

REVIEW OF *STREAMING SPEECH*

Title	<i>Streaming Speech</i>
Author	Richard Cauldwell
Platform	PC
Minimum hardware requirements	Windows based multimedia computer, Processor at least 350MHz, Windows 95, 98, NT (SP4 or above), 2000, ME, XP with Internet Explorer 5 or later, CD-ROM and Windows-compatible sound card, microphone, headset or speakers.
Producer	Richard Cauldwell, P. O. Box 10662, Birmingham B17 0ZE, United Kingdom +44 (0) 121 240 9804, e-mail: Richard@speechinaction.com
Support offered	Comprehensive Web site http://www.speechinaction.com/
Target language	English
Target audience	High Intermediate to advanced learners
Price	Prices start at 40.00UK for a single user license, plus VAT at 17.5 % for UK & EC; Student's Book 12.50UK plus shipping and handling
ISBN	0-9543447-0-7

Review by Andrew Lian, Rice University

Streaming Speech by Richard Cauldwell addresses, in an interesting and theoretically-argued way, problems of speech perception and production as experienced by foreign language learners. The approach used, while focusing on English, is not language-specific.

TECHNICAL MATTERS

Streaming Speech comes in the form of a CD inside a modest-looking package with minimal paper documentation. It installs effortlessly. The material takes the shape of a Web-based interactive system (it actually runs in Internet Explorer 5 or above) and is written in a Web-based authoring system called Fabris (<http://www.fab24.net>). This system is loosely based on *Toolbook* by Asymetrix. *Streaming Speech* has the potential to be deployed across the Internet (a version of it is, in fact, already available as a demonstration http://www.fab24.net/SS_Demo/guest1.htm). In the long term, the ability to be served across the Internet leading to the potential to integrate with other systems will enable *Streaming Speech* to increase its user base as well as increase its range of related teaching and learning support. The system, which consists of lesson materials and a recording applet installs itself painlessly and with no detectable software problems.

PEDAGOGIC MATTERS

Streaming Speech is a self-study computer package arranged in 10 chapters. It presents itself as a series of attractive pages, which are meant to be navigated sequentially within each chapter. Learners are also able to connect to any part of the program at any time through access to a layered series of menus some with drop-down boxes. The interface provided is simple yet information-rich, allowing learners to travel easily from point to point in response to their perceived needs. The page in [Figure 1](#) shows the range of options usually available at any one time, including a drop-down menu of the chapter content.

