Elicitation and documentation of tone

What is the relation between qualitative and quantitative approaches in the study of tone?

1. Dimensions of investigating tone

The challenge of investigating tone

The segmentals as a point of reference

In relation to segmental speech sounds such as stop consonants, for most researchers the auditory impressions match up with places and manners of articulation. In other words, most researchers know how this phonetic space is carved up, and can relate this knowledge to their auditory impressions.

The use of phonetic symbols and brackets – // vs. [] – enables us to relate the phonetic realisations to the system of contrasts. Often the perception is more fine-grained than the system contrasts.

Then tone…

In relation to tone, auditory perception is typically less fine-grained, that is, most researchers are less able to pick up phonetic detail, as compared to segmentals.

[[Illustration]]

It is not just our auditory perception of the phonetic space that is less good than on the segmental side. Some important descriptive tools – i.e., the IPA and autosegmental representations – are also less well fleshed out in relation to tone. Up to four phonological levels or up to five? One or two patterns of timing in two-target contour tones?

In descriptive studies of tonal phenomena, it is not uncommon to omit discussion of the phonetic realisation, going straight to the system of contrast. One factor here is autosegmental theory, where the main concern is to represent contrast. Chao tone letters enable us to represent great phonetic detail. What approach to notation works for you?
Phonological complexity in tone

“Tone is like segmental phonology in every way – only more so!” (Hyman 2011)

Tone systems show huge cross-linguistic diversity in terms of the degree and the nature of phonological complexity – e.g. intricate systems of tone sandhi; interaction with prosodic domains or intonation; etc.

[[Illustration from discussion of Chumburung in Snider (2014:717-719)]]

Chumburung citation forms:

\[
\begin{align*}
\text{k̀pà} & \quad \text{‘want’} & \text{kpbá} & \quad \text{‘path’} \\
\text{njì} & \quad \text{‘know’} & \text{njí} & \quad \text{‘mother’}
\end{align*}
\]

Chumburung nominalisations of verb stems:

\[
\begin{align*}
\text{ki-kpbá} & \quad \text{‘wanting’} \\
\text{kJ-jní} & \quad \text{‘knowing’}
\end{align*}
\]

The case to postulate abstract underlying representations is compelling here.

In another, the inventory may be rich but the tonal specifications may be very stable, so that there is little evidence for a divergence between underlying and surface-phonological specifications.

[[Illustration from Shilluk in Remijsen & Ayoker (2014)]]

The phonological diversity among tone systems is striking. And this diversity is to be expected not just between related languages but also among mutually intelligible dialects. Because of this diversity and complexity, we can dearly use the fine-grained level of discernment that is at the disposition of most of us in relation consonantal place or manner of articulation.
What does a description and documentation of a tone system include?

A comprehensive description covers:

a) the inventory of lexical and morphological specifications;
b) the phonotactic distribution of these specifications;
c) the contextual processes involving word-level specifications (sandhi);
d) the interaction between word-level tone and intonation;
e) the realisation of the tone patterns.

What does a documentation of tone involve?

What does it mean to not just describe but also document a tone system? Accountability is the key concept in language documentation (Himmelmann 1998, Woodbury 2003). Given our trouble with ascertaining tonal melodies, what we need more than anything else is sound: the phenomenon from which phonological representations abstracts away, in the case of tone more so than in the case of segmentals. This is particularly valuable because our auditory impressions are not always reliable.

Beyond accountability, the sound examples illustrate the tone patterns better than diacritic tone marks. The diacritics do not get the message across, in part because they are vague: the IPA convention underrepresents the range of possible tone patterns (cf. below on alignment in contour tones).

[[Illustration from Shilluk in Remijsen & Ayoker (2014) – paradigm forms in Table I.]]

Best practice in including sounds

… is embedding rather than including as supplementary materials. *Language Documentation & Conservation, Linguistic Discovery, Journal of the International Phonetic Association* (Illustrations of the IPA), *Language* and *Phonology* are among the journals that have a track record.
How important are spontaneous speech data in relation to documentation of tone?

Corpus data are center-stage in language documentation in general – how about specifically in relation to tone? If a tone language is rich in terms phonological processes, a transcribed corpus will not reveal how the system works, unless both underlying and surface-level specifications are represented. The key issue is how to satisfy the requirement of accountability.

Summarizing – the advantage of recording sound data and including them in publications:

When the evidence in support of the various components of the description of a tone system is recorded and made available, then the analysis becomes more accountable and more intelligible.

The importance of qualitative exploration to the study of tone

Controlled elicitation

Controlled elicitation is essential to the discovery of a tone system. This methodology enables us to pry apart the system, one potential factor at a time. The procedure of controlled elicitation in relation to tone is laid out in detail in Pike (1948), for many decades the primary didactic resource available to students of tone. In her appraisal of Pike (1948), Yu (2014) explains that this is in essence an experimental procedure: whereby a range of factors as possible are kept constant, while one or more are manipulated systematically. Examples can be found in Hyman (2014) and Snider (2014), among many others.

Because tone can interact with so many parts of the language system, this exploration is often closely connected with general descriptive work on the sound system and the morphosyntax (see e.g. Van de Velde 2009).

Didactic materials on how to study tone

Volume 8 of Language Documentation and Conservation (2014) includes a set of papers focusing on this topic, guest-edited by Steven Bird and Larry Hyman. It covers a range of
angles, topics, and case studies. A very welcome resource: there was not much beyond Pike (1948). Ear-training materials can also be useful – the most detailed resource in this respect is Smalley (1963). It includes a wide range of sound-based exercises specifically on tone.

The first suspicion of tone

All spoken languages has f0 / pitch in voiced segments, so whether tone is specified at the word-level is not evident at face value just from listening to an utterance. The first step in this process is to start entertaining the hypothesis that there could be a lexical specification. Sometimes a melody just strikes us. Sometimes there is a minimal pair. And at times it is the contextual phenomenon, whereby the melody of a given word is different depending on an adjacent word.

[[Illustration from Ambel (Austronesian [SHWNG], Indonesia). Data from Laura Arnold.]]

Exercise

[[Exercise on distinguishing between tone patterns based on their realisation, and on tone sandhi.]]
How did you go about it?

If we can lump forms together reliably, we are off to a good start

At the initial stage, it does not matter whether we get the surface-phonological description of the tone patterns right. What is key is that we can lump forms together with reliability. Bird & Lee (2014) offer a software tool that can help with this, although the crucial challenge remains with the researcher.

So having a solid grip on the tone categories does not require knowing the phonetic realisation.

At this point, the underlying representation – if different from surface phonology – is still beyond reach. That is ok. But it is important to realise the tone patterns at issue now are surface-phonological (at best). It is only from a comparison across a range of contexts that we can infer re. the underlying forms.

Of course, the underlying forms may be the same as the surface-phonological patterns, as in Shilluk.

Systematicity across speakers is key, and offers important corroboration.

Another advantage of recording sound data: better listening

Just hearing the patterns when the speaker utters them a few times is not enough, they won’t make enough impact. By recording the realisations and working with them, our practice becomes so much better. We can then hear an utterance more often, compare realisations across speakers, compare realisations across contexts, and mimick the tonal shape. All of these techniques are very useful in developing a grip on a tone system.
2. What about quantitative analysis in studying of tone?

The process of discovering a tone system hinges on qualitative analysis: auditory pitch impressions, collected through a controlled procedure as outlined in Pike (1948) and in Bird & Hyman (2014). Quantitative analysis – acoustic measurements accompanied with statistical analysis – will not offer the answers. Hyman (2014:553) advises against the use of quantitative evidence (f0) early on: “The problem with looking at the f0 properties of tones too early is the tendency to interpret them literally.”

Why are traditional methods better in the initial stages? Because we are looking for contrast, both in the sense of minimal sets and also in the sense of patterns that sound saliently different.

What would be the alternative?

F0 traces are useful – if we can interpret them appropriately

[[Illustration from Hyman 2014]]

“If we look at Figure 1, it clearly starts out fairly low and ends rather high. However, the second syllable seems to have a continuous rise in it, suggesting maybe a L-LH transcription. Listening to it, however, it was clear that it was perceptually [L-H]. To have been L-LH, the transition from L to H would have had to take place later in the syllable. […] we know that pitch targets are reached late on a tone-bearing unit […]” [Hyman 2014:553]

Fundamental frequency (F0) traces can reveal the phonetic characteristics of tone patterns we have previously hypothesized. For example, F0 traces could reveal that “this pattern that I have distinguished from the low pattern: it is a shallow fall / it is mid level / etc.”

We hear pitch patterns as low, high, rising or falling. We can only check these impressions with the acoustic evidence if we know what an impressionistic fall is supposed to look like in terms of its F0 trace. In what follows I relate auditory impressions of pitch to f0. The goal here is to enable you to use f0 traces to support the discovery and description of tone.¹

¹ I leave aside here the question of how to handle perturbations (e.g. the influence of obstruent voicing on f0 height, or spikes at the boundaries between sonorants and vowels) and tracking errors (e.g. due to vocal fry). There are straightforward best-practice procedures to deal with these. Feel free to ask.
Level tones in acoustic and auditory perspectives

Think of the organisation of consonants in the IPA table. What would be the tonal parallel? In terms of tone, there are primarily two dimensions: vertical (tone height) vs. horizontal (tonal alignment or timing).

**Vertical dimension (tone height):**

Up to five levels of tone height – e.g. Benchnon (Wedekind 1983) and Black Miao (Kuang 2013). In the latter non-modal phonation is also involved. Many tone languages only have two levels. A third level is not uncommon; richer systems are rare.

It is already clear from Hyman’s example that ‘level tones’ and ‘contour tones’ are often used as perception-based notions. What does ‘level tone’ mean in acoustic terms?

It is the f0 pattern during the vowel, the part of the syllable where intensity is greatest, that determines the pitch percept of a syllable.

**Horizontal dimension (tonal alignment):**

Timing or alignment of targets. Alignment is invariably part of the realisation of tone patterns, even if the configuration is just Low vs. High. How? Because it is essential to the realisation of level tones in context.

