LANGUAGE DOCUMENTATION & CONSERVATION

Vol. 8 (2014), pp. 453–461
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/24613

Computational support for early elicitation and classification of tone

Steven Bird[1],[2] and Haejoong Lee[2]
[1]*Department of Computing and Information Systems, University of Melbourne*
and [2]*Linguistic Data Consortium, University of Pennsylvania*

Investigating a tone language involves careful transcription of tone on words and phrases. This is challenging when the phonological categories – the tones or melodies – have not been identified. Effects such as coarticulation, sandhi, and phrase-level prosody appear as obstacles to early elicitation and classification of tone. This article presents open source software that can assist with solving this problem. Users listen to words and phrases of interest, before grouping them into clusters having the same tonal properties. In this manner, it is possible to quickly annotate words of interest in extended recordings, and compare items that may be widely separated in the source audio to obtain consistent labelling. Users have reported that it is possible to train one's ear to pick up on the linguistically salient distinctions. The approach is illustrated with data from Eastern Chatino (Mexico) and Alekano (Papua New Guinea).

1. INTRODUCTION. During early elicitation, transcription practice evolves as we tune into the linguistically salient contrasts. For segmental distinctions, it is usually straightforward to begin with narrow phonetic transcriptions and gradually leave out details once they are found to be non-contrastive. For instance, after noting that voiceless obstruents are aspirated in syllable onset position, we may decide to stop marking aspiration. Over time, such conventions make it possible for transcription to proceed more quickly, and for the results to be more readable. Yet all the time, we try to remain open to detecting new contrasts (cf Hyman 2001).

The situation is often more acute for tone. To begin with, the IPA notation for tone is cumbersome, and it is also arbitrary with its five levels and the corresponding contours. In the experience of many, it is more effective to draw stylized contours, e.g [– _ /]. The use of elicitation frames may effect the target word in unpredictable ways, and we have to sort out the various contributions of phrase-level prosody (e.g. phrase boundary tones), local phonological alternations, and phonetic interpretation (e.g. tonal coarticulation). Eyeballing F_0 traces sometimes helps, but these are often misleading.

In short, we are trying to identify discrete surface tonal categories without knowing the underlying tonal inventory or what gave the tones their observable phonetic realization. In the early stages of description, we may perceive a pitch difference between a pair of syllables or words, but we may not know whether this difference indicates an underlying contrast. Later, perhaps after a week or a month, our language acquisition device becomes engaged and we start to "hear" the tone, to tune into the salient distinctions. Ideally, we would reach this stage more quickly and reliably so that we can produce useful transcriptions in a shorter amount of time. Field trips often have a short duration, and so speeding up this ear-training process may have a significant impact on the quality and quantity of the transcriptions that can be made in the field, and this may in turn help identify gaps where more data can be collected while there is still time. Our work is intended to occupy this

E-ISSN 1934-5275

niche of early elicitation. Note that there is no claim that this process uncovers underlying tones; the focus is on grouping forms by their surface shape, and on discovering which superficial details must be heeded and which can be ignored.

This paper presents a free, open source software tool called Toney that is intended to support the early elicitation and classification of tone language data. Toney displays forms on a canvas, and the user can listen to the forms and group them into clusters. By reviewing the clustered items, it is easy to learn to hear the tonal categories and identify misclassified items. By using this software, the user can quickly learn the linguistically salient tonal categories, and annotate extended audio recordings.

This paper is organised as follows. Section 2 presents an extended example of an early elicitation problem in Alekano. Section 3 shows how the tool is used to classify words in isolation. In section 4, this is broadened to include sentence frames, multiple speakers, and further categorizations that may be useful. The paper closes with a discussion and conclusions.

2. BACKGROUND: EARLY ELICITATION OF TONE. In order to motivate our approach, we begin with an example of early elicitation in Alekano (ISO gah), a language spoken by about 20,000 people in the Eastern Highlands Province of Papua New Guinea. Consider the following sentence, transcribed using the phonemic orthography where symbols have their IPA values.