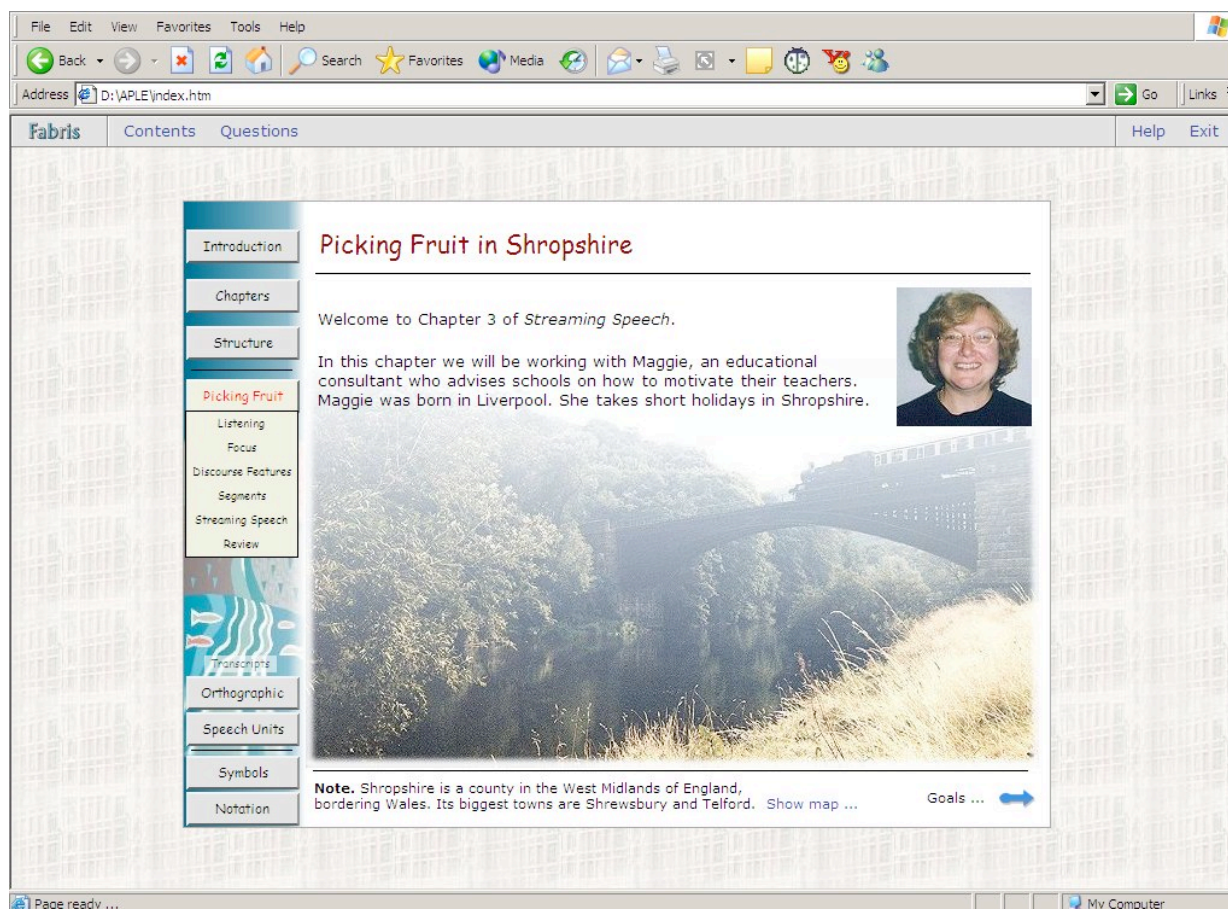


Figure 1. Menu structure

The richness of the menu structure is supplemented with some presentation niceties. For instance, one of the introduction pages (see Figure 2) features a photograph of the author (top right) and small photographs of the speakers. When the mouse passes over each speaker some biographical information is displayed in a popup window and the speaker's voice is heard. This side-by-side arrangement allows quick contrasting and comparing of voice quality, speech rhythms, and delivery, while personalising the learning experience and reinforcing the sense of friendship/closeness which the author clearly tries to generate between himself, informants, and learners. Disappointingly, the author's picture remains silent.

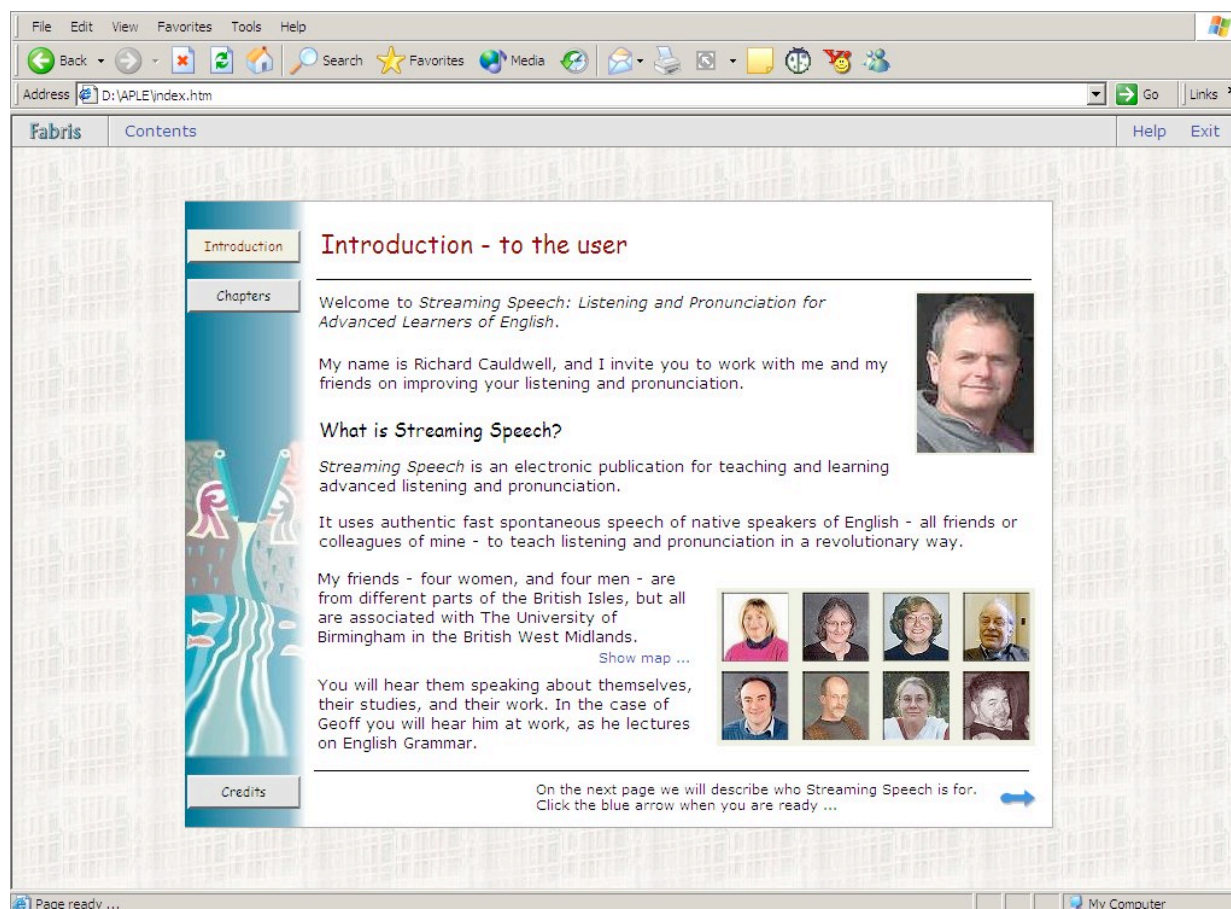


Figure 2. Introducing the approach

The Introduction to *Streaming Speech* identifies potential users (upper intermediate/advanced learners of English and language teacher trainees) and presents some key concepts. These include a description of "fast spontaneous speech," the main object of study in the program, which is defined as being between 250 and 500 words per minute with significant inter and intra speaker variations depending on context and communicative needs. The introduction also describes the format of each chapter.

The stated objective of the resource is to use "authentic fast spontaneous speech of native speakers of English -- all friends or colleagues of mine -- to teach listening and pronunciation in a revolutionary way."

Further, the author states,

If you have problems handling fast speech in listening, and problems in communicating fluently, *Streaming Speech* will help. You will learn to bridge the gap between slow and fast speech. *Streaming Speech* will present extracts of fast speech to do two things:

- first, train your ears to hear and understand
- second, train your voice to speak at speed with accuracy and fluency

You will use expert speakers (my friends) as your model: you will imitate them as accurately as you can, at the same time and at the same speed as their original speech.

You will learn to handle fast speech, so you can understand it when you hear it, and speak it when you need to.

This sets the scene, establishes priorities, and reveals the author's theoretical framework. Perception comes first, then production. What is not stated explicitly but is implied, is that perception and production are viewed as mutually reinforcing processes and that working with them in the way described above results in a cyclical process of progressive refinement of both.

The first eight chapters of the program are based on the discussion and analysis of eight speakers of British Isles English all identified as friends of the author and all having a connection with the University of Birmingham. With one exception, they all produce unscripted speech as part of an interview and talk spontaneously about their personal lives and the things that matter to them. The exception just mentioned is in chapter 5 and consists of a fragment of a university lecture. Thus the corpus of speech samples is fairly small and consists of the voices of educated people marked by some common regional and cultural variations. This choice of example voices, though probably offering too little spoken language for a full listening course and a full sensitisation to the range of British Isles English, is a reasonable one given the practicalities of courseware production.

According to the author,

Streaming Speech teaches you to listen by using a standard three stage procedure.

First you get information about the speaker and the topic; second, you get an activity to do while listening; and third, you focus on those parts of the recording that contain the answers for the activities. You will do this in the first two sections of each chapter.

The focus will be on the fastest meaningful sections of the recording, or sections which illustrate important features of the stream of speech. ... To help you focus on these stretches of speech, you will see an extract in speech-units: you will be able to click on any line of the extract and hear it.

In addition, in the third section of each chapter (named "Discourse Features") you are taught about a feature of the stream of speech (such as loss of, and merging of sounds; rhythm, level or falling tones) and then you have to identify the same features in another extract of speech. This section involves ear-training to make you comfortable with the features of fast speech.

This structure is adhered to consistently with awareness-raising exercises provided at each step of the way. For instance, the screen below ([Figure 3](#)) is meant to bring to the learner's notice the prosodic features of a speech-unit. It does that by using a special notation which identifies stressed syllables (uppercase letters or large circles) and which is arranged in the shape of an intonation curve. Clicking the speaker icon will play the speech-unit and each syllable is highlighted as it is pronounced. While there is plenty of evidence that learners have difficulty in perceiving correct prosody, the feedback provided here can begin to make learners aware of where stress patterns are supposed to occur and how speech melody is organised.

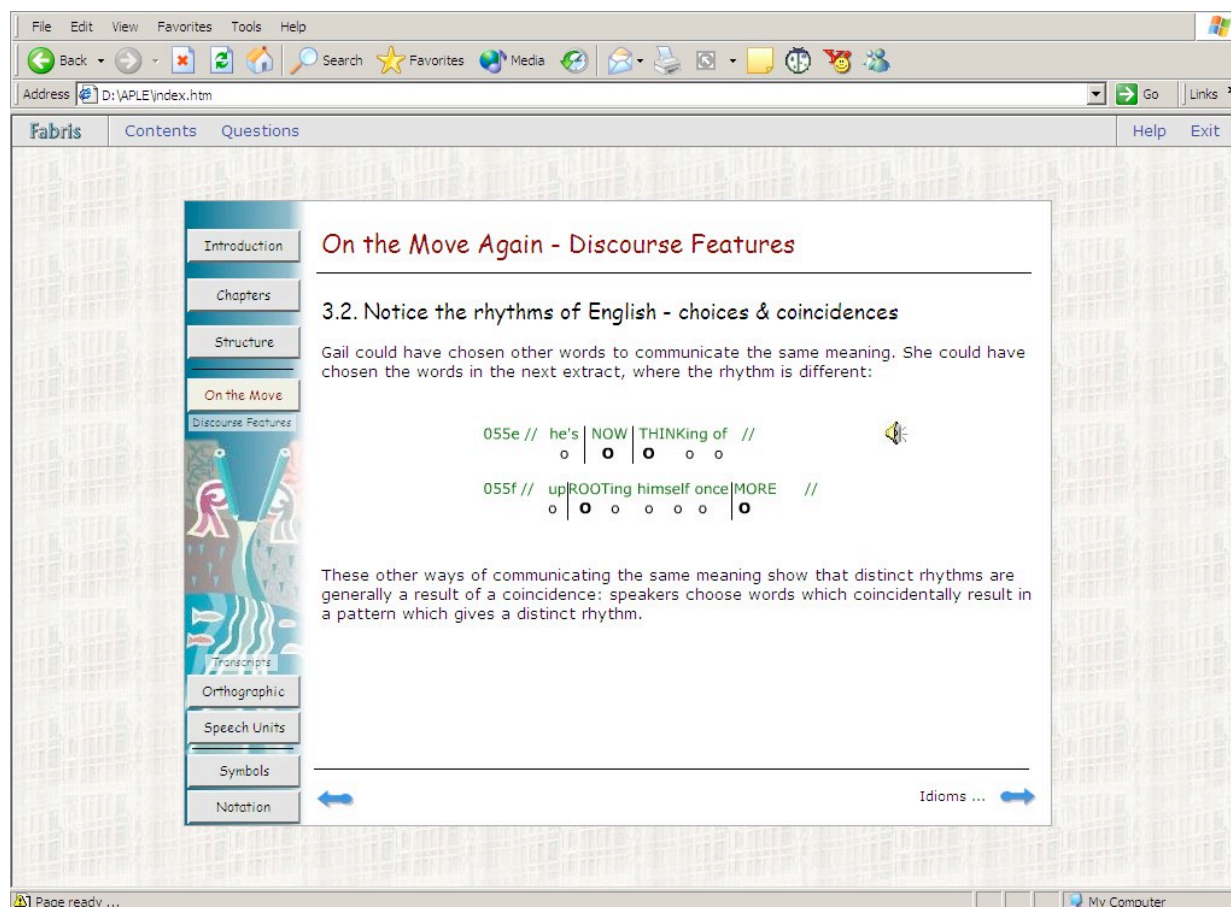


Figure 3. Raising awareness of stress and intonation patterns

There are many similar awareness-raising exercises built into this program. The two screen displays which follow represent some of the more complex ones.

The screenshot shows a web browser window with the address bar displaying 'D:\APLE\index.htm'. The browser's menu bar includes File, Edit, View, Favorites, Tools, and Help. The main content area is titled 'You were at Oxford? - Focus' and displays a transcript of a conversation. The transcript is organized into sections: Introduction, Chapters, Structure, At Oxford?, and Focus. The 'Focus' section is currently selected, showing a transcript of a conversation between Philip and Richard. The transcript includes time markers and a focus on the oral examination. The transcript text is as follows:

2.3. The Oral examination
 The answer is that Richard was interested in **B. the fact that people could not fail the oral exam**, as indicated in units 095-107:

Philip	095 // → in <u>FACT</u> you could	// 259
	096 // → <u>YOU</u> could	// 140
	097 // → go <u>RIGHT</u> through and get a de <u>GREE</u> with	// 338
	098 // → in <u>THOSE</u> days	// 257
	099 // → <u>CERTAINLY</u>	// 128
	100 // → without <u>SPEAKING</u> a <u>WORD</u> of the <u>LANguage</u>	// 161
	101 // → because you <u>COULDN't</u> fail the <u>Oral</u>	// 250
	102 // [pause]	//
Richard	103 // → you <u>COULDN't</u> fail the <u>Oral</u>	// 300
Philip	104 // → you <u>COULDN't</u> fail [the <u>Oral</u>]	// 356
Richard	105 // → [there <u>WAS</u> one]	// 783
	106 // → but you couldn't <u>FAIL</u> it	// 492
Philip	107 // → <u>YES</u>	// 330

Below the transcript, there is a section titled 'And you did not answer!' with a blue arrow pointing to the right. At the bottom of the page, there is a status bar with the text 'Page ready ...' and 'My Computer'.

Figure 4. Analysis of section of speech

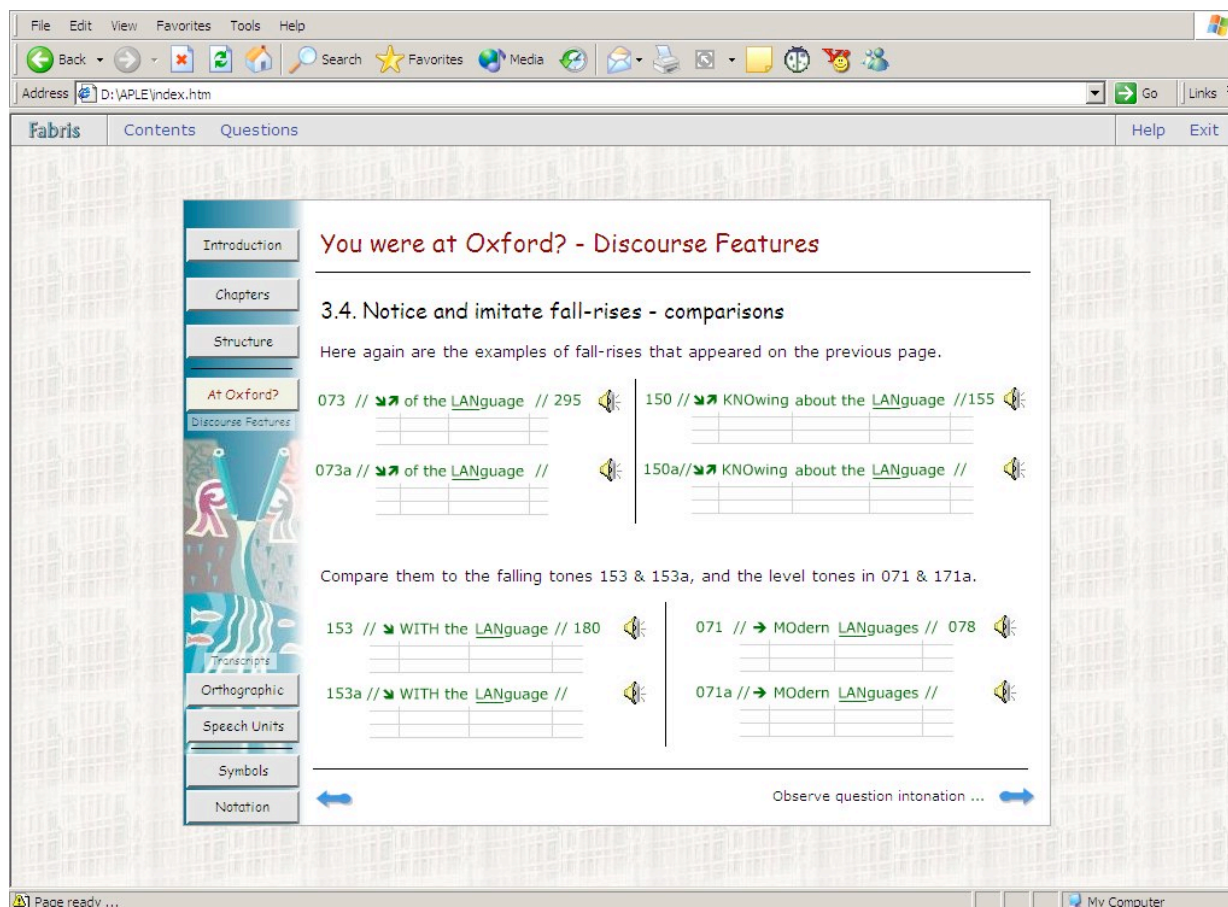


Figure 5. Raising awareness to prosodic features

While awareness of prosodic features is crucial to the achieving the author's objectives, the "fast speech" aspect of the programs is further developed through specific exercises which, for instance contrast slow "dictionary" pronunciations of individual words with the same words produced in natural language. This kind of exercise is valuable in that it heightens learners' awarenesses that "words" simply do not exist in natural spoken language. In time, armed with their new sense of awareness, learners' expectancies as to the content of natural spoken language will change.

Awareness and listening exercises are reinforced through the pronunciation practice which is available throughout the program. This is based on a record and compare approach. Learners are required to match the models provided in every way, thus hopefully developing their sense of the phenomena under scrutiny. However, they are asked not only to record a final product such as "fast speech" in action, but also to record some of the language which contrasts strongly with "fast speech," for example, "dictionary" pronunciations of words. This is so as to give them opportunities for experiencing contrasts not only as listeners but as producers, too. These changes in forms of perception are designed to give learners the opportunity to listen and perceive differently, thus increasing the probability of changing the ways they both hear and produce. While learners are likely to be able to detect gross errors in their performances (e.g., speed mismatches between the model and themselves), such an approach is problematic in the more subtle areas of pronunciation and runs the risk of reinforcing incorrect pronunciation habits unless the learner's perceptions have been sufficiently sharpened to enable some degree of self-correction. To counteract the problem, it might have been possible to incorporate a speech recognition engine of one kind or another (e.g., the popular Auralog *Tell Me More* <http://www.auralog.com>). These programs, while improving in their functionality probably have some way to go before they are fully reliable. It is possible

to envisage other ways of providing feedback such as incorporating visual displays. However, speech recognition systems and visual displays are both expensive to incorporate and potentially difficult to manage.

Though clearly linked to the production of correct sounds and prosody, the speed-matching exercises are less likely to be problematic as ample support is provided to develop awareness of the global timing issues involved. While the provision and nature of any feedback remains problematic in all programs such as this one, the level of information and awareness-raising provided by this particular system through its exercises, animations, and tests is likely to enable learners to begin making inroads into their perceptual mechanisms. Of course, the ultimate effectiveness of the package will need to be tested over time.

The content for the approach just outlined is based on an integrated view of speech phenomena. Each of the first eight chapters deals with a combination of features. For instance, chapter 2, "On the Move Again," says it is about long vowels but lists the following activities: "speak in a distinct rhythm; speak without a distinct rhythm; use prominences, non-prominences and pauses -- 360 words per minute."

Thus, this program adopts a standard intellectual position in relation to articulation phenomena, appearing to focus on speech *sounds* but in fact doing something else: working on prosodic features, rate of delivery and articulation phenomena. In a world where the study of pronunciation is still very strongly focused on the study of individual sounds, such an approach makes discursive sense.

Chapter 9 consists of a "segment workshop: choose a speaker to work on the vowels and consonants of English." Learners are asked to select from one of six speakers on whom they would like to model their voices. This is in itself a positive original approach designed to enhance learners' comfort levels and enable them to "tune in" to the kind of English that they wish to produce. The same approach is used here in relation to fast and slow speech but this time there is special emphasis on particular groups of vowel and consonants grouped together.

There is also a special section called the "cluster-buster" where an attempt is made to deal with problems of consonant cluster, a difficult area for many learners of English.

Finally, chapter 10 provides theoretical and practical information in relation to "speech units" as well as some practice in transcribing spoken language according to the system used throughout the program. While this chapter may seem to be of primary relevance to potential teachers or persons wishing to develop a better understanding of the prosodic features of English, learners will undoubtedly benefit from the pencil and paper transcription exercises offered. This is because transcription activities provide one of the few possibilities for learners to confront what they *think* they heard with what is actually there without the need to resort to the support of another person. The accuracy of their transcription will give them at least the beginnings of an understanding of where their *individual* problems lie in a way which is meaningful to them.

CONCLUDING REMARKS

This review has been written from a perspective that positions learners as central in the process of learning and that identifies language learning as no more than a special case of learning in general. A further principle is that the act of learning entails the construction of internal logical and representational systems for making sense of the phenomena which surround us. In that perspective, it is further assumed that learning is neither linear nor tree-structured but, rather, is rhizomatic in nature with connections and flows being made by learners in ways which are individual and an outcome of their personal histories.

Streaming Speech is an ambitious program which deals with important issues of listening and speaking. Unlike many other commercially-available products it draws on a specific theoretical base (the work of [David Brazil](#)) to legitimize its actions. It is clear that the author has an in-depth understanding of the issues raised by Brazil and makes good argued use of Brazil's work. The author is also highly committed

to continuing intellectual activity through his work in his [Centre for Discourse Intonation Studies](#). While Cauldwell clearly admires the work of David Brazil and may even think that he is making the case for Brazil's work he is, in my view, rather making the case for a form of interactive learning based on constructivist-like approaches.

While Brazil's work offers, in this specific context, a set of categorisations which provide a focus and a way of talking about speech phenomena, the major strength of the program lies in the ways in which awarenesses or perceptions are raised and adjusted through a feedback-modified listening-speaking loop -- much as it occurs in real life but with enhancements. The problem is to assess the effectiveness of that loop in the context of this program.

Importantly, the program focuses on the realities of language in action. We no longer have simplified, "sanitised" language as the basis for study. In a theoretically-related perspective, the program, correctly in my view, steers away from the impossible Krashenian task of adapting input to the learners' level in a mass market context. Rather, it offers the following statement/question: "This is what you've got, how do we deal with it?" To re-formulate this in the context of language-learning, the reality is that natural speech is fast, complex, and fuzzy whereas teachers often tend to focus, in a commonsensical and therefore discursively powerful way, on the slow, simple, and clear, to the detriment of their learners (at least in my view). Rather than providing learners with comprehensible input, the approach here is to provide learners with the tools which will enable them to develop internal mechanisms for making ordinary, everyday, language comprehensible. These mechanisms can be generalised beyond specific texts and should help learners to become self-managing in due course.

The fundamental problem with mastering (or even competently approximating) the sounds of a foreign language depends largely on the ways in which the learner makes sense of the sound input. Both native speakers and non-native speakers are exposed to the same sound waves. But they do not make sense of those sound waves in the same way. This is because the non-native speaker's past (language) experience enforces a specific way of organising sounds. The perceptual mechanisms accept interpretations that have been legitimised by experience and refuse interpretations rejected by experience. Thus the trick to correct structuring, and therefore learning, is to make legitimate that which had, so far, been refused legitimisation by our perceptual/organising mechanisms.

The question to ask in the context of *Streaming Speech* is whether, the program does enough to change learners' ways of organising input. On the one hand, the program makes considerable and valuable use of a number of procedures for raising awarenesses ranging from transcription to the use of animation. The intention seems to be that the awarenesses generated will in fact enable the beginnings of a re-structured processing. However, it should be noted that at no time is the student's output verified in any way by any system or anyone other than themselves. Any analysis which occurs is in the form of what might be called analysis by proxy, e.g. when the learners perform a transcription and discover that what they wrote down was in fact not what was expected. This form of feedback then gives them the opportunity not only to listen again but to listen *differently*. It is through the performance of this kind of work that perceptual structures can begin to be modified when using a system such as *Streaming Speech*. In addition, though, the provision of large amounts of information at every point of the program will certainly maximise the potential of the awareness-raising exercises to make inroads into the learners' perceptions.

It is arguable that an approach which focuses on fast speech is possible only because the program is dealing with intermediate/advanced learners. This is based on the notion of teachers "knowing" what is difficult and what is not. While that may be true in a statistical sense, there is no guarantee that the "statistics" will apply to any individual learner. Rather than "statistically" determining difficulty and setting the program sequence, it might be more valuable to provide learners with opportunities for grappling with a range of problems, discovering what is difficult for them and at least contribute to setting their learning agenda. In effect, this is what *Streaming Speech* does. Further, the approach developed here

can be adapted to many different texts (with different kinds of speech-units) in many different contexts to match needs and motivations. To put it another way, it may well be possible for less advanced learners to derive benefits from this approach.

Finally, will learners become fluent listeners and producers of fast speech? That will depend on many factors, including more opportunities to process speech and also access more of the same kind of materials (the corpus of speech samples in the program is quite small). What is likely, however, is that they will all have developed some sensitivity to some of the issues involved, that they will be able to generalise some of their experiences and have the opportunity better to develop their comprehension and production processes through interaction with real-world texts.

In that context *Streaming Speech* makes an interesting and valuable contribution to language teaching and learning.

ABOUT THE REVIEWER

Andrew Lian specialises in the methodology of teaching foreign/second languages and has a special interest in the uses of modern technology to enhance learning. He is Professor of Humanities and Director of the Center for the Study of Languages at Rice University, Houston, Texas, and Emeritus Professor of Languages and Second Language Education at the University of Canberra, Australia. Previously, he had been Professor of Modern Languages at James Cook University and Professor of Computer-Enhanced Language-Learning at Bond University, Australia.

E-mail: lian@rice.edu