The percept of a level pitch corresponds to a single target, and this target is reached late in the syllable. F0 changes during the onset consonant do not impact much on perception.

[[Illustration: sound examples and f0 traces from Shilluk; and schematic representations.]]

Xu (1999) shows how the Mandarin tones are most stable at the end of the syllable.

**Key insights from speech production, speech perception, and typology**

*The speed of f0 changes*

F0 changes take time to realise. The most accurate account of how f0 changes are implemented is the landmark study by Xu & Sun (2002). This is an experimental study on
the maximum speed of f0 change, based on data from speakers of English and of Mandarin. Key finding: it takes about 100 ms to implement a change in f0 of 4 semitones.

[[Illustration from Xu & Sun 2002]]

In practice: in the realisation of the Low tone in the Shilluk example, the f0 change goes from 165 Hz (8.67 ST) at 0.526s to 118 Hz (2.82 ST) at 0.658s. So there has been a 5.85 ST drop over 132 ms. Note how the f0 change often extends over stretches of speech that are greater than a speech segment: The onset consonant [l] is only 83 ms long; 132 ms corresponds to a vowel with substantial duration.

The glissando threshold

Another consideration: the glissando threshold (Rossi 1978; Greenberg & Zee 1978; ‘t Hart, Collier & Cohen 1990): the smaller the time domain is small, the greater an f0 change needs to be in order to be perceived as a pitch change.

Formula – for a given duration (D), the glissando threshold, in semitones (ST) per second

\[ \text{ST} \text{ per second} = \frac{0.16}{D^2} \]

Example – a change from 150 to 135 Hz (1.83 ST). Perceived as a pitch contour over 100 ms, but as a level pitch over 50 ms.

Typology

The evidence from speech production and perception discussed above suggests that f0 changes take time to implement. They may be difficult to perceive as pitch contours if there is little time.

The typological evidence is in line with this. Zhang’s (2001) PhD includes a typological survey on the distribution of contour tones. He finds that: “Of the 187 languages, 22 had no restriction on the distribution of contour tones.” [Zhang 2001:166]

The distribution of contour tone is restricted to syllables that have one of the following characteristics: stress; a long vowel; greater weight; that are utterance-final. The common denominator is sonorous duration, i.e., duration that conducts f0.
Contour tones

In a contour tone, a change between tone targets is implemented during the syllable rhyme. The phonetic constraints restrict the possibility for languages to have contour tones (cf. Zhang 2001).

In perceptual terms, that is saliently different from a level tone: in that case, any f0 change is implemented at the beginning of the target syllable or at the beginning of the next one.

The difference between a level tone and a contour is not necessarily about whether there is an f0 change. It is often about alignment, i.e., where the f0 change takes place.

How is the f0 change lined up in contour tones? Is contrast possible?

Many languages present contour tones where the initial target is lined-up right at the beginning of the vowel. Also very common are contour tones where the f0 changes takes place well into the vowel. Is contrast possible here?

- Recent evidence: yes (Remijsen 2013 on Dinka, Dicanio, Amith & García 2014 on Mixtec, Remijsen & Ayoker 2014 on Shilluk)

When investigating contour tones, it is worthwhile to pay attention to where the f0 change sets in: near the beginning of the rhyme, or well into the vowel? Working hypothesis based on House (1990): there is a threshold at 30 milliseconds into the vowel. Whether the f0 change sets in before or after this threshold determines whether it is perceived as an early-aligned contour tone or instead as a late-aligned one.²

² This hypothesis of David House can be interpreted as a quantal threshold in the sense of Stevens (1989) and Stevens & Keyser (2011).
Putting the descriptive analysis to the test

The exploration of a tone contrast using traditional field methods, with special attention to the need to control for potential confounds, and supported by recordings and acoustic representations, yields a working hypothesis. There are various ways to put it to the test through quantitative analysis. Beyond the intrinsic value of testing the hypothesis, there is, if the analysis is confirmed, the advantage of making the analysis more persuasive. This matters, because the risk that an analysis based solely on auditory impressions is wrong is factored in the field. Two kinds of quantitative analyses can be distinguished:

Production studies

A quantitative acoustic analysis. Key characteristics: (a) the sample of speakers is representative of the population (>7 people is typically safe); (b) the inventory of tone categories in different phonotactic forms; (c) the factors governing contextual tone processes are manipulated. In this way, we can corroborate both the hypothesized inventory and the hypothesized phonological processes.

Perception tests

A basic test is to investigate whether speakers accurately recognise members of minimal sets distinguished primarily by tone. A more advanced approach is analysis through resynthesis, that is, a perception test using (re)synthesized stimuli. If we know the acoustic ‘recipe’ of a word-level specification for tone, we can start out with one member of a minimal set, modify the f0 shape to that of another member of the same set, and native speakers will perceive it as the latter.

References


Kuang, Jianjing (2013). The tonal space of contrastive five level tones. Phonetica 70, 1-23.


Remijsen, Bert (2013). Tonal alignment is contrastive in falling contours in Dinka. Language 89(2), 297-327.