(1) gènēzá àní'gùvè 'see tongue'

The target word *gènēzá* 'tongue' appears to have a rising sequence which we have transcribed as low-mid-high. However, the position of the mid between L and H is suspicious: perhaps it is really a low tone that has been raised in the context of H. If so, we may may be able to drop the mid tone category, and write instead *genezá aní'guve* (leaving low tone unmarked), and posit a rule of phonetic interpretation in which a low tone is raised in the L_H environment. Alternatively, the middle syllable of *gènēzá* may be toneless, and we would need to posit a rule of tonal interpolation.

After further elicitation we build up a picture of the inventory of tonal melodies on words with a fixed syllable shape, in this case CVCVCV words:

(2)	a. LLH	genezá aní'guve	'see tongue'
	b. LHH	golání aní'guve	'see blood'
	c. LHL	gosíha aní'guve	'see snake'
	d. HLH	lágahá aní'guve	'see fish'

Further possibilities for these words show up when we add a definiteness marker.

(3)	a. LLL	geneza-má aní'guve	'see the tongue'
	b. LHH	golání-má aní'guve	'see the blood'
	c. LHL	gosíha-má aní'guve	'see the snake'
	d. HLL	lágaha-má aní'guve	'see the fish'

Here, the final syllable of the target word in (3a) and (3d) becomes L, and we could posit a rule H->L/L_#H. The rule which raises this L seems to be variable, and we get both level and rising variants, e.g. ,

There are problems with this approach. It establishes a sequence of hypotheses which purport to account for a selection of the data. However, we would like more than this. First, we want to be faithful to the data, confident that we are not deluding ourselves by transcribing the materials opportunistically in order to support our early hypotheses. Second, we want to be accountable, retaining the link between a transcribed form, the audio recording on which it is based, and the full set of forms that are transcribed with the same sequence of tone labels. Third, we would like to tune our ears to linguistically salient aspects of pitch (the usual perceptual correlate of tone) so that we can transcribe more quickly and reliably over time. These goals are challenging when we have not identified the tones and don't know how to attribute putative contrasts to phonological categories or phonetic effects. The following sections introduce the software and show how it addresses these problems.

3. CLASSIFYING WORDS IN ISOLATION. Consider the case of an early elicitation session in which a set of words and glosses have been transcribed, reviewed, and recorded. If multiple speakers are involved, the wordlist would ideally be recorded separately with each speaker, to minimize the risk that one will copy the intonation of the other. Armed with these recordings, we proceed by manually labelling the words using acoustic analysis software (see Figure 1).

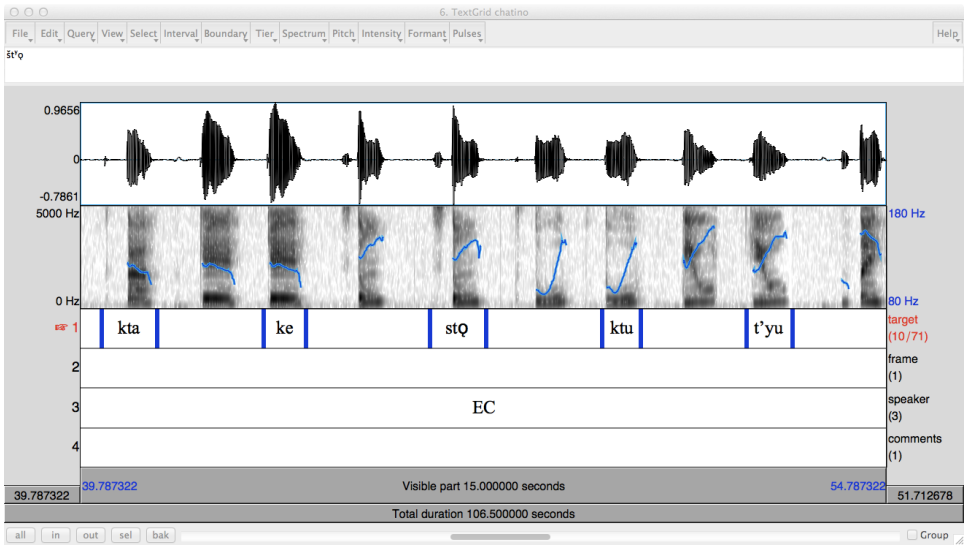

FIGURE 1: Praat annotation of selected lexemes, using the "target" tier. In this case, only one instance for each lexeme has been labelled.

The annotation is required to have at least the following three tiers: target, frame, and speaker. In this instance, the words were produced in isolation, and so no frame is specified.

The speaker is the same for the entire file, so there is a single annotation with the speaker's initials (here, EC). Once the recording has been annotated in this way, we open the Praat file using Toney and see the words scattered on a canvas, as shown in Figure 2.

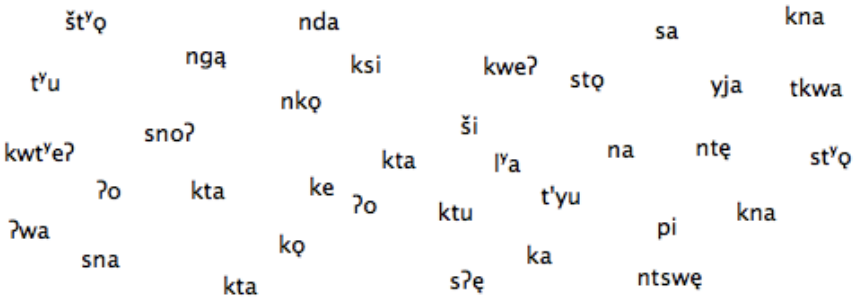

FIGURE 2: Toney's initial display of the Eastern Chatino nouns

The user can click on the words to hear them, and drag similar-sounding words into groups on the canvas. In Figure 3, the forms have been arranged into three groups, roughly corresponding to rising, level, and falling melodies. The choice of groups and membership is arbitrary and depends on the user's perception of tone melodies, optionally with guidance from a native speaker. (Note that the three instances of *kta* have different melodies. The two instances of *ʔo* are also slightly different, though it is fine to put them together for now.)

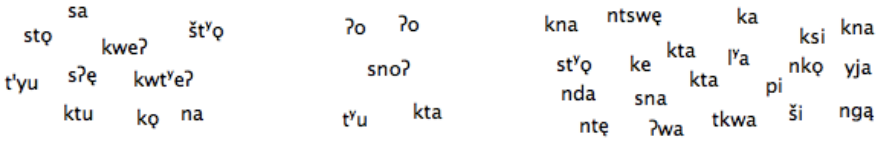

FIGURE 3: Manually arranging words into clusters on the canvas

Once a set of words with the same tone melody has been identified, it is moved into a cluster in the top half of the display (Figure 4). Each cluster has an arbitrary user-assigned label. There is a row of four buttons at the bottom of each cluster. The first button plays all of the forms of the cluster in sequence. Usually, any misclassified forms are obvious at this point, and they can be moved to another cluster or back to the canvas. The second button plays all the forms with their elicitation frames, permitting them to be heard in context. The remaining buttons are for stopping playback and for deleting the cluster (respectively). Users can create an arbitrary number of clusters.

FIGURE 4: Establishing tone clusters

Each cluster has two tabs, one for the individual item labels (Figure 4) and one for the F_0 contours (Figure 5). The F_0 contours for all forms in the cluster are overlaid. Here a falling-tone word has been added to the mid cluster, and it stands out in the F_0 display for the mid tone. We can click on the contour to hear it, and switch back to the "Label" tab to see which word is highlighted and remove that word from the cluster.

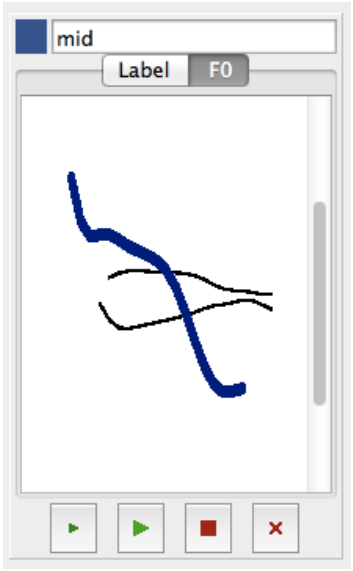

FIGURE 5: Display of F_0 with smoothing and interpolation

At the end of a session, the labels are saved back to the original Praat file, and any forms that have been clustered will appear with a cluster label. For instance, *kweʔ*, from the first cluster will now appear in the Praat file as *kweʔ:high-rise*. (NB. it is useful to adopt compact labels for maximum readability in Praat, e.g. *HR* instead of *high-rise*).

4. ADDING INFORMATION ABOUT SENTENCE FRAMES AND SPEