic scale. Subjects in the Yuba Method were also not specifically told when specific pitches were exactly in tune or "a match." Matthias also did not recognize the specific vocal registers in the treatment exercises. It is also unknown if the Matthias study employed the use of the head voice range in the exercises. Matthias did conclude that the use of games helped to maintain interest during treatment of the subjects. The Yuba Method was written specifically for adults and in contrast to the Matthias study, may not have seemed as interesting in comparison to elementary aged subjects.

Welch, Howard, and Rush (1989) explored the use of visual feedback using a microcomputer. Subjects reportedly improved over the treatment period, and it was concluded that "verbal feedback on its own appears to be less powerful in promoting learning than real-time, meaningful visual feedback" (156). The treatment exercises employed the use of matching single pitches only and the investigators added that there was no evidence to suggest that matching single pitches results in singing accuracy of song phrases. The treatment by Welch, Howard, and Rush differed from the Yuba Method in that the exercises employed the use of matching single pitches and not melodic phrases as in the Yuba Method. The study by Welch, Howard and Rush also did not recognize the different vocal registers. This was evident by their comment that subjects in their study could learn to match pitch to a visual image using improper vocal technique and straining at the laryngeal level—which would not occur as in the Yuba Method if the different vocal registers had been utilized to avoid straining.

Jones (1971) investigated the use of a vertically-arranged keyboard instrument. The results indicated that: 1) subjects were more accurate in matching single pitches when those pitches were more than a step apart; 2) accuracy increased when stimulus pitches were at least a minor third apart; and 3) wider skips were found to be useful in helping subjects to gain the use of the upper range or the head voice (85). The Jones treatment differed from the Yuba Method in that it utilized a visual vertical keyboard along with pitch discrimination exercises and reinforcement techniques. The Jones treatment suggests psychological rather than physiological reasons for singing inaccuracy. Jones also began pitch matching within a subject's personal range. The practice of playing all pitches and patterns before singing was used. The investigator concluded that pitch matching and aural discrimination are skills which develop along with singing skills and should not be consisered prerequisites for singing. This conclusion supports the Yuba Method premise that pitch matching and aural discrmination skills may be secondary to the physiological cricothyroid functioning as a primary source of singing inaccuracy. The Yuba Method did not employ the playing of a vertical keyboard or utilize beginning exercises in pitch matching within a subject's personal range. Yuba Method exercises were begun on the same pitches for all subjects.

Buckton (1977) investigated the effects of vocal and instrumental instruction. Results indicated that the vocal program did significantly improve the vocal accuracy scores of the vocal group while the instrumental and control groups' scores slightly deteriorated on the posttest (43).

Roberts and Davies (1976) investigated vocal range extension. The results indicated that the remedial group showed greater improvements on single note production and

interval production (40). Unlike the Yuba Method, the treatment employed by Roberts and Davies utilized pitch discrimination and tonal memory exercises, whereas the Yuba Method focused on the physiological training of the cricothyroid muscle and coordination of the vocal registers. Kodaly hand signs were also used during the Roberts and Davies exercises. The Yuba Method did not use Kodaly hand signs to correspomd with the diatonic scale degrees used in the exercises. The glockenspiel was selected by Roberts and Davies for instrumental instruction and pitch discrimination exercises. The Yuba Method does not advocate the use of the playing of any instrument other than the voice.

While some studies indicated significant improvement of inaccurate singers, there is still a need for further research to corroborate findings. Gordon (1985) concluded that most areas of research have only one or two studies to support their findings. With the passage of almost two decades since his conclusions, there is still a great need for further research. The remedial methods with significant results should be used by music educators. Music educators should be trained in those methods to help inaccurate singers. However, with the limited amounts of time music educators have to work with inaccurate singers, the methods which are most effective and which take the least amounts of time should be utilized first. Further research needs to be conducted if methods are to be combined, for example breath control management with the Yuba Method because the effects of combining methods is not presently known or empirically tested. Combining methods would essentially be creating an entirely new method.

None of the aforementioned studies has utilized a treatment specifically involving the training of the cricothyroid muscle. As discussed in the section *The Act of Singing*, the

cricothyroid muscle serves as a pitch adjuster. This study empirically tested that Yuba Method and found it to significantly improve the inaccurate singer, especially those in the highly inaccurate subgroup.

Some observations were made by the investigator during the course of the treatment sessions that were noteworthy but not significant. While individualized training sessions were utilized, the process of training thirty subjects was time consuming. It is not feasible that a music educator would have that amount of time to provide individual remediation sessions to individual students. Shorter, group sessions, would be more ideal.

The investigator agrees with Greenberg (1970) who concluded that faulty singing can lead to disinterest in music as a whole subject. It was noted that high treatment subject 61310 was always a monotone singer from Kindergarten. The subject would sing individually in class year after year, still singing in a monotone voice. The subject's attitude towards music worsened with each passing year and no improvement in his singing ability. After the treatment, the subject went from a VPA score of 755.47 to 25.32. The subject's attitude towards music and self-esteem improved dramatically.

The treatment sessions as a whole were viewed positively by almost all students who indicated a desire to improve their singing. Many students continually inquired about their *Posttest Singing Stimulus* scores, which indicated an interest in their improvement.

No changing voice boys were included in the treatment or control groups, but there were at least two in the sample population. One of the changing voice boys was an accurate singer with a *Pretest Singing Stimulus* VPA of 42.60 and the other changing voice boy had a *Pretest Singing Stimulus* VPA of 333.33. The investigator perceived that

the accurate changing voice boy was always an accurate singer from Kindergarten and that the inaccurate changing voice boy was always an inaccurate singer from Kindergarten. A separate remedial method might be used with those students whose voices are changing.

Due to the fact that the treatment sessions were scripted, no repetitions were possible in the singing exercises to allow for individualized treatment. The investigator felt that some subjects may have needed to repeat some of the exercises for better improvement but it was necessary to give each subject equal treatment.

It was effective to have the investigator also be the regular music teacher due to the fact that the students were familiar with the instructor and more at ease during the testing and treatment sessions. There was no need for preliminary introduction and acquaintance sessions prior to the study.

It might also be beneficial to classify inaccurate singers into various types as Yuba (1998) had done previously. This would help to identify more specific remedial needs prior to treatment. For example, some students sang inaccurately throughout the singing stimuli, while others sang accurately up until the register break where they just "didn't make it." Others sang continually sharp or flat. Some continued to scoop pitches.

Singing stimuli in various keys might also help determine the type of inaccurate singer. For example, students who are inaccurate only out of their comfortable singing range, versus students who are inaccurate in all keys. Further research might be conducted utilizing Yuba's eight-type classification system prior to the treatment session to identify specific deficiencies of the inaccurate singer.

Several singing stimuli tests would be beneficial rather than just two. Some students were nervous for the tests. Having more practice taking the tests might help to ease the discomforts, since students would then know that they have more than one chance.

It was rewarding for the investigator to see the inaccurate students in the treatment group improve in their singing abilities. Several of those inaccurate singers were able to join the school's select choir after treatment. The students who improved found new joy in singing and expanded their musical potentials. The Yuba Method has provided new hope for those students thought to be unchangeably inaccurate singers.

Yuba mentioned that inaccurate singers who undergo treatment need to continually practice the exercises in order to maintain their newly developed singing skills. The investigator observed that at least for the remainder of the school year, and throughout the next school year, the students who received the Yuba Method treatment and had improved to become accurate singers, were still able to sing accurately. Further research needs to be conducted to confirm this observation and to see how long the effects of the Yuba Method treatment lasts.

CONCLUSIONS

Based on the results of this study, it is concluded that the Yuba Method as adapted for elementary-aged singers, was effective in correcting the vocal pitch accuracy of inaccurate elementary singers ($p < .0001$). The method was especially effective in correcting highly inaccurate singers ($p < .0024$).

RECOMMENDATIONS

Following are the recommendations by the investigator for future research based on the findings of this study:

1. The participant population in further studies be expanded to include students younger than the fourth grade and older than the sixth grade so that developmental influences can be investigated with different age groups.
2. Research studies should be conducted to determine more precise differences between echo singing phrases versus free song singing. It may be that these are two different skills altogether and would therefore need to have different types of treatment based on assessment.
3. Work with a laryngologist to assess cricothyroid function of subjects and therefore determine a specific cause for inaccurate singing for individual subjects.
4. Test for possible causes of inaccurate singing other than cricothyroid function prior to the study.
5. Repeat the study using a different and larger population to improve generalizability.
6. Repeat the study with a longer treatment period to see if additional treatment results in better vocal pitch accuracy improvement.
7. Retest treatment subjects at various intervals after treatment to see if the treatment effects last. Yuba noted that the treatment exercises need to be

repeated to maintain its effects, however, no specific time periods have been established.

8. Determine if it is reasonable to think that inaccurate singers always sing at the same pitch deviance. Test and retest subjects on more than one occasion and on different pitch levels.

9. Conduct further research to determine if inaccurate singers perform better with or without accompaniment. This research can then be taken a step further and discover a treatment for the singing inaccuracy due to singing with accompaniment and with the absence of accompaniment because obviously, there are going to be situations where students will have to sing with and without accompaniment.

10. Test the correlation between tonal memory and free song singing.

11. Test the correlation betweeng singing accuracy, attention span and behavior problems.

12. Adapt the Yuba Method for changing voice boys. There were only two changing voice boys in this study (in the sampled population only and not in the control or treatment groups), but this could be a potentially diverse group which would need specialized treatment. It was observed in the pilot study, for example, that one of the changing voice boy subjects was unable to produce any tones in the area between C4 and G4. His voice would skip over this area when singing. Yuba mentioned that this is a problem for some changing voice boys.

13. Determine what the actual vocal pitch accuracy threshold is for the inaccurate singer by the audible ear of music educators. Goetze (1985) arbitrarily picked the VPA of 100 cents or greater as the inaccurate singer but its validity has not been tested.
14. Repeat the Yuba Method to see if it is effective in larger groups. A large group treatment may be more feasible given the limited amounts of time available to music educators to remediate.
15. Repeat the study using different pretest and posttest singing stimuli to see if results are the same.
16. Repeat the study using singing stimuli at various keys to see if vocal registration is a factor in singing inaccuracy.
17. Correlate the *Pretest Singing Stimulus* with the *Posttest Singing Stimulus* to see if the levels of singing abilities required for each are comparable.
18. Repeat the study using various treatment time sessions—e.g. broken up into more, and shorter sessions. If effective, this would allow music educators to work with inaccurate singers during short pockets of time rather than having to utilize a whole forty-five minute class period.
19. Continue to corroborate past research findings on the correlation between pitch discrimination, tonal memory, and inaccurate singing.
20. Develop a standardized test for inaccurate singing.
21. Develop standardized assessment tests for inaccurate singing to better determine causality.

22. Perform a longitudinal study to see how much VPA fluctuates from grade to grade.
23. Conduct a study to determine the age when a subject becomes an inaccurate singer.
24. Conduct a study to see how the Yuba Method compares with other vocal treatment methods.

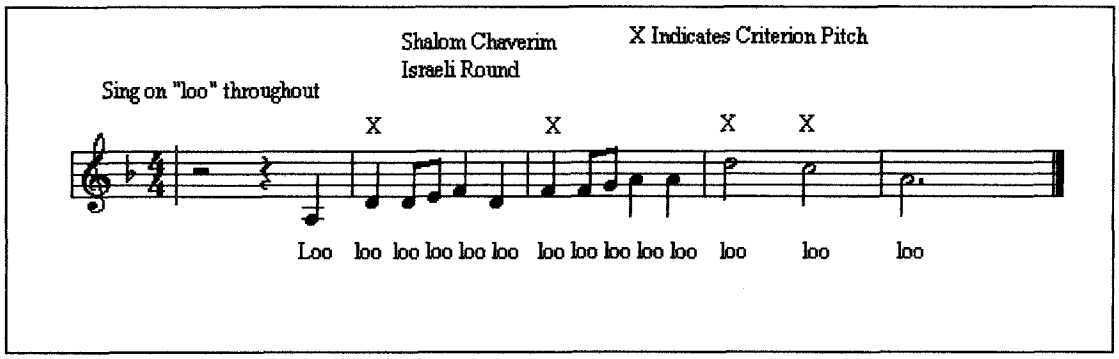
There have been some significant results reported by prior research studies to aid the inaccurate singer. Gordon's (1985) survey indicated that music educators at that time were not using proven methods to help inaccurate singers. The investigator doubts that this situation has changed in the nearly two decades that have since passed. It should be the responsibility of the music education profession to educate and inform music educators on significant research findings. It is the responsibility of the music educator to aid the inaccurate singer. The purpose of research should be to enhance the teaching of music educators and not to remain as a separate field amongst researchers alone. A connection needs to be made between those two seemingly dichotomous entities if music education is to progress beyond the information age of the past.

Results from this study indicated that the Yuba Method was most effective on the most inaccurate singers. Although additional studies need to corroborate results of this study, there is hope that all students can master singing simple songs in tune. It is recommended that music educators become trained in the Yuba Method so that it can be used as a remedial method with inaccurate singers. The possibility that the music education profession might someday empirically establish a cure for inaccurate singing provides hope that no child would be left behind in vocal music. Vocal music is often the

door to further music experiences and so its improvement will perhaps lead to music literacy for each and every person in our society today, where every person can experience the joys and great fulfillment that only music can provide. Music is the unspoken aesthetic that makes all of us human, humane, and cultured. It is a great gift that only mankind can experience.

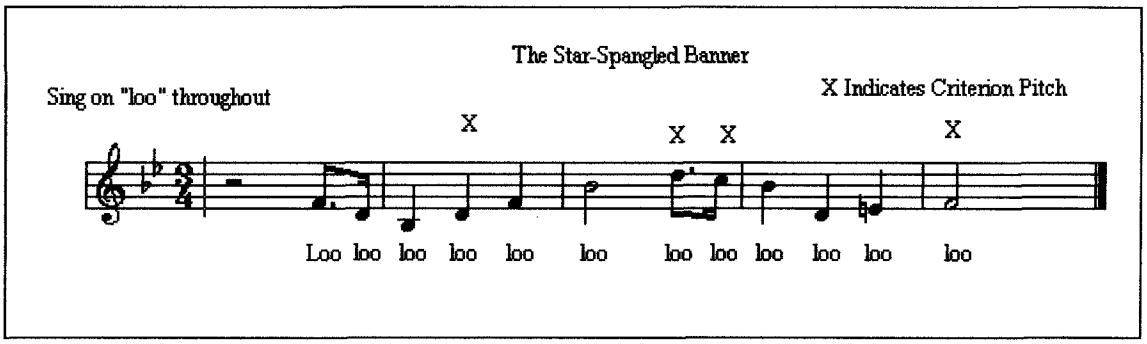
APPENDIX I

THE PRETEST SINGING STIMULUS



APPENDIX II

THE POSTTEST SINGING STIMULUS



APPENDIX III

SINGING STIMULI SCRIPT

Investigator: "Good morning boys and girls. Remember the song that we learned in class? Today we will review singing that song. Let's sing it together." (Students then proceed to sing it with the piano accompaniment.)

Investigator: "Now let's sing it a capella. I will give you the first two pitches of the song and then you come in. Ready, go." (Students then sing the song "a capella.")

Investigator: "That was very good. Now we are going to sing only a part of the song on the word "loo" to make it easier for you." (Teacher sings the song phrase on "loo.")

Investigator: "Now you try it." (Teacher plays the first two pitches of the song and students sing the stimulus phrase on the word "loo.")

Investigator: "Very good. Today you will be singing this one by one as a way to assess how well you can sing. This will not affect your grade in music. Are there any questions?" (Teacher then takes students one by one for individual testing. All students are to be praised regardless of their singing success.)

APPENDIX IV

SAMPLE LETTER TO THE SCHOOL ADMINISTRATION

Karen A. Miyamoto
2506 Akepa St.
Pearl City, HI 96782
(808) 455-7592
miyamotok001@hawaii.rr.com

March 7, 2002

Ms. Patricia Dang, Principal
Kapalama Elementary School
1601 North School Street
Honolulu, HI 96817

Dear Ms. Dang,

My name is Karen Miyamoto. I am a doctoral candidate in Music Education at the University of Hawaii at Manoa Music Department. I am currently in the last phase of my doctoral dissertation at the University of Hawaii at Manoa Music Department. I would like to kindly request permission to conduct research at Kapalama Elementary School. The intended subjects are students in grades 4, 5, and 6.

In this study, I am investigating a voice training method to see if it helps students who are considered to be inaccurate singers to sing better in tune. A random sample of 60 students from grades 4, 5, and 6 will take part in the study.

If a child participates, he or she will take a short singing test (to see whether he or she is an accurate or an inaccurate singer) and will be recorded on a computer to be scored. Thirty students will go through the vocal training method and thirty students will be in a control group with no extra training. All students will take two post-singing tests which will also be recorded on a computer and scored. These recordings will be used only for this project and will be erased at the completion of the study.

The total amount of time each child will participate does not exceed 60 minutes, and results will be made available upon request. Students will need to be pulled out of class and worked with on an individual basis. Each of the thirty students in the experimental group will receive one 60-minute training session of the vocal training method.

The projected time frame for the study is April 15, 2002 through April 26, 2002 for the experimental instruction and the first Post-Singing Stimulus Test. The Second Post-Singing Stimulus Test will be administered from May 27-31, 2002. The Second Post-Singing Stimulus Test will be administered during their regular music period.

Attached is a copy of the parental permission slip form that will be sent home with the students. Please note that participation is voluntary, and that students may withdraw at any time. Although I will take down names of participants for my own use as I collect the data, no names will be used in reporting the data. All participants will remain anonymous. If you have any questions about the study, please contact me at 832-3290 x272.

Thank you very much for your assistance.

Sincerely,

Karen A. Miyamoto
Music Education Doctoral Candidate
University of Hawaii at Manoa
Music Department

APPENDIX V

SAMPLE CONSENT FORM TO PARENTS/GUARDIANS

PARENTAL/GUARDIAN CONSENT FORM
AGREEMENT TO PARTICIPATE IN
"THE EFFECT OF THE TORU YUBA VOCAL TRAINING METHOD
ON THE VOCAL PITCH-ACCURACY OF INACCURATE ELEMENTARY
SINGERS"

Karen A. Miyamoto, Principal Investigator
Kapalama Elementary School
1601 North School St.
Honolulu, HI 96817
(808) 832-3290 x272

Date:_____

Dear Parents and Legal Guardians,

Hi! My name is Karen Miyamoto. I am the Music Teacher for students in Grades 4, 5, and 6 at Kapalama Elementary School. I am also a Doctoral Candidate in Music Education at the University of Hawaii at Manoa Music Department.

During the next few weeks, students from Kapalama Elementary School will be participating in a research study of children's singing. In this study, I am investigating a voice training method to see if it helps students to sing better in tune. Although only students who are inaccurate singers will participate in the actual experimental or control group, this letter is being sent home to all students. All students will take the singing pretest.

If your child participates, he or she will take a short singing pretest to see whether or not he or she is an accurate or an inaccurate singer (based on past research criteria). The pretest will be recorded on a computer to be scored. A random sample of 60 inaccurate singers from Grades 4 and 5 (and 6 if there are not enough students in Grades 4 and 5 who respond) will be selected. Thirty students will go through the experimental vocal training method and thirty students will be in a control group with no extra training. All students will take a post-singing test which will also be recorded on a computer and scored. These recordings will be used only for this project and will be erased at the complete of the study. The total amount of time each of the thirty children in the experimental group will spend in the vocal training method will not exceed 60 minutes, and results will be made available upon request. Students who go through the vocal training lesson will take a post-test immediately following the session. All other students will take a Post-Singing Test at the completion of the training period. The Post-Singing Test will also be recorded and scored on a computer. The 60-minute experimental vocal training method will be video-taped for verification of procedures only. All recordings will be erased at the conclusion of the research project. Students will be given a number for the research project. Strict confidentiality of records identifying the student will be maintained. The time period to complete the Pretest, training period, and Posttests is approximately six weeks.

There are no foreseeable risks or discomforts to the student. All students who participate in the research project will be helping to identify ways in which students who are inaccurate singers can be helped. Students who participate in the experimental vocal training method may benefit with a lifetime of accurate singing. All students will be congratulated and praised after the singing tests and training method (if one of the thirty randomly selected) regardless of singing improvement or performance. There will be no cost to the student.

Please note that participation is voluntary, and that you or your child may withdraw at any time with no penalty or loss to the student of any kind. Although I will take down names of participants for my own use as I collect the data, no names will be used in reporting the data. All participants will remain anonymous. Results will be available upon request.

If you have any questions about the study, please contact me at 832-3290 x272. Please contact me if you have any questions on alternative methods to help with your child's vocal pitch accuracy. If your child can participate, kindly complete the permission slip below and return it to me at your child's music class by _____.

I certify that I have read and that I understand the foregoing, that I have been given satisfactory answers to my inquiries concerning project procedures and other matters and that I have been advised that I am free to withdraw my consent and to discontinue participation in the project or activity at any time without prejudice.

I consent to the participation of my minor child or minor ward in this project with the understanding that such consent does not waive any of my legal rights, nor does it release the principal Investigator or the institution or any employee or agent thereof from liability for negligence.

_____

Name of Minor (First, Last)

_____

Signature of Authorized Legal Representative

_____

Date

(If you cannot obtain satisfactory answers to your questions or have comments or complaints about your treatment in this study, contact: Committee on Human Studies, University of Hawaii, 2540 Maile Way, Honolulu, Hawaii 96822. Phone: (808) 539-3955.)

c: Signed copy to subject

APPENDIX VI

SAMPLE ASSENT FORM TO STUDENTS

STUDENT ASSENT FORM
AGREEMENT TO PARTICIPATE IN
"THE EFFECT OF THE TORU YUBA VOCAL TRAINING METHOD
ON THE VOCAL PITCH-ACCURACY OF INACCURATE ELEMENTARY
SINGERS"

Karen A. Miyamoto, Principal Investigator
Kapalama Elementary School
1601 North School St.
Honolulu, HI 96817
(808) 832-3290 x272

Date:_____

Dear Students,

Hi! I am Mrs. Miyamoto, your music teacher. I am also a Doctoral Candidate in Music Education at the University of Hawaii at Manoa Music Department.

During the next few weeks, I will be conducting a study of children's singing. In this study, I am testing a voice training method to see if it helps students to sing better in tune. I am asking for your assistance in helping me to do this study. If successful, this study will help other music teachers teach their students how to sing.

If you participate, you will take a short singing pretest. The pretest will be recorded on a computer to be scored. Sixty students from Grades 4 and 5 (and 6 if there are not enough students in Grades 4 and 5) will be randomly selected. Thirty of these students will go through the experimental vocal training method and thirty students will be in a control group with no extra training. If you are selected for the vocal training method group, you will be pulled out of class for one 60-minute lesson.

All students will take a post-singing test which will also be recorded on a computer and scored to see if there is any improvement. These recordings will be used only for this project and will be erased at the complete of the study. The 60-minute experimental vocal training lesson will be video-taped to make sure that I am teaching the lesson correctly. All recordings will be erased at the conclusion of the research project. Your records and participation will be kept anonymous.

All participation is voluntary and you may withdraw at any time without penalty or loss of any kind. Participation in this study will not affect your music grade. Results will be available upon request.

If you have any questions about the study, please contact me at 832-3290 x272. If you would like to participate, please complete the permission slip below and return it to me at the school office by _____.

I certify that I have read and that I understand the foregoing, that I have been given satisfactory answers to my inquiries concerning project procedures and other matters and that I have been advised that I am free to withdraw my consent and to discontinue participation in the project or activity at any time without prejudice.

I consent to participation in this project with the understanding that such consent does not waive any of my legal rights, nor does it release the principal Investigator or the institution or any employee or agent thereof from liability for negligence.

_____

Name of Minor (First, Last) [Please Print]

_____

Signature of Minor Student

_____

Date

(If you cannot obtain satisfactory answers to your questions or have comments or complaints about your treatment in this study, contact: Committee on Human Studies, University of Hawaii, 2540 Maile Way, Honolulu, Hawaii 96822. Phone: (808) 539-3955.)

c: Signed copy to subject

APPENDIX VII

USE OF HUMAN SUBJECTS FORM

UNIVERSITY OF HAWAI'I

Committee on Human Studies

MEMORANDUM

April 12, 2002

TO: Karen A. Miyamoto
Principal Investigator
Department of Music

FROM: William H. Dendle
Executive Secretary

SUBJECT: CHS #11734- "The Effect Of The Toru Yuba Vocal Training Method On The Vocal Pitch-Accuracy Of Inaccurate Elementary Singers"

Your project identified above was reviewed and has been determined to be exempt from Department of Health and Human Services (DHHS) regulations, 45 CFR Part 46. Specifically, the authority for this exemption is section 46.101(b)(1). Your certificate of exemption (Optional Form 310) is enclosed. This certificate is your record of CHS review of this study and will be effective as of the date shown on the certificate.

An exempt status signifies that you will not be required to submit renewal applications for full Committee review as long as that portion of your project involving human subjects remains unchanged. If, during the course of your project, you intend to make changes which may significantly affect the human subjects involved, you should contact this office for guidance prior to implementing these changes.

Any unanticipated problems related to your use of human subjects in this project must be promptly reported to the CHS through this office. This is required so that the CHS can institute or update protective measures for human subjects as may be necessary. In addition, under the University's Assurance with the U.S. Department of Health and Human Services, the University must report certain situations to the federal government. Examples of these reportable situations include deaths, injuries, adverse reactions or unforeseen risks to human subjects. These reports must be made regardless of the source funding or exempt status of your project.

University policy requires you to maintain as an essential part of your project records, any documents pertaining to the use of humans as subjects in your research. This includes any information or materials conveyed to, and received from, the subjects, as well as any executed consent forms, data and analysis results. These records must be maintained for at least three years after project completion or termination. If this is a funded project, you should be aware that these records are subject to inspection and review by authorized representatives of the University, State and Federal governments.

Please notify this office when your project is completed. We may ask that you provide information regarding your experiences with human subjects and with the CHS review process.

2540 Maile Way, Spalding 252, Honolulu, Hawai'i 96822-2303
Telephone: (808) 539-3955/(808) 956-5007 Facsimile: (808) 539-3954, Web site: www.hawaii.edu/irb

Upon notification, we will close our files pertaining to your project. Any subsequent reactivation of the project will require a new CHS application.

Please do not hesitate to contact me if you have any questions or require assistance. I will be happy to assist you in any way I can.

Thank you for your cooperation and efforts throughout this review process. I wish you success in this endeavor.

Enclosure

OMB No. 0925-0418
Approved for use through 01/31/2001

Protection of Human Subjects
Assurance Identification/Certification/Declaration
(Common Federal Rule)

Policy: Research activities involving human subjects may not be conducted or supported by the Departments and Agencies adopting the Common Rule (56FR28003, June 18, 1991) unless the activities are exempt from or approved in accordance with the common rule. See section 101(b) the common rule for exemptions. Institutions submitting applications or proposals for support must submit certification or appropriate Institutional Review Board (IRB) review and approval to the Department or Agency in accordance with the common rule. Institutions with an assurance of compliance that covers the research to be conducted on file with the Department, Agency, or the Department of Health and Human Services (HHS) should submit certification of IRB review and approval with each application or proposal unless otherwise advised by the Department or Agency. Institutions which do not have such an assurance must submit an assurance and certification of IRB review and approval within 30 days of a written request from the Department or Agency.

1. Request Type	2. Type of Mechanism	3. Name of Federal Department or Agency and, if known, Application or Proposal Identification No.
☐ ORIGINAL ☐ FOLLOWUP ☒ EXEMPTION	☐ GRANT ☐ CONTRACT ☐ FELLOWSHIP ☐ COOPERATIVE AGREEMENT ☐ OTHER:	

4. Title of Application or Activity	5. Name of Principal Investigator, Program Director, Fellow, or Other
"The Effect Of The Toru Yuba Vocal Training Method On The Vocal Pitch-Accuracy Of Inaccurate Elementary Singers"	Karen A. Miyamoto

6. Assurance Status of this Project *(Respond to one of the following)*

☒ This Assurance, on file with Department of Health and Human Services, covers this activity:
Assurance identification no. M- 1217 IRB identification no. 01

☐ This Assurance, on file with *(agency/dept)*_____, covers this activity.
Assurance identification no._____ IRB identification no._____ *(if applicable)*

☐ No assurance has been filed for this project. This institution declares that it will provide an Assurance and Certification of IRB review and approval upon request.

☒ Exemption Status: Human subjects are involved, but this activity qualifies for exemption under Section 101(b), paragraph (1).

7. Certification of IRB Review (Respond to one of the following IF you have an Assurance on file)

☐ This activity has been reviewed and approved by the IRB in accordance with the common rule and any other governing regulations or subparts on (date) by: ☐ Full IRB Review or ☐ Expedited Review

☐ This activity contains multiple projects, some of which have not been reviewed. The IRB has granted approval on condition that all projects covered by the common rule will be reviewed and approved before they are initiated and that appropriate further certification will be submitted.

8. Comments

CHS #11734

9. The official signing below certifies that the information provided above is correct and that, as required, future reviews will be performed and certification will be provided.		10. Name and Address of Institution
11. Phone No. *(with area code)* (808) 956-5007	12. Fax No. *(with area code)* (808) 539-3954	University of Hawaii at Manoa Office of the Chancellor 2444 Dole Street, Bachman Hall Honolulu, HI 96822
13. Name of Official William H. Dendle		14. Title Compliance Officer
15. Signature		16. Date April 10, 2002

Authorized for local Reproduction OPTIONAL FORM 310 (Rev. 1-98) Sponsored by HHS/NIH

APPENDIX VIII

TREATMENT SCRIPT

INTRODUCTION

- Goodmorning/afternoon. You are here to help to test out this vocal training method to see if it helps students to sing better in tune. You will be going through some singing exercises. This will take 45 minutes.
- This lesson will be videotaped to be sure that all students receive the same lesson.
- You will be given a short singing test at the end of the lesson to see if your singing has improved, gotten worse, or stayed the same. Your grade for music will not be affected by this lesson or the singing test. This test will be recorded and scored by a computer.
- The exercises are recorded on a CD for you to echo back.
- You may ask questions at any time during the period if you do not understand something. Do you have any questions?

Instructions for the Instructor

- Throughout the teaching of the method, the instructor will make corrections as needed in accordance with the written script/instructions to the student.
- Instructor checks for the following: (1) Do not nod the head. Instead tap the foot or hand at the side to keep the beat. (2) Tongue position forward and just below the lower teeth with relaxation. If the tongue is in a backward position, the tone becomes vague and obscure. (3) For the falsetto voice there is open space between the back molars. Check this by placing the thumb and middle finger into the cheeks to feel the space.
- Instructor checks for the following: (1) Make sure the jaw position does not move. "A" and "Ah" for example, should have the same jaw position. (2) Have the student project farther—fixing his/her eyes to a distance will help to visualize this. (3) Both the natural and falsetto voices need to be strong. (4) Be sure to praise the student every step of the way.
- Instructor moves on when the student has mastered both the Natural and Falsetto Voices. At any time during the exercises the facilitator may use hand motions and verbal instructions to help correct pitches to make them either higher or lower. If the pitch is too high, the subject should decrease air pressure and use more of the natural voice. If the pitch is too low, the subject should increase air pressure and use more of the falsetto voice.

- Instructor should use hand signs to facilitate pitch contour movement visualization. For "up" the right hand should be held above the head as if holding a pizza. For "down" the right hand should face palm down by the hip.
- Instructor checks that the subject links the vowels "hu-ah" smoothly using a glissando from the falsetto to the natural voice in a portamento style.

Mouth Configuration

- Pay attention to your mouth shape when singing. Check your mouth position in the mirror with the pictures given below.
- Open your mouth with a light smile, vertically and in an upright position. [Figure 1]
- Do not bring wrinkles to the middle of your forehead. [Figure 2]
- Do not raise your shoulders. [Figure 3]
- Do not open your mouth excessively. [Figure 4]
- Do not raise your chin. [Figure 4]
- Do no place your tongue at the back of the mouth.

Figure 1. (Yuba, 1998; 42)

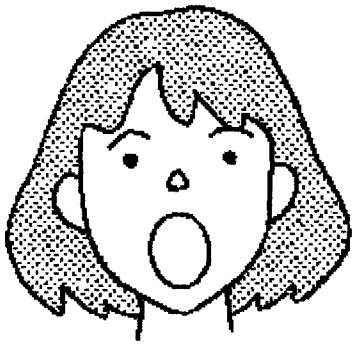


Figure 2. (Yuba, 1998; 42)

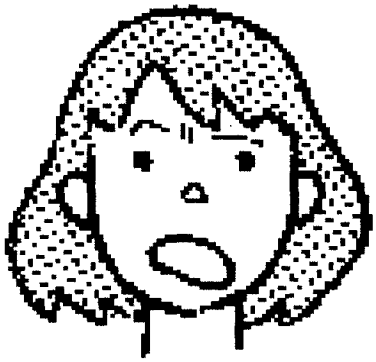


Figure 3. (Yuba, 1998; 42)

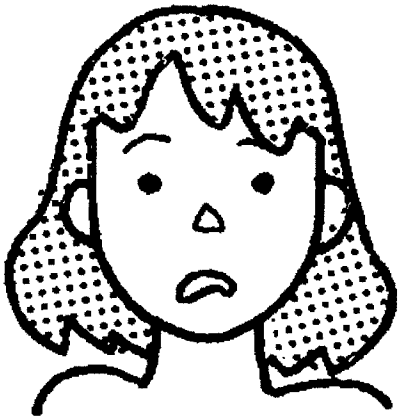
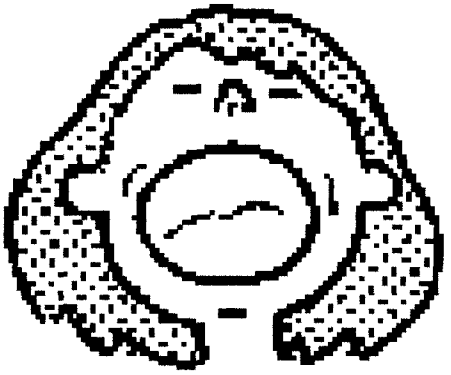


Figure 4. (Yuba, 1998; 42)



Proper Standing Position

- Now we will work on the proper standing position.
- Compare your position with the pictures below.
- Place your heel firmly on the ground. [Figure 5]
- Place your weight slightly forward. [Figure 6]
- Fix your eyes looking at eye level height and look to a distance. [Figure 6]
- Stand with your legs approximately a shoulder width apart. [Figure 6]

Figure 5. (Yuba, 1998; 43)

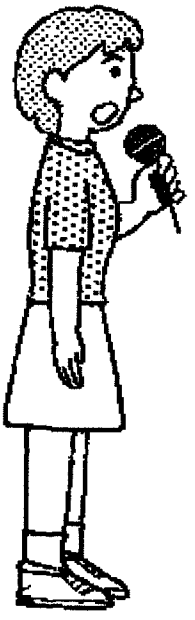


Figure 6. (Yuba, 1998; 43)

- Do not raise the shoulders. [Figure 7]
- Do not bend backwards. [Figure 7]
- Do not round the shoulders or hunch over. [Figure 8]

Figure 7. (Yuba, 1998; 43)

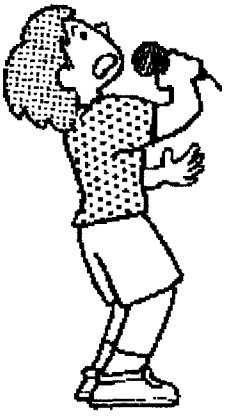
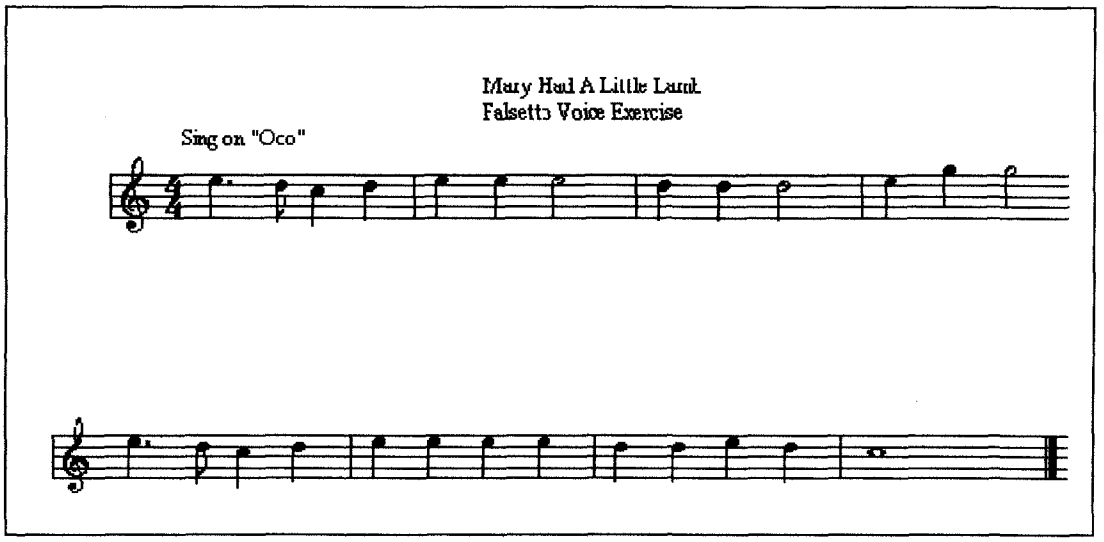


Figure 8. (Yuba, 1998; 43)

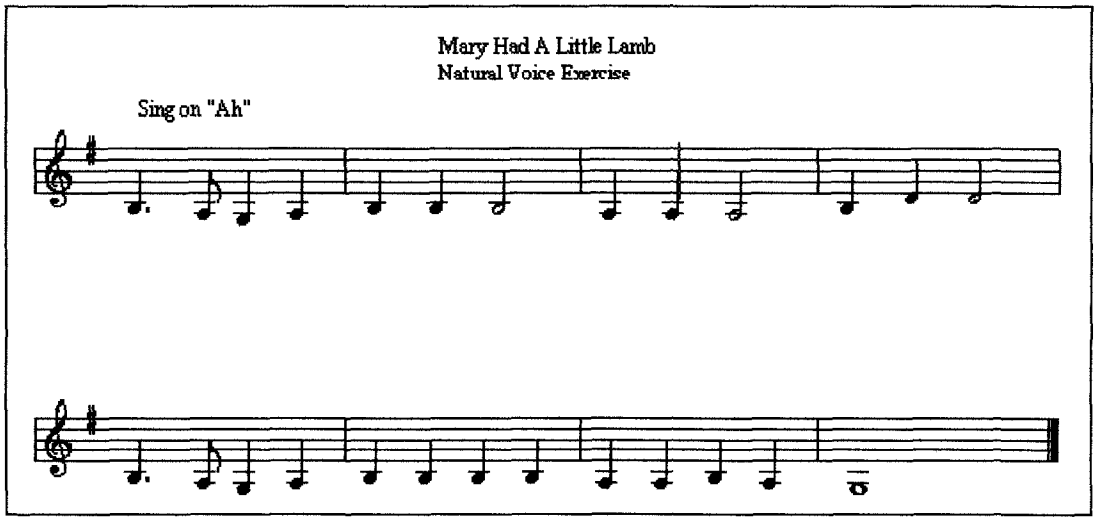
Falsetto And Natural Voice Exercises

- You have a falsetto and a natural voice. Each voice has its own distinct sound quality.
- Listen to the sample of the falsetto voice. [Track 1: "Yahoo! Hoo, hoo!]
- Now listen to the sample of the wrong way to produce the falsetto voice. [Track 2]
- Now with the falsetto voice, project your voice outward as you echo the recording. Open the back of the throat as if lightly yawning.
- "Yahoo!" "Hoo, hoo" [Track 3]
- "Woo" [Track 4]
- Now we will practice the natural voice. Project your voice with authority in a rather low but continuous breath.
- Pronounce the vowel "ah" sharply and clearly. It should feel as if tightening the throat muscles.
- Echo the recording.
- "Ah, ah, ah" [Track 5]
- "Ah--, ah--, ah—" [Track 6]
- We will now practice the falsetto and natural voice.
- Make sure each sound is clear.
- "Wo, ah, wo, ah" [Track 7]
- Let's practice the falsetto voice once more.
- Echo the recording. The easiest was to get the falsetto voice is by making the sound of a dog howling.
- Echo the recording.
- "Wo!" [Track 8]
- Let's practice the natural voice once more.
- Echo the recording of the strong cry of a sheep.
- "Mee, mee" [Track 9]
- Echo the recording of the creaking sound of a door.
- "Gii, gii" [Track 10]
- You can also make up falsetto and natural voice exercises of your own.
- Can you think of some? (Do some exercises that the student thinks of or the instructor will recommend some.)
- Continue these exercises repeatedly until you have gained a thorough understanding of the proper use of each.

- Now let's sing a simple song using the Falsetto Voice. (Mary Had A Little Lamb)

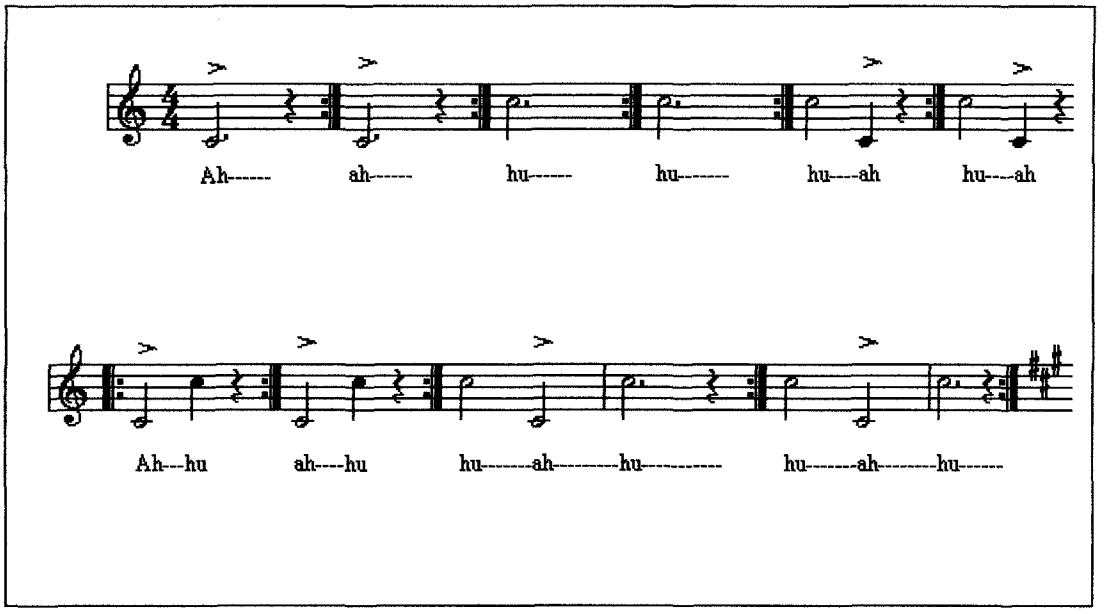


- Now let's sing the same song using the Natural Voice. (Mary Had A Little Lamb)



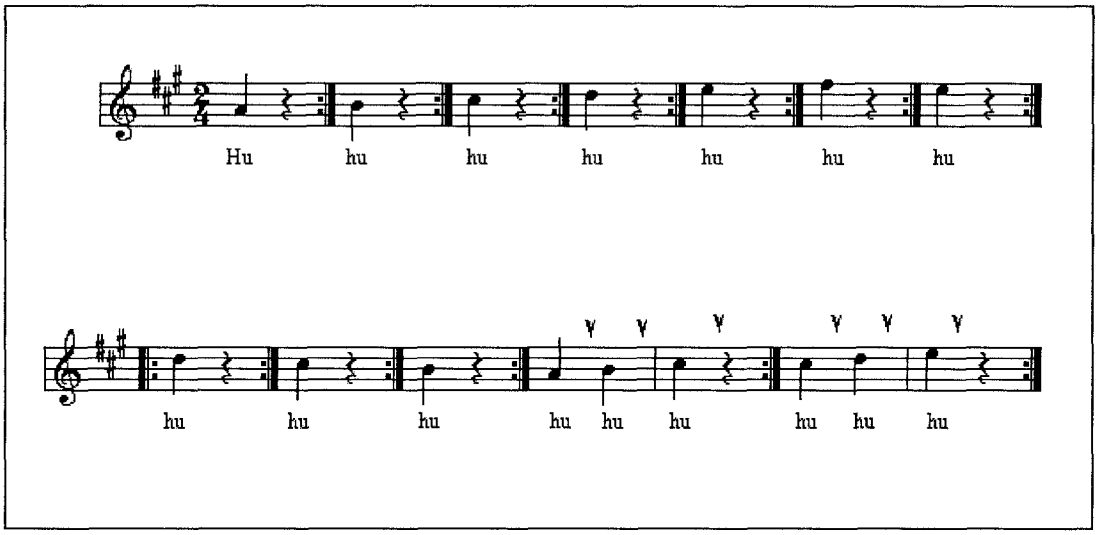
Falsetto/Natural Voice Separation And Establishment

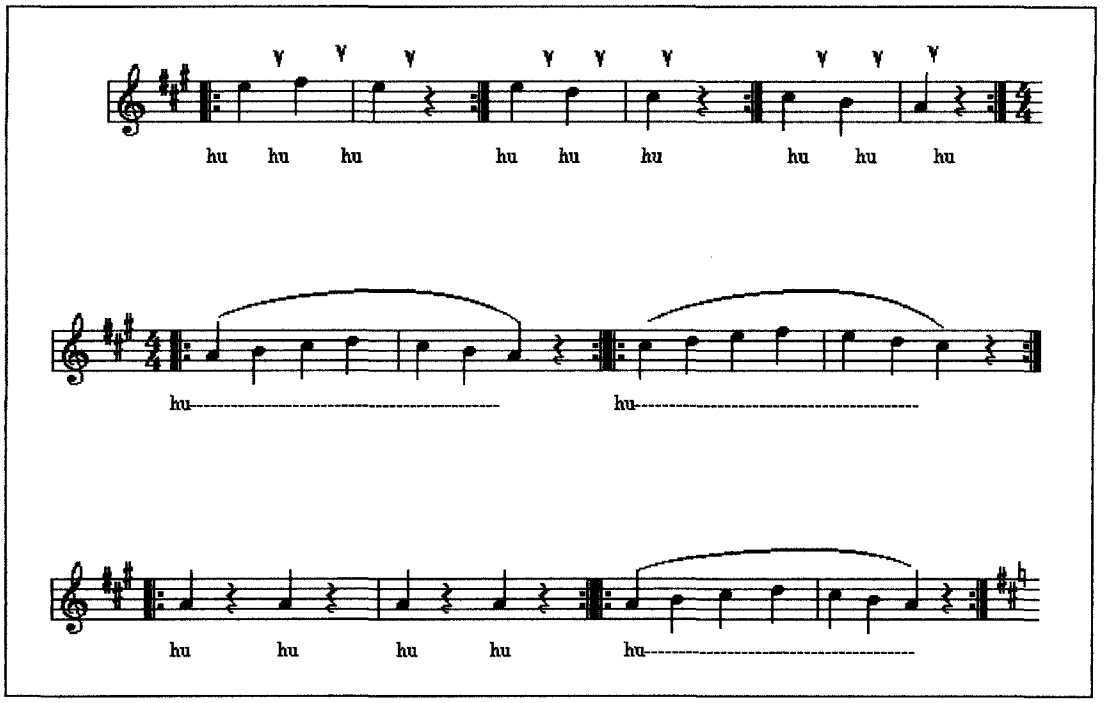
- These next training steps help you to produce the high sounds of a falsetto voice (hu).
- Do not use the natural voice tone quality.
- Use a continuous breath of air when projecting the natural voice low sound (ah). Be sure not to have the falsetto sound.
- High tones do not always mean it is a falsetto voice. This is a tonal quality difference.
- Echo the exercises on the CD (Track 11). Follow along with the music notation below.



Falsetto Strengthening And Stabilization

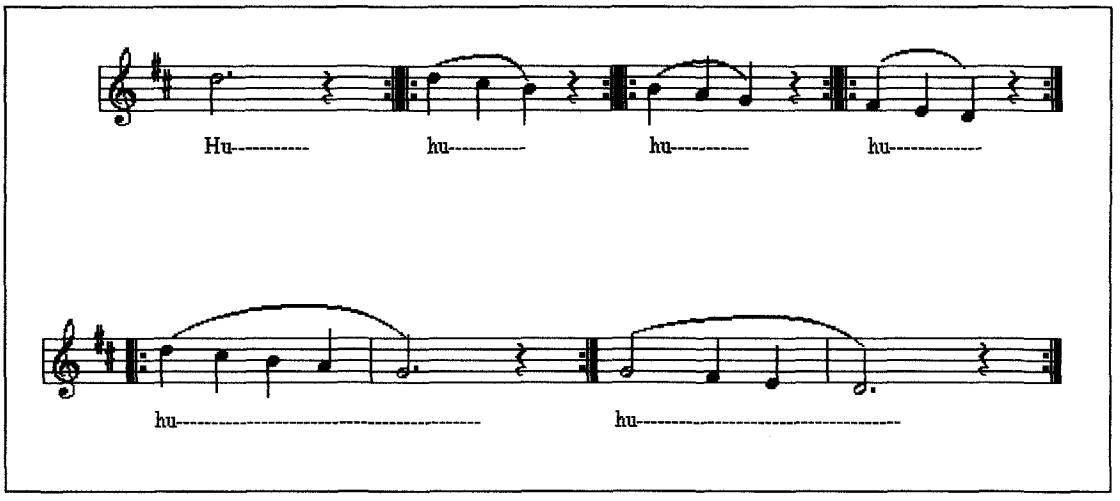
- This exercise is to build interval training of the cricothyroid muscle.
- The cricothyroid muscle is the muscle that controls the vocal cords to make the tones higher or lower.
- Inhale enough air and with a falsetto breath of air (as if through a small opening). Say "Hu."
- Be sure not to use the natural voice.
- Take the breath at the 'V' sign and continue to hold the high tone.
- Take a breath of great energy as you go up and weaken as you go down in pitch.
- The sound produced is not a singing voice.
- This exercise focuses on first getting the pitches, so do not worry about the tone quality right now.
- Halfway through, sing the falsetto vowel "uu" smoothly not allowing much breath to escape.
- Now follow along with the music notation below and echo the CD exercises (Track 12).

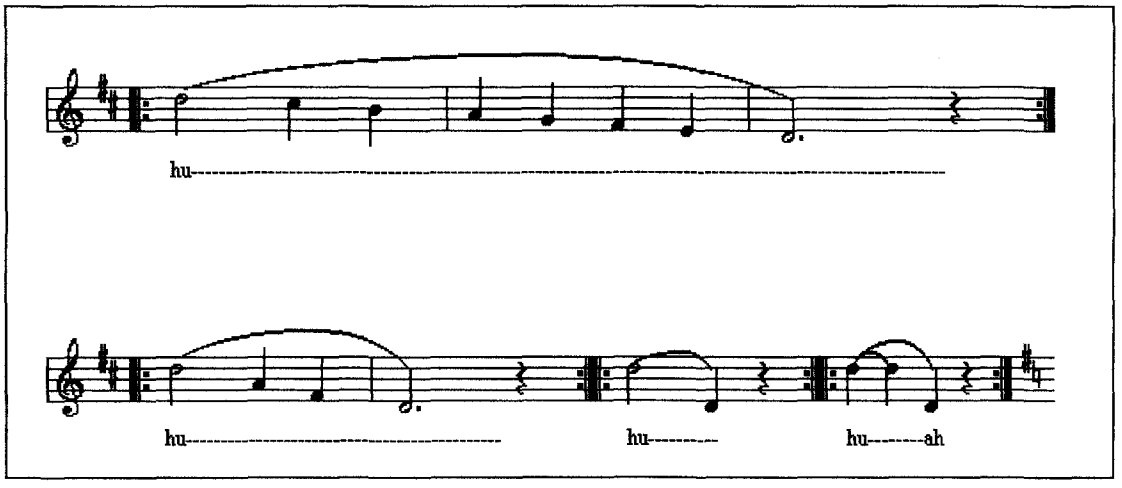



hu hu hu hu hu hu hu hu hu
hu
hu
hu hu hu hu hu

The Transference From The Falsetto To The Natural Voice

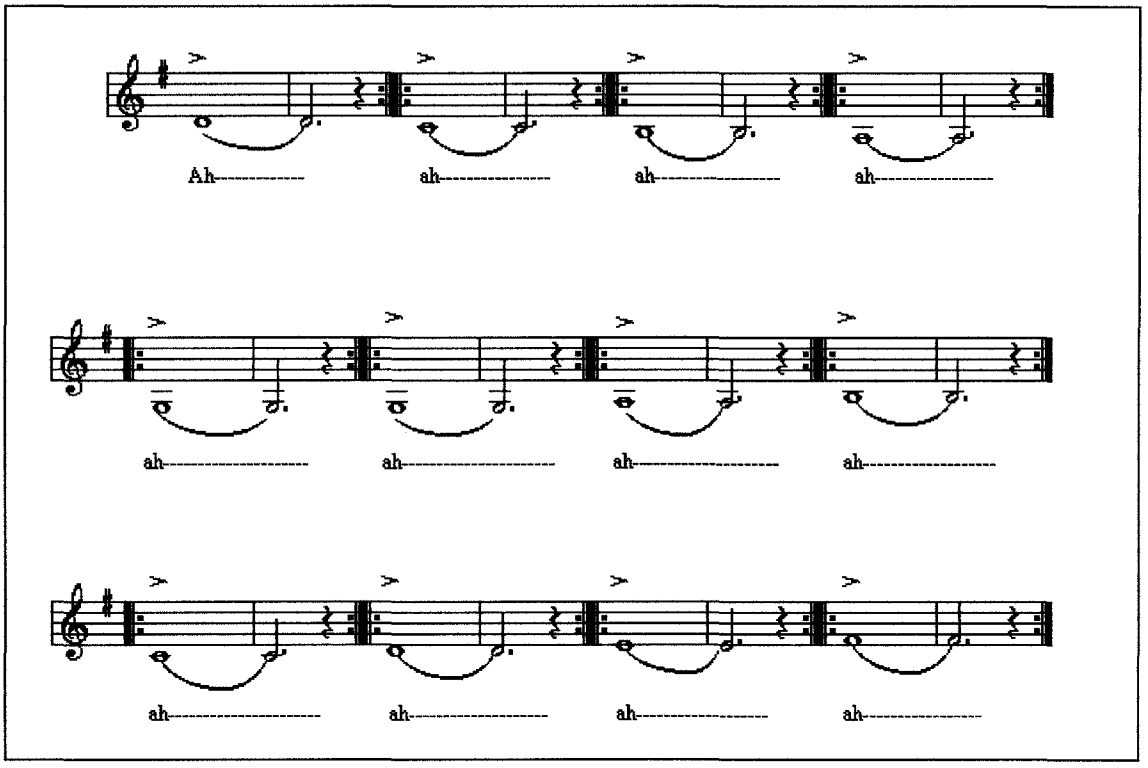
- Next we are going to learn how to change from the Falsetto to the Natural Voice.
- There are times when singing becomes difficult due to the point of voice change (the turning point of the falsetto and the natural voice).
- Do not worry about this problem at first. As long as you are able to sing the intervals, it doesn't matter if the point of voice change stands out.
- One secret of switching to the Natural Voice is to focus on the outer regions of the throat and progressively move inward as though constricting or getting smaller.
- It is good for there to be a smooth transition in singing at the point of voice change without big differences in tone.
- We will start at first with the high tones of the Falsetto Voice.
- Echo back the exercises that you hear on the CD (Track 13). Follow along with the music notation below.

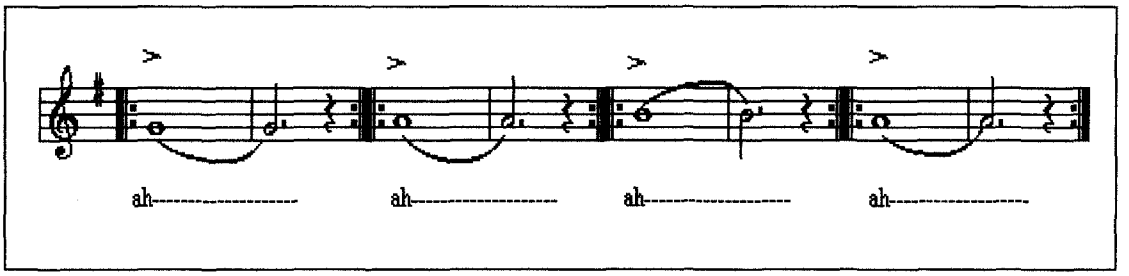


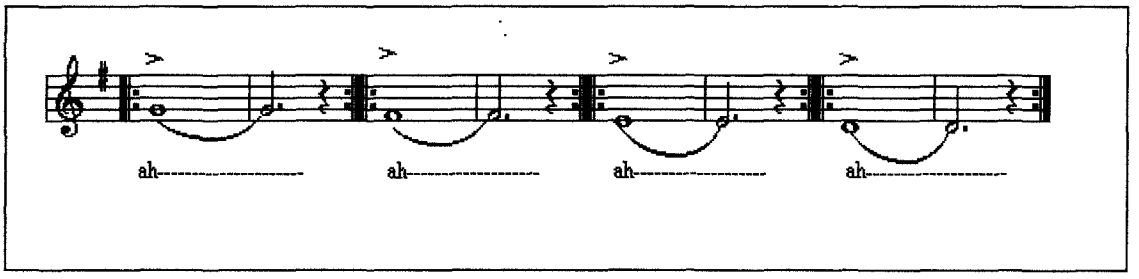


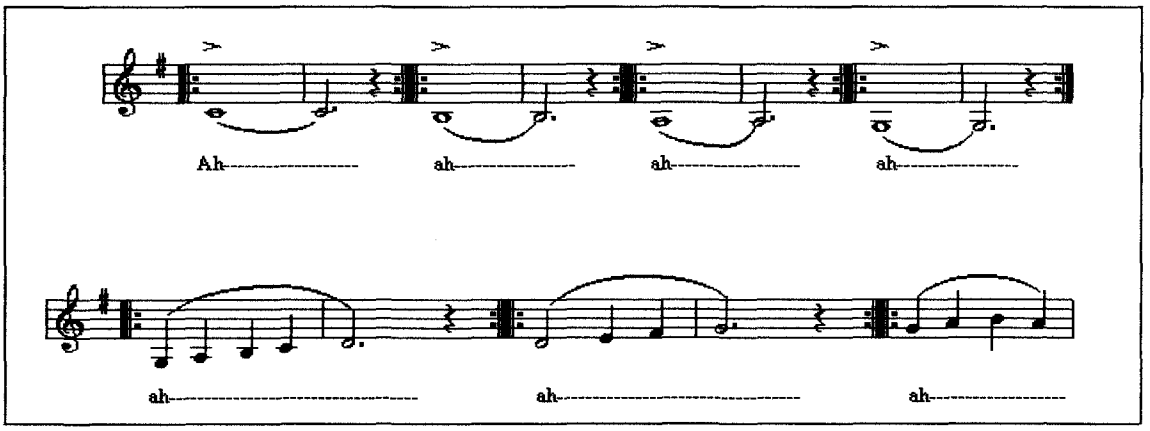
The Transference From Strengthening The Natural Voice To The Falsetto Voice

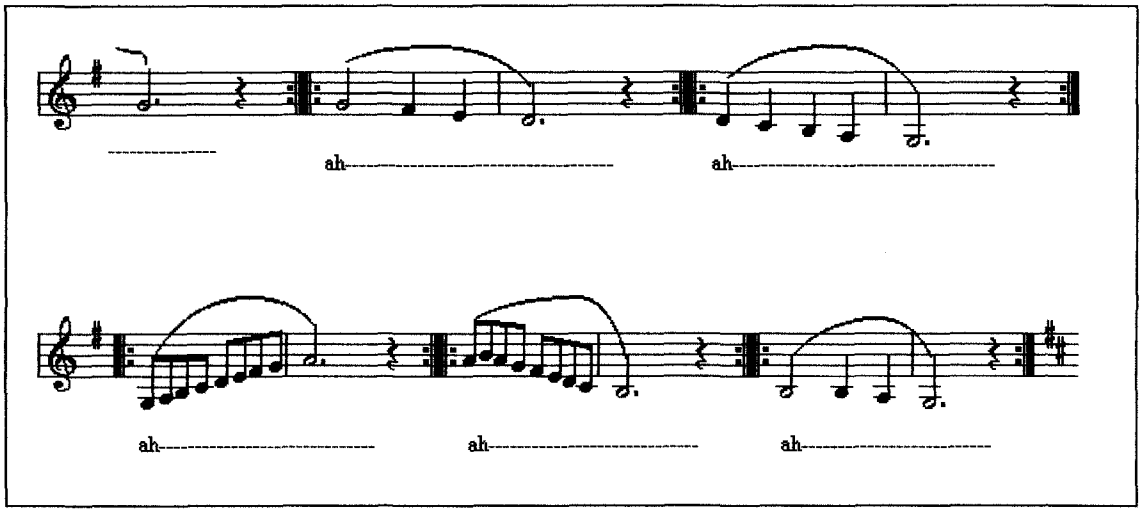
- Now we are going to learn how to change from the Natural Voice to the Falsetto Voice.
- Train the Natural Voice by projecting the vowel "Ah" clearly.
- Do not yell with all your might. Little by little get the Falsetto Voice sound without strain and project your voice.
- Let the sound ring out and sing with energy.
- Echo back the exercises that you hear on the CD (Track 14). Follow along with the music notation below.




ah
ah
ah
ah

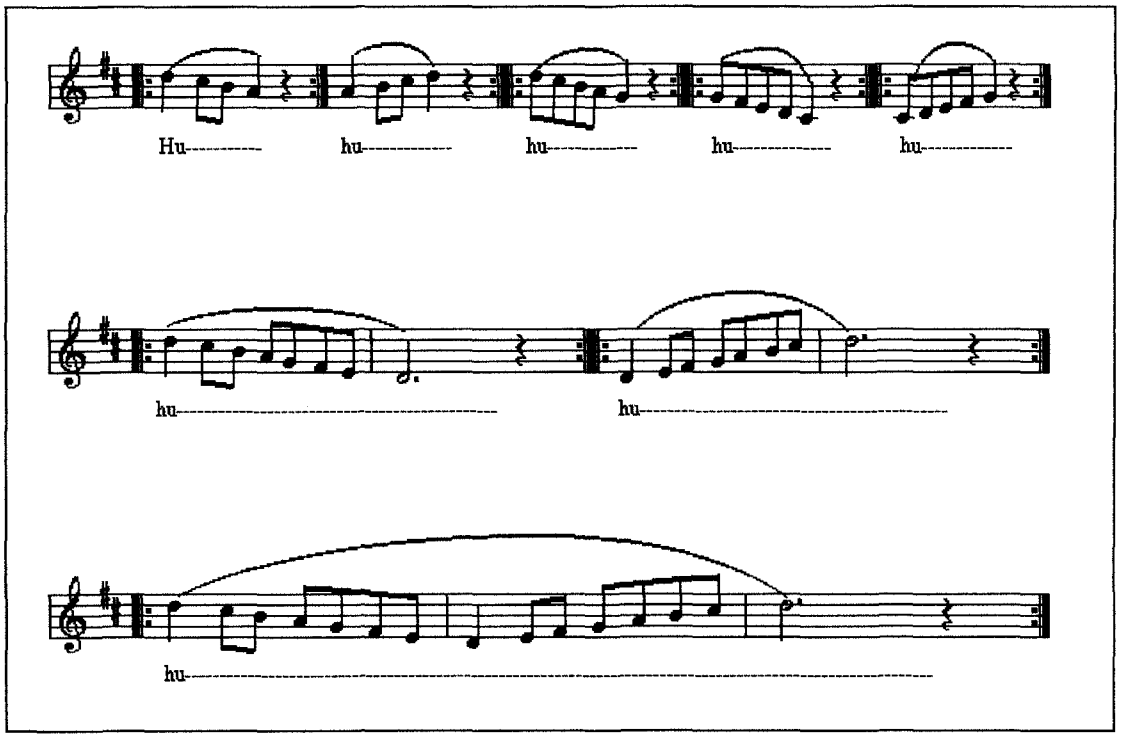

ah
ah
ah
ah

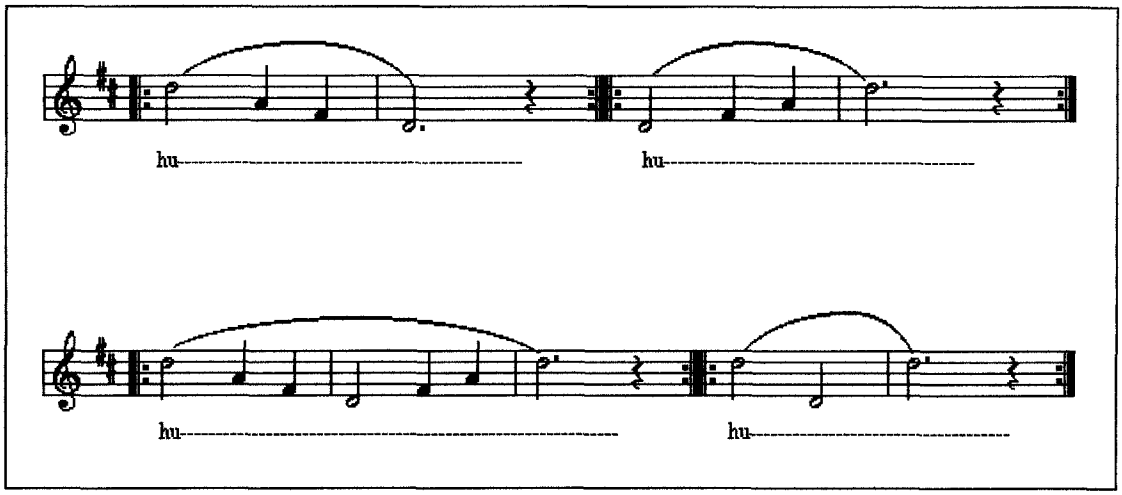

Ah
ah
ah
ah
ah
ah
ah


ah
ah
ah
ah
ah

Increased Singing Ability: Falsetto And Natural Voice Fusion 1

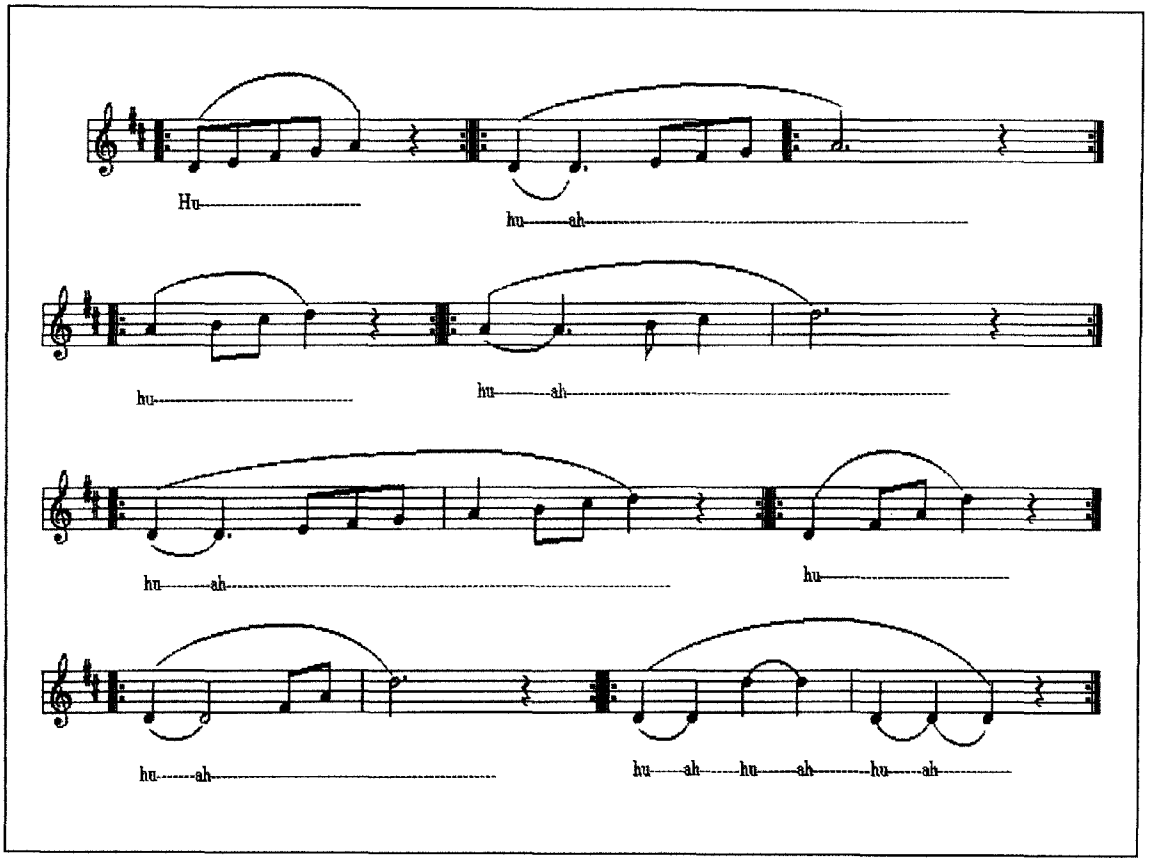
- Now we will work on blending the falsetto and the natural voice.
- Sing using the falsetto voice for both low and high tones.
- Inhale with enough air and be sure not to weaken your voice as you exhale.
- You will then increase the length of singing little by little.
- Echo back the exercises that you hear on the CD (Track 15). Follow along with the music notation below.




hu
hu
hu
hu

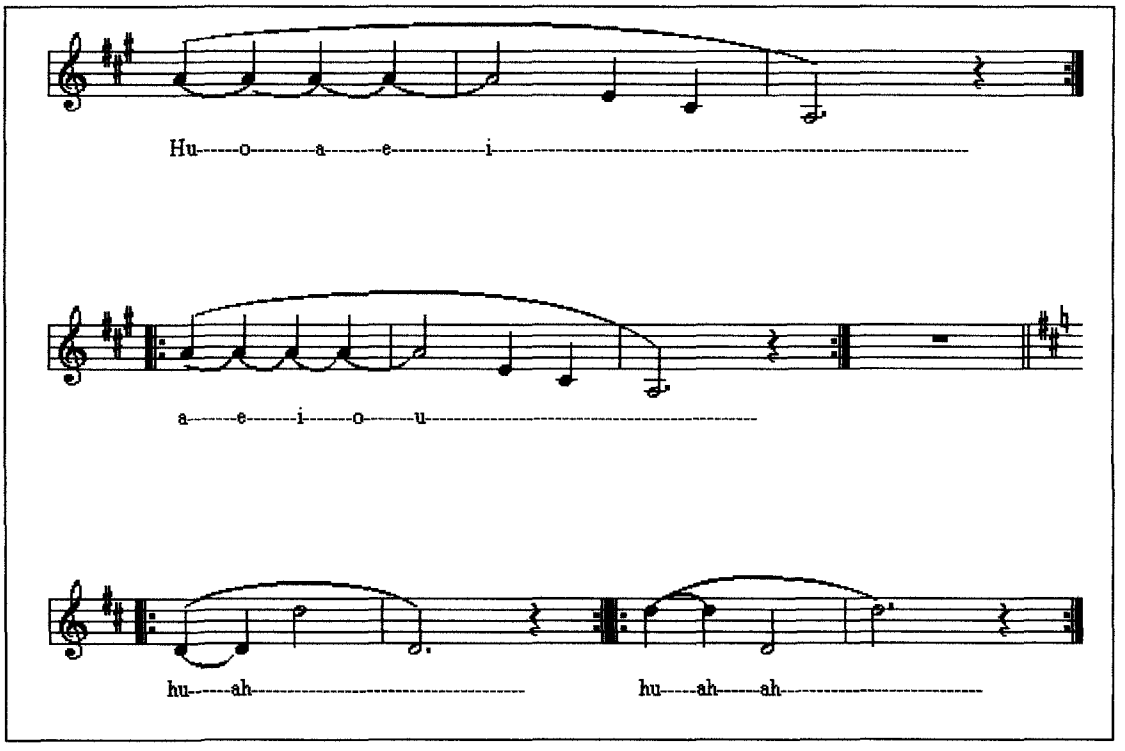
Falsetto And Natural Voice Fusion 2

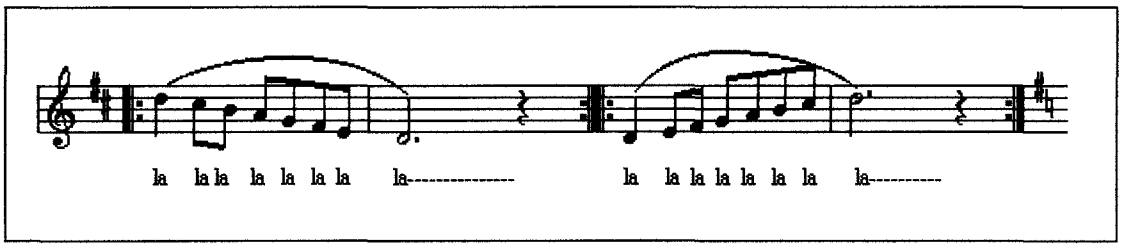
- These next exercises will continue to work on blending the falsetto and the natural voice.
- Start by singing with a clear, low tone voice. It is easier to sing higher tones when beginning, so sing as if using the falsetto voice.
- Carefully and smoothly link the vowels (Hu-ah).
- Be sure to pronounce the "ah" vowel sound clearly.
- Echo back the exercises that you hear on the CD (Track 16). Follow along with the music notation below.



Vowel And Consonant Stabilization

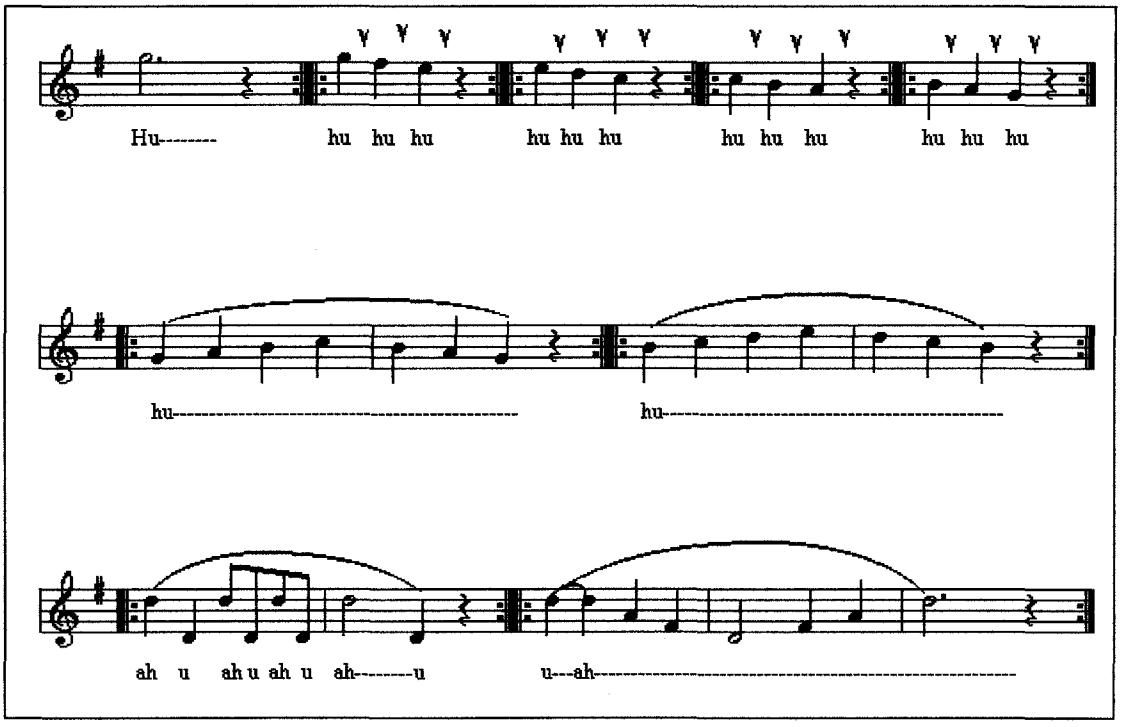
- The next exercises will help you to pronounce both vowels and consonants.
- Be sure not to overemphasize each of the vowels but to have a slight falsetto sense, smoothly linking each vowel in song.
- Stabilize your voice when projecting and be sure not to chatter your jaw when singing the "la" segment.
- Sing "la la la la la" using a falsetto voice. Be sure to keep the jaw still.
- The pronunciation of "l" depends on the motion of the jaw and tongue.
- If you are able to project your voice during the "la" section of it, then it should be very easy to sing the lyrics of a practical song.
- Sing "hu-o-a-e-i" in the falsetto voice "like a woman's voice up high." Be sure not to move the mouth shape.
- Echo back the exercises that you hear on the CD (Track 17). Follow along with the music notation below.

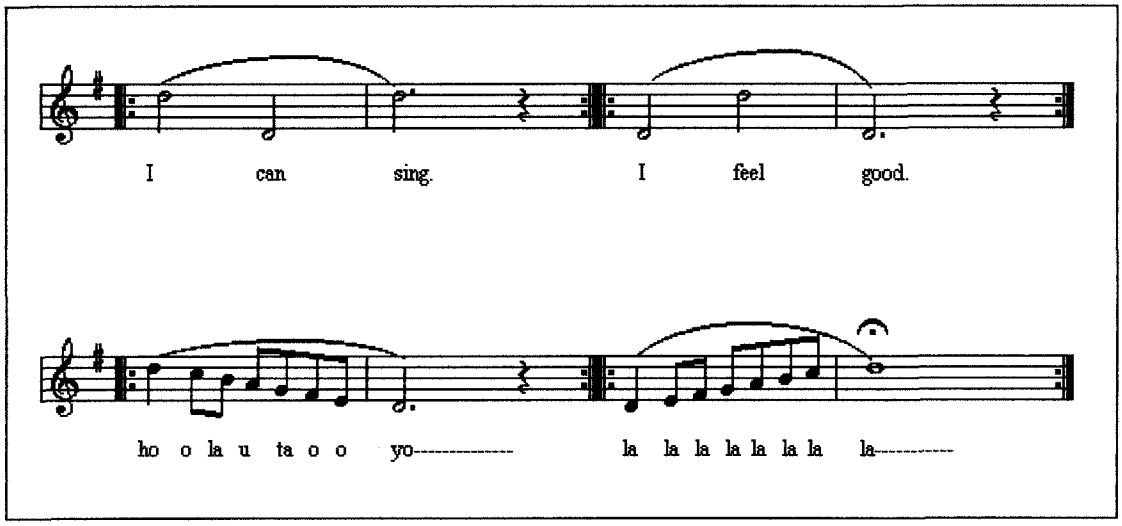



la la la la la la la la----------
la la la la la la la la---------

Singing Muscles Final Check

- Everything you have learned so far is used in this next exercise.
- Using the techniques learned in all of the previous training exercises, sing with confidence.
- The goal is to be able to sing smoothly.
- Use these and other maintenance techniques as warm-up exercises before singing various songs.
- Echo back the exercise that you hear on the CD (Track 18). Follow along with the music notation below.




I can sing. I feel good.
ho o la u ta o o yo---------- la la la la la la la la----------

Congratulations!

- Give yourself a hand for singing skillfully and with confidence!
- Great job on your hard work!
- Here is your "Seal of approval in singing ability."
- "Bam! Pass!"

APPENDIX IX

SONA-SPEECH SOFTWARE MODEL 3600 INFORMATION

Visi-Pitch III/Sona-Speech Model 3900/3600 is by Kay Elemetrics Corp. At 2 Bridgewater Lane, Lincoln Park NJ 07035-1488, www.kayelemetrics.com.

Visi-Pitch III and Sona-Speech have been designed to make standard operations readily accessed by a way of a mouse and pull-down menus. This enables a new user to obtain results easily and quickly. Functions may be entered in one of two ways: selections in the pull-down menus and defined key commands (e.g., the [F12] key will acquire speech data in most modules).

Further flexibility of operation is achieved by configuring the Visi-Pitch II and Sona-Speech software to run through built-in protocols (i.e., a series of command files, or macros) that combine many functions into one instruction. For example, there is one built-in protocol that performs all of the functions necessary to determine and report a subject's habitual fundamental frequency. This built-in protocol and others are selected from the menus and provide a simple way to quickly perform a series of commonly used tasks.

The Real-Time Pitch module has a circular memory buffer that allows a maximum of 60 seconds of data to be captured. Once you record past 60 seconds, the buffers begin to overwrite so that only the last, most recently acquired 60 seconds of data are kept in memory when capture is halted. This circular buffer capability is very handy for acquiring data. The subject can continue to capture data and wait until the desired sample is produced before halting data capture. In other words, there is no need to stop recording

if the subject does not produce the desired sample; simply keep recording until the desired result is achieved and then stop the capture.

The program records up to 60 seconds of data regardless of the duration displayed in the view screen. If a view screen shows a duration less than 60 seconds, 60 seconds or more can still be captured, but only the last part of the data will appear on the screen. To view the earlier portion of data, scroll through it after capture. Both pitch and energy can be displayed simultaneously, or either can be displayed alone.

The Real-Time Pitch module allows the user to mark data for more selective analysis. This task is performed by bracketing the desired data with blue vertical cursors. The user may wish to mark a portion of data to analyze that portion alone, or you may wish to measure the duration of a particular segment.

To analyze the data that you have selected, click on the *Analyze* menu and then *Compare Result Statistics*. All of the statistics pertain only to selected data. The statistical information provided is: Mean Frequency (Hz); Mean Fo; Mean Period (msec); Range (Hz); Minimum (Hz); Maximum (Hz); Standard Deviation (Hz); Vfo; RAP (%); Periodicity; Semitone Range; and Semitones.

Once the data has been marked, you can listen to the marked area by selecting Speak on the Menu Bar, then *Speak Selected*. There are five available Speak operations. All of the Speak operations will redraw and play back the signal or a specific portion of a signal. Speak capabilities, which can all be found in the Speak menu, include: Speak All, Speak Displayed, Speak Selected, From Start to Cursor, and From Cursor to End.

It must be noted that no statistics are available by the company on the *Visi-Pitch* or the *Sona-Speech*. They do offer a bibliography of extant research using the *Visi-Pitch*. Of the research reviewed in music education that has used the *Visi-Pitch*, no statistical information on the accuracy has been cited.

New Product Announcement

Sona-Speech, Model 3600

A Powerful, Low-Cost Speech Therapy Product for Your PC

Applications

- **Visual Feedback in Therapy**
- **Articulation Training**
- **Voice Disorders**
- **Fluency**
- **ESL**
- **Motor Speech Disorders**

KAY

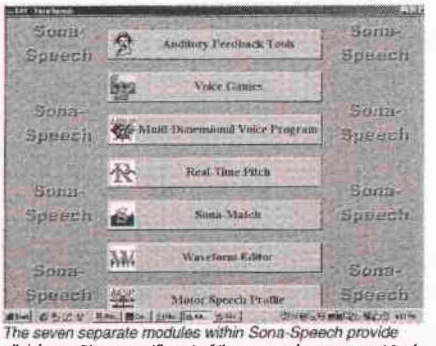
Introduction

Sona-Speech is a low-cost software package based on the acclaimed Visi-Pitch III, also offered by Kay. Instead of using the robust hardware of Visi-Pitch III, Sona-Speech relies on standard sound cards for data acquisition and playback. Clinicians can use Sona-Speech as a therapy tool for a broad range of speech and voice problems. Real-time visual feedback of important speech/voice parameters and quantitative measurements to track patient performance make Sona-Speech an ideal choice for the budget-minded clinician. Due to the performance limitations of most PC sound cards (e.g., compromised dynamic range), Kay recommends Sona-Speech primarily for routine therapy tasks, rather than for research or rigorous voice lab analysis.

Sona-Speech offers the clinician tremendous versatility with seven separate modules (and two optional programs) designed for assessment and treatment of voice, articulation, fluency, motor speech (dysarthria), and other communication disorders. Graphically interesting games are also provided to make therapy tasks appealing to children. Sona-Speech brings the sophistication of Kay software found in Visi-Pitch III to the sound card hardware environment. Clinicians will find Sona-Speech a great bargain and a powerful addition to their clinical tools.

Importance of Speech Biofeedback

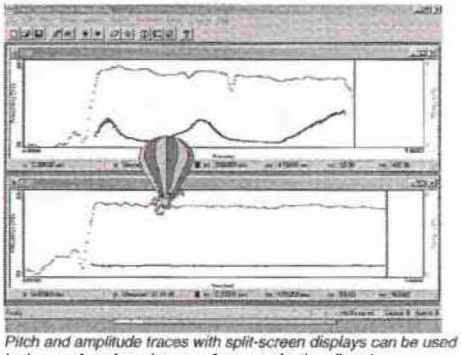
Sona-Speech extracts acoustic parameters (e.g., pitch, amplitude, and spectral characteristics) during speech/voice production and presents these in real time, providing clients with clear, intuitive visual displays. Using target vocalizations provided by a clinician, client attempts can be directly compared both graphically and with auditory playback. Monitoring important speech behaviors with concrete visual displays helps clients reach therapy goals more easily.



The seven separate modules within Sona-Speech provide clinicians with a versatile set of therapy and assessment tools.

Measuring Speech and Voice Behavior

To help objectify a client's baseline performance and subsequent progress, Sona-Speech extracts an extensive range of measurements relevant to speech and voice. In addition to the key speech parameters of fundamental frequency and amplitude, Sona-Speech contains modules with tasks and measurements specific to dysphonia, motor speech problems (i.e., dysarthria), and articulation of selected phonemes. Kay does issue caution regarding the robustness of many sound cards for data acquisition because they can affect the precision of certain measurements. For this reason, Kay recommends Sona-Speech primarily for therapy as opposed to usage in research or voice laboratories.



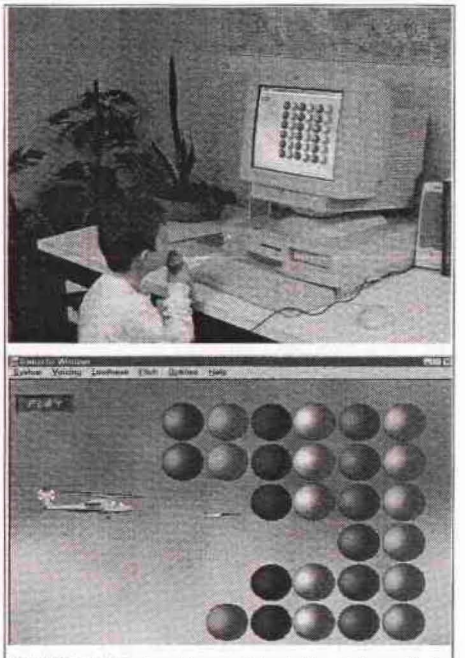
Pitch and amplitude traces with split-screen displays can be used in therapy for a broad range of communication disorders.

Real-Time Pitch

A key program of Sona-Speech is Real-Time Pitch, which displays fundamental frequency and relative intensity in real time. Stress, timing, and intonational patterns, as well as target pitch and/or amplitude values during running speech, can be seen as they are said by the client. Split screens allow target vocalizations to be compared to a client's attempt to imitate critical features (e.g., amplitude levels, speech rate, prosodic patterns, etc.). The target and client's attempt can then be analyzed visually and quantitatively; high-fidelity audio playback allows the clinician and client to listen critically to important speech behaviors.

Waveform Editor

This module is a convenient tool for speech waveform acquisition, editing, and playback. Selected portions of a speech sample can be edited and saved to disk. It can also be used for dual-channel data acquisition. For example, a speech sample can be acquired concurrently with Kay's electroglottograph signal in the second channel.



Innovative games relevant to speech production are fun and help to motivate children in therapy.

Voice Games

The Voice Games module provides animated graphics to represent important speech parameters. The games are particularly effective as therapy tasks for children because they motivate clients and serve as rewards for good speech behavior. Each game can be modified by the clinician to make a task more or less challenging, depending on the child's level of performance.

Sona-Match

Multiple innovative displays of the spectral patterns of sustained phonation (e.g., vowels and sibilants) are offered with the Sona-Match module. These are particularly useful for articulation training, second language acquisition, and the singing voice. Target patterns for each sustained sound can be placed on the screen to allow the client to see the acoustic patterns of their productions against the model. Plots of first formant vs. second formant, as well as true FFT spectral displays, provide the clinician with powerful tools to help clients understand and improve speech articulation.

Multi-Dimensional Voice Program (MDVP)

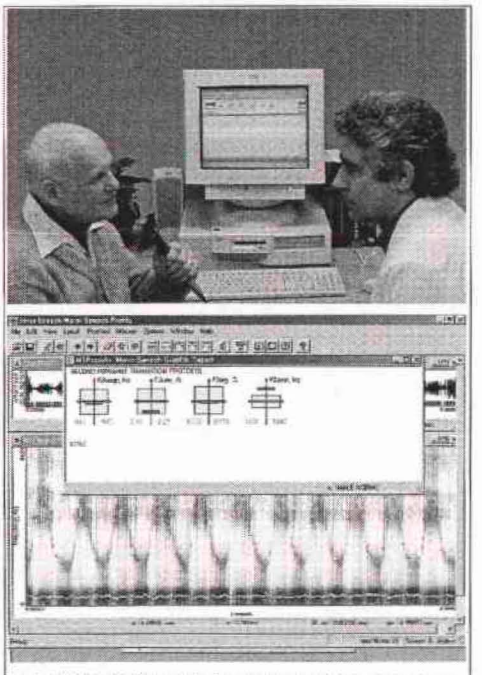
MDVP has been cited in numerous peer-reviewed articles that tout its robust performance for quantitatively analyzing dysphonic voices. In Sona-Speech, MDVP calculates a subset of the measurements offered with Kay's CSL platform. Care is needed when non-professional microphones and generic sound cards are used for sensitive voicing measurements The parameters extracted on a patient's sustained phonation are presented in numerical format or they can be displayed graphically in comparison to a built-in database. An optional CD-ROM of some 700 normal and pathological voices is available to familiarize clinicians with voicing profiles.

Auditory Feedback Tools

Auditory Feedback Tools (AFT) is a module included for critical listening feedback (using headphones) and various types of auditory feedback for achieving desired speech behaviors. A key premise of the Auditory Feedback Tools is the seminal role of auditory feedback in speech therapy. Included are high-fidelity amplification, looped playback of selected speech tokens (e.g., word, phrase, or sentence level), a limited range of Delayed Auditory Feedback (DAF), white-noise masking, speech-rate modification (e.g., prolonging an acquired utterance without altering its pitch on playback), and a metronome pacer. Most of these tools are a subset of features provided in Kay's acclaimed Facilitator, which contains five modes of auditory feedback in a portable (wearable) instrument.

Features

- **Low-cost software for use with standard sound cards**
- **Easy-to-use, Windows-compliant software**
- **Powerful software based on Visi-Pitch III, the leading speech therapy instrument**
- **Real-time pitch and energy displays**
- **Single-keystroke operation of "protocols" commonly used in therapy**
- **Games and graphic rewards to motivate children in therapy**
- **Vowel and sibilant training**
- **Quantitative measurements of voice quality**
- **Special modules for auditory feedback and motor speech analysis**
- **Report generator for outcome analysis**



In the Motor Speech module, the clinician elicits speech tasks to objectify the severity of dysarthria and to help track improvement over time.

Motor Speech Profile (MSP)

Previously an option strictly for Kay's CSL platform, a simplified version of Motor Speech Profile is incorporated into the powerful set of analysis features of Sona-Speech. This program assesses, in depth, the speech performance of patients with motor speech problems (i.e., dysarthria). Distinct tasks are elicited in systematic protocols to help profile the patient's performance; results of each task are quantified and graphed against a normative database. For example, a task for analyzing second formant transitions is sensitive in tracking vowel neutralization often observed in dysarthric patients; similarly, diadochokinetic rate and periodicity have been shown to be closely associated with articulatory agility. As with all Sona-Speech parameters, the data can be included in a report summarizing patient performance.

Optional Programs Available

Sona-Speech users may also purchase separately Real-Time Spectrogram and the Voice Disorders Database options. The Real-Time Spectrogram option displays excellent wideband spectrograms in real time (i.e., as the subject speaks). The Voice Disorders Database option provides an array of disordered voice samples on a CD-ROM to help clinicians achieve an auditory familiarization with a broad range of voice problems.

Host Computer Requirements

Sona-Speech requires a host computer operating Windows 98. It should be a Pentium II or higher with speed >400 MHz. A high-quality sound card, speakers, and a condenser microphone are needed with the PC. Users may wish to consider a USB microphone for improved signal quality. The computer's video graphics card should include accelerators for high-speed graphic display. Finally, the computer should have sufficient hard disk space for storing client speech samples and reports (4 Gbytes or greater recommended).

Summary

Sona-Speech is a low-cost version of the powerful Visi-Pitch III used in sound card hardware environments. Visi-Pitch is used in many hundreds of clinics internationally and has become a standard clinical tool because of its broad applicability to speech/voice disorders. Budget-minded clinicians can take advantage of the very affordable Sona-Speech and apply it to many therapy applications including voice, articulation, auditory feedback, fluency, and motor speech (dysarthria). Sona-Speech complements Kay's family of instrumentation in the areas of acoustics, aerodynamics, laryngeal imaging, electroglottography, and swallowing.

KAY
World Leaders in Speech, Voice, and Swallowing Instrumentation

Kay Elemetrics Corp.
2 Bridgewater Lane ❖ Lincoln Park, NJ 07035-1488 USA
Tel: 1-800-289-5297 (USA and Canada) ❖ (973) 628-6200 ❖ Fax: (973) 628-6363
E-mail: sales@kayelemetrics.com ❖ Web: www.kayelemetrics.com K05M0201

APPENDIX X

PITCH TO HERTZ FREQUENCY CONVERSION CHART

(Kay Elemetrics, 1995; 177)

Cross-Reference for Piano Note/Key Number/Frequency

The octave labeling starts at A in the VRP labeling scheme. For example, the 4th octave (A4 to G#4) contains middle C (C4). By common reference, middle C is described as c'. This is the 28th key on the VRP piano keyboard, and has a frequency of 261.63 Hz.

Octave	Note	Key No.	Frequency (Hz)
	A2	1	55.00
	A#2	2	58.27
	B2	3	61.74
C	C2	4	65.41
	C#2	5	69.30
	D2	6	73.42
	D#2	7	77.78
	E2	8	82.41
	F2	9	87.31
	F#2	10	92.50
	G2	11	98.00
	G#2	12	103.83
	A3	13	110.00
	A#3	14	116.54
	B3	15	123.47
c	C3	16	130.81
	C#3	17	138.59
	D3	18	146.83
	D#3	19	155.56
	E3	20	164.81
	F3	21	174.61
	F#3	22	185.00
	G3	23	196.00
	G#3	24	207.65
	A4	25	220.00
	A#4	26	233.08
	B4	27	246.94
c'	C4	28	261.63
	C#4	29	277.18

Octave	Note	Key No.	Frequency (Hz)
	D4	30	293.66
	D#4	31	311.13
	E4	32	329.63
	F4	33	349.23
	F#4	34	369.99
	G4	35	392.00
	G#4	36	415.30
	A5	37	440.00
	A#5	38	466.16
	B5	39	493.88
c"	C5	40	523.25
	C#5	41	554.37
	D5	42	587.33
	D#5	43	622.25
	E5	44	659.26
	F5	45	698.46
	F#5	46	739.99
	G5	47	783.99
	G#5	48	830.61
	A6	49	880.00
	A#6	50	932.33
	B6	51	987.77
c'''	C6	52	1046.50
	C#6	53	1108.73
	D6	54	1174.66
	D#6	55	1244.51
	E6	56	1318.51
	F6	57	1396.91
	F#6	58	1479.98
	G6	59	1567.98
	G#6	60	1661.22

APPENDIX XI

LETTER OF COPYRIGHT PERMISSION FROM JVC

Karen A. Miyamoto
2506 Akepa St.
Pearl City, Hawaii 96782
Phone: (808) 455-7592
Email: miyamotok001@hawaii.rr.com

January 19, 2003

Mr. Toshi Suzuki
Business Affairs
Legal Business Affairs Department
Victor Entertainment, Inc.

Dear Mr. Suzuki:

This letter is in regards to obtaining legal permission from Victor Entertainment, Inc. to use examples of the Toru Yuba Vocal Training Method in my dissertation entitled: "The Effects Of The Toru Yuba Vocal Training Method On The Vocal Pitch-Accuracy Of Inaccurate Elementary Singers." I am currently a doctoral candidate at the University of Hawaii at Manoa Music Department working towards the degree of Doctor Of Philosophy in Music Education. My committee chairperson is Dr. Arthur Harvey.

I plan to complete the dissertation this May 2003 and will need to obtain a letter of permission to use examples of the Toru Yuba Vocal Training Method in the dissertation. No sound samples will be included in the dissertation. Only musical notation examples with English translation as it appears in the attached unedited copy of what is requested to be included in the dissertation. All illustrations and samples will have the appropriate bibliographical reference with page number to the Yuba Method book (1998) "Muscles for singing." Tokyo, JAPAN: Victor Entertainment, Inc.

The dissertation study selects 60 inaccurate elementary singers, grades four through six at one public elementary school from a population of 320. All 320 students were given a singing pretest to determine their vocal pitch accuracy score, recorded and scored on a computer using Sona Speech Software—a speech therapy product for Personal Computer, by Kay Elemetrics. Students with an average pitch deviance of 100 cents or greater were determined to be inaccurate singers. The inaccurate singers were then placed into three groups—low scores, medium scores, and high scores. Ten students were selected from each of the groups to be in the Experimental Group (Toru Yuba Vocal Training Method) and ten students were selected from each of the groups to be in the Control Group (no extra remediation). The thirty students in the Experimental Group were given the Toru Yuba Vocal Training Method for one forty-five-minute session and given a post singing test to determine whether or not there was any improvement. The Control Group subjects were also given the same post singing test for comparison with the Experimental group singing scores. The results will be analyzed with a double classification analysis of variance and reported.

Enclosed is the unedited copy of what is requested to be included in my dissertation. Please let me know if you have any questions. I look forward to hearing from you soon. Thank you for your kind considerations.

Sincerely,

Karen A. Miyamoto

Karen A. Miyamoto
University of Hawaii at Manoa
Music Department
Doctoral Candidate

Dear Ms. Miyamoto:

We permit you to use Toru Yuba Vocal Training Method in your dissertation within the extent specified in your request letter dated January 19th, 2003.

Masahide Tanino
Group Manager / Legal and Business Affairs
Victor Entertainment, Inc.

BIBLIOGRAPHY

_____ (2003). *The report of a candidate for distinguished researcher medal.* University of Tokyo. Unpublished paper.

Aaron, J. C. (1991). The effects of vocal coordination instruction on the pitch accuracy, range, pitch discrimination, and tonal memory of inaccurate singers. *Dissertation Abstracts International*, 51(09), 2912A. (University Microfilms No. AAC 9103179)

Alderson, R. (1979). *Complete handbook of voice training.* West Nyack: Parker Publishing Co.

Apfelstadt, H. (1984). Effects of melodic perception instruction on pitch discrimination and vocal accuracy of kindergarten children. Journal of Research in Music Education, 32, 15-24.

Appelman, R. D. (1985). *The science of vocal pedagogy: Theory and application.* Bloomington: Indiana University Press.

Atterbury, B. & Silcox, L. (1993). The effect of piano accompaniment on kindergartner's developmental singing ability. *Journal of Research in Music Education*, 41 (1), 40-47.

Bentley, A. (1957). *Children who sing out of tune, 2nd ed.* London: Novello.

Bentley, A. (1966). *Musical ability in children and its measurement.* London: Harrap.

Bentley, A. (1968). *Monotones: A comparison with "normal singers" in terms of incidence and musical abilities.* Music Education Research Papers, No. 1. London: Novello.

Berg, J. V. (1980). Vocal ligaments versus registers. In J. Large (ed), *Contributions of voice research in singing* (123-133). Texas: College-Hill Press.

Birge, Edward Bailey (1955). *History of public school music in the United States.* Reston, VA: Music Educators National Conference.

Blount, R. (1982, February). The singing impaired. *The Atlantic*, 77-78.

Boardman, E. L. (1964). An investigation of the effect of preschool training on the development of vocal accuracy. Doctoral dissertation, University of Illinois. Dissertation Abstracts 25, 1245 (University Microfilms No. 64-1123).

Brophy, T. S. (1997). Authentic assessment of vocal pitch accuracy in first through third grade children. *Contributions to Music Education*, 24, 57-70.

Brown, C. J. (1988). The effect of two assessment procedures on the range of children's singing voices. Unpublished master's thesis, Indiana University, Bloomington.

Buckton, R. (1977). A comparison of the effects of vocal and instrumental instruction on the development of melodic and vocal abilities in young children. *Psychology of Music*, 5, 36-47.

Campbell, M., & Greated, C. (1987). *The musician's guide to acoustics*. New York: Schirmer Books.

Clayton, L. (1986). *An investigation of the effect of a simultaneous pitch stimulus on vocal pitch accuracy*. Unpublished master's thesis, Indiana University, Bloomington.

Cleall, C. (1970). *Voice production in choral technique*. (Revised and enlarged). London: Novello.

Cobes, C. J. (1969). The conditioning of a pitch response using uncertain singers. *Dissertation Abstracts International*, 31, 1829A-1830A. (University Microfilms No. 70-19398).

Collins, D. L. (1981). *The cambiata concept*. Conway, AK: Cambiata Press.

Collins, G. E. B. (2000). Breath-management instruction: Effects on the vocal accuracy and attitudes toward singing of sixth-grade choral students. (Unpublished Doctoral Dissertation, University of North Carolina at Greensboro).

Cooksey, J. M. (1992). *Working with the adolescent voice*. St. Louis: Concordia Publishing House.

Cooper, I. (1963). *Changing voices in junior high: Letters to Pat*. New York: Carl Fischer, Inc.

Cooper, N. (1993). Selected factors related to children's singing accuracy (Doctoral dissertation, Indiana University, 1992). *Dissertation Abstracts International*, 53, 4244A.

Cooper, N. A. (1995). Children's singing accuracy as a function of grade level, gender, and individual versus unison singing. *Journal of Research in Music Education*, 44(4), 222-231.

Culpepper, L. R. (1961). A study of hearing impairments in defective singers. (Doctoral dissertation, George Peabody College, 1960). *Dissertation Abstracts International*, 22, 082816.

Davidson, L. (1985). *Tonal structures of children's early songs*. Music Perception, 2(3), 361-374.

DeYarman, R. M. (1972). An experimental analysis of the development of rhythmic and tonal capabilities of kindergarten and first grade children. In E. E. Gordon (Ed.), *Experimental research in the psychology of music*: 8 (pp. 1-44). Iowa City: University of Iowa Press.

Dittemore, R. M. (1969). An investigation of some musical capabilities of elementary school students. *Dissertation Abstracts International*, 29, 4516A. University Microfilms No. DA8410243.

Eikum, R. L. (1963). Singer vs. non-singer: An educational research study of the Clarkston, Washington elementary school system concerning the students' ability to sing and factors affecting singing ability. Unpublished master's thesis, Unviersity of Idaho, Moscow.

Feierabend, J. M. (1985). The effects of specific tonal pattern training on singing and aural discrimination abilities of first grade children (Doctoral dissertation, Temple University, 1984). *Dissertation Abstracts International*, 45, 110A.

Flowers, P. J., & Dunne-Sousa, D. (1990). Pitch-pattern accuracy, tonality, and vocal range in preschool children's singing. *Journal of Research in Music Education*, 38, 102-114.

Fracht, J. A. & Emmette, R. (1960). *You, too, can sing; voice; its techniques and improvement*. New York: Tudor Publishing Company.

Franklin, E., & Franklin, A. D. (1983). The uncertain singer. *Update*, 1 (3), 3-6.

Franklin, E., & Franklin, A. D. (1988). The uncertain singer. *UPDATE: Applications of Research in Music Education*, 7(1), 7-10.

Gackle, M. L. (1987). The effect of selected vocal techniques for breath management, resonation, and vowel unification on tone production in the junior high school female voice. Unpublished doctoral dissertation, University of Miami, Coral Gables.

Gackle, M. L. (1991). The adolescent female voice: Characteristics of change and stages of development. *The Choral Journal*, 31(8), 17-25.

Garcia, M. (1894). *Hints on singing*. New York: Schuberth.

Gates, R. K. (2003). The owner's manual to the singing voice. Ohio State University. *Dissertation Abstracts International*, 64 (2), 328.

Geringer, J. M. (1983). The relationship of pitch-matching and pitch discrimination abilities of preschool and fourth-grade students. *Journal of Research in Music Education*, 31, 93-99.

Goetze, M. (1985). Factors affecting accuracy in children's singing. *Dissertation Abstracts International*, 46, 2955A. (University Microfilms No. DAI8528488)

Goetze, M., Cooper, N., & Brown, C. (1990). Recent research on singing in the general music classroom. *Bulletin of the Council for Research in Music Education*, no. 104, 16-37.

Goetze, M. & Horii, Y. (1989). A comparison of the pitch accuracy of group and individual singing in young children. *Bulletin of the Council for Research in Music Education*, 99, 57-73.

Gordon, D. S. (1985). A survey of literature and practice in assisting the pitch-defective singer in the elementary school. *Pennsylvania Music Educators Association: Bulletin of Research in Music Education*, 16, 11-18.

Gordon, E. (1978). *A factor analytic description of tonal and rhythm patterns and objective evidence of pattern difficulty level and growth rate*. Chicago: G.I.A. Publications.

Gordon, E. (1979). *Primary measures of music audiation*. Chicago: G. I. A. Publications.

Gordon, E. (1980). *Learning sequences in music: Skill, content, and patterns*. Chicago, IL: GIA Publications.

Gordon, E. E. (1984). *Learning sequences in music*. Chicago: GIA Publications, Inc.

Gould, A. O. (1968). *Finding and learning to use the singing voice: A manual for teachers* (Report No. OEC-6-10-016). Macomb, IL: Western Illinois University. (ERIC Document Reproduction Service No. ED 025 531)

Gould, O. (1969). Developing specialized programs in singing in the elementary school. *Bulletin of the Council for Research in Music Education*, 17, 57-73.

Gratton, M. (1992). The effect of three vocal models on uncertain singers' ability to match and discriminate pitches (Masters thesis, McGill University, Canada, 1989). *Masters Abstracts International*, 30 (4).

Green, G. A. (1990). The effect of vocal modeling on pitch-matching accuracy of elementary school children. *Journal of Research in Music Education*, 38, 225-232.

Green, G. A. (1994). Unison versus individual singing and elementary students' vocal pitch accuracy. *Journal of Research in Music Education*, 42, 105-115.

Greenberg, M. (1970). Musical achievement and self-concept. *Journal of Research in Music Education*, 18(1), 57-64.

Guerrini, S. C. (2002). The acquisition and assessment of the developing singing voice among elementary students. *AAT* 3040318.

Haan, Charles, T. (1977). *Statistical methods in hydrology*. Iowa: Iowa State University Press.

Heaton, C. P. (1992). Air-ball: Spontaneous large-group precision chanting. *Popular Music and Society*, 16(1), 81-83.

Hirano, M., Kirchner, J., and Bless, D. Ed. (1991). *Neurolaryngology: Recent Advances*. California: Singular Publishing Group, Inc.

Humpherys, M. S. (1967). The out-of-tune singer at the college level: Twenty case studies. Unpublished master's thesis, Brigham Young University.

Jaffurs, S. E. (2000). The relationship between singing achievement and tonal music aptitude. Unpublished Masters Thesis, Michigan State University.

Jarjisian, C. S. (1981). The effects of pentatonic and/or diatonic pitch pattern instruction on the rote-singing achievement of young children. *Dissertation Abstracts International*, 42, 2015A. (University Microfilms No. 8124581)

Jones, M. (1971). A pilot study in the use of a vertically-arranged keyboard instrument with the uncertain singer. *Journal of Research in Music Education*, 19(2), 183-194.

Jones, M. (1979). Using a vertical keyboard instrument with the uncertain singer. *Journal of Research in Music Education*, 27, 173-184.

Jones, M. (1993). An assessment of audiation skills of accurate and inaccurate singers in grades 1, 2, and 3. *Update*, 11, 14-17.

Joyner, D. R. (1969). The monotone problem. *Journal of Research in Music Education*, 17, 115-124.

Joyner, D. R. (1971). Pitch discrimination and tonal memory and their association with singing and the larynx. Unpublished master's thesis, University of Reading, England.

Kahane, J. C. (1978). A morphological study of the human prepubertal and pubertal larynx. *American Journal of Anatomy*, 151, 11-19.

Kalmus, H., & Fry, D. (1980). On tune deafness (dysmelodia): Frequency, development and musical background. *Annals of Human Genetics*, 43, 36.

Kay Elemetrics Corp. (1995). *Visi-Pitch application manual*. Lincoln Park, NJ: Kay Elemetrics Corp.

Klemish, J. (1974). Treating the uncertain singer through the use of the tape recorder. *Bulletin of the Council for Research in Music Education*, no. 37, 36-45.

Kramer, S. J. (1985). The effects of two different music programs on third and fourth grade children's ability to match pitches vocally. Unpublished doctoral dissertation, Rutgers University, Rutgers.

Levinowitz, L. M. (1987). An experimental study of the comparative effects of singing songs with words and without words on children in kindergarten and first grade. *Dissertation Abstracts International*, 48, 16497A.

Levinowitz, L, M., Barnes, P., Guerrini, S., Clement, M., D'April, P., & Morey, M. J. (1998). Measuring singing voice development in the elementary general music classroom. *Journal of Research in Music Education*, 46, 35-47.

Mathias, S. L. (1997). A teaching technique to aid the development of vocal accuracy in elementary school students. Unpublished Doctoral dissertation, Ohio State University.

Mizener, C. (1993). Attitudes of children toward singing and choir participation and assessed singing skill. *Journal of Research in Music Education*, 41, 233-245.

Montgomery, T. (1988). A study of the associations between two means of vocal modeling by a male music teacher and third grade students' vocal accuracy in singing pitch patterns. Ann Arbor, Michigan: University Microfilms International 8822404.

Moore, R. S. (1991). Comparison of children's and adults' vocal ranges and preferred self-selected pitches in singing familiar songs. *Bulletin of the Council for Research in Music Education*, no. 107, 13-22.

Moore, D. S., and McCabe, G. P. (2003). *Introduction to the practice of statistics, Fourth Edition*. New York: W. H. Freeman and Company.

Moss, W. A. (1973). Effectiveness of bone conduction hearing in improving pitch accuracy of out-of-tune singers. *Dissertation Abstracts International*, 34, 353A.

Murphy, J. C. (2002). The use of external and internal feedback for the regulation of pitch in different types of singers. *AAT* 1407874.

Music Educators National Conference (1994). *What every young American should know and be able to do in the arts: National standards for arts education.* Reston, VA: Music Educators National Conference.

Myers, L. K. (1965). *Teaching children music in the elementary school, 3rd ed.* Englewood Cliffs, N.J.: Prentice-Hall.

National Assessment of Educational Progress (1974). *The first national assessment of music performance* (Report 03-MU-01). Denver, CO: Education Commission of the States.

Pedersen, D. M., & Pedersen, N. O. (1970). The relationship between pitch recognition and vocal pitch production in sixth-grade students. *Journal of Research in MusicEducation*, 18, 265-272.

Petzold, R. (1966). *Auditory perception of musical sounds by children in the first six grades*. Cooperative Research Project No. 1051, Madison: University of Wisconsin.

Phillips, K. H. (1983). The effects of group breath control training on selected vocal measures related to the singing ability of elementary students in grades two, three, and four. Unpublished doctoral dissertation, Kent State University, Kent.

Phillips, K. H. & Aitchison, R. E. (1995). Effects of psychomotor instruction on elementary general music students' singing performance. *Journal of Research in Music Education*, Volume 45, (2), 185-196.

Phillips, K. H. (1996). *Teaching kids to sing*. New York, NY: Schirmer Books.

Phillips, K. H., & Aitchison, R. E. (1997). The relationship of singing accuracy to pitch discrimination and tonal aptitude among third grade students. *Contributions to Music Education*, 24, 7-22.

Phillips, K. H., & Vispoel, W. P. (1990). The effects of class voice and breath management instruction on vocal knowledge, attitudes, and vocal performance among elementary education majors. *The Quarterly Journal of Music Teaching and Learning*, 1 (1 & 2), 96-105.

Porter, S. Y. (1977). The effect of multiple discrimination training on pitch-matching behaviors of uncertain singers. *Journal of Research in Music Education*, 25(1), 68-82.

Rainbow, E. L. and Hildegard C. Froehlich (1987). *Research in music education: An introduction to systematic inquiry*. New York: Schirmer Books.

Reuter, G. S. (1956). Remedial treatment of inaccurate singers. *Educational Music Magazine*, 35 (3), 19-20, 56-57.

Richner, S. S. (1976). The effect of classroom and remedial methods of music instruction on the ability of inaccurate singers, in the third, fourth, and fifth grades, to reproduce pitches. *Dissertation Abstracts International*, 37, 1447A. (University Microfilms No. 76-19,898)

Roberts, E., and Davies, A. (1975). Poor Pitch Singing: The response of "monotones" to a program of remedial training. *Journal of Research in Music Education*, 23, 227-239.

Roberts, E., & Davies, A. D. M. (1976). A method of extending the vocal range of "monotone" schoolchildren. *Psychology of Music*, 4, 29-43.

Rooks, J. M. (1987). The effect of range training on the vocal accuracy of restricted and complete range K-3 singers. Unpublished doctoral dissertation.

Rutkowski, J. (1986). The effect of restricted song range on kindergarten children's use of singing voice and developmental music aptitude. *Dissertation Abstracts International*, 47, 2072A.

Rutkowski, J. (1990). The measurement and evaluation of children's singing voice development. *The Quarterly: Center for Research in Music Learning and Teaching*, 1 (1-2), 81-95.

Sataloff, Robert T. (1998). *Vocal health and pedagogy*. London: Singular Publishing Group, Inc.

Seashore, C. E., Lewis, D. and Saetveit, J. S. (1939). *Manual of instructions and interpretations for the Seashore measures of musical talents*. Camden, New Jersey: R.C.A. Victor, and 78 r.p.m. record.

Silver Burdett & Ginn (1991). *World of Music*. New Jersey: Silver, Burdett & Ginn, Inc.

Sims, W. L., Moore, R. S., and Kuhn, T. L. (1982). Effects of female and male vocal stimuli, tonal pattern length and age on vocal pitch-matching abilities of young children from England and the United States. *Psychology of Music*, Special Issue: *Proceedings of the IX International Seminar on Research in Music Education*, 104-108.

Smale, M. J. (1987). An investigation of pitch accuracy of four- and five-year-old singers. *Dissertation Abstracts International*, 48, 2013A. (University Microfilms No. DA8723851)

Small, A., and McCachern, F. L. (1983). The effect of male and female vocal modeling on pitch-matching accuracy of first-grade children. *Journal of Research in Music Education*, 31(3), 227-233.

Smith, R. B. (1961). A study of the effect of large-group vocal training on the singing ability of nursery school children. *Dissertation Abstracts International*, 21, 3811. (University Microfilms).

Snedecor, George W. (1956). *Statistical methods, 5th edition.* Iowa: The Iowa State College Press.

Stauffer, S. (1986). An investigation of the effects of melodic and harmonic context on the development of singing ability in primary grade children (Doctoral dissertation, University of Michigan, 1985). *Dissertation Abstracts International*, 46, 1862A.

Sundberg, J. (1987). *The science of the singing voice*. Illinois: Northern Illinois University Press.

Tatem, F. (1990). Effects of selected timbres, tasks, grade level, and gender on vocal pitch-matching accuracy of kindergarten through third grade children. Doctoral Dissertation, University of North Carolina at Greensboro, 1990. *Dissertation Abstracts International*, 51 (09), A3009.

Titon, J. T. (1996). *Worlds of music: An introduction to the music of the world's peoples*, Third edition. New York: Schirmer Books.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley Publishing Co.

Vaughan, M. M. (1980). Intercultural studies in children's natural singing pitch and walking tempo. *Bulletin of the Council of Research in Music Education*, 66, 96-101.

Wassum, S. (1979). Elementary school children's vocal range. *Journal of Research in Music Education*, 27 (4), 214-224.

Welch, G. F. (1986). A developmental view of children's singing. *Brittish Journal of Music Education*, 3 (3), 295-303.

Welch, G. F., Howard, D. M., and Rush, C. (1989). Real-time visual feedback in the development of vocal pitch accuracy in singing. *Psychology of Music*, 17, 146-157.

Whitman, C. F. (1970). The effects of piano instruction on vocal pitch accuracy in fourth, fifth, and sixth grade students. Unpublished Doctoral dissertation, University of Florida.

Wilson, D. S. (1970). A study of the child voice from six to twelve. *Dissertation Abstracts International*, 31, 5453.

Wing, H. D. (1947). *Wing standardized test of musical intelligence*. Windsor, England: NFER Publishing Co.

Wyatt, S. C. (1993). The effect of group instruction in pitch accuracy and in head voice on the intonation and range of second-grade children. Unpublished Doctoral Dissertation, University of South Carolina.

Yarbrough, C., Bowers, J., & Benson, W. (1992). The effect of vibrato on the pitch-matching accuracy of certain and uncertain singers. *Journal of Research in Music Education*, 40, 30-28.

Young, R. W. (1952). *A table relating frequency to cents*. Elkhart, IN: C. G. Conn.

Yuba, Toru (1998). *Dream voice training*: *Muscles for singing* (with CD). Tokyo, JAPAN: Victor Entertainment, Inc.

Yuba, Toru (2000a). Method for Curing Motor-Related Off-Key Singing: Is Tone Deafness Really Deafness? *The Japan Journal of Logopedics and Phoniatrics*, 41, (4).

Yuba, Toru (2000b). *How to make a count of off-key singers*. Okuyama, Japan. Unpublished Paper.

Yuba, Toru (2001). *How to Train the Muscles for Singing*. Okuyama, Japan. Unpublished paper.

Yuba, Toru (2002). Cause of Off-Key Singing and Its Curative Education. *Journal of Otolaryngology, Head and Neck Surgery*, Tokyoigakusha Inc., June.

Yuba, Toru (2003a). Cause of Off-Key Singing and Its Curative Education. *Journal of the Indian Musicological Society*, August.

Yuba, Toru (2003b). *The Definitions of the Natural Voice and Falsetto Voice*, June. Okuyama, Japan. Unpublished paper.

Zimmerman, M. (1971). *Musical characteristics of children*. Reston, Virginia: Music Educators National Conference.

Zwissler, R. N. (1972). An investigation of pitch discrimination skills of first grade children identified as accurate singers and those identified as inaccurate singers. *Dissertation Abstracts International*, 32, 4056A-4057A. (University Microfilms No. AAC 7202947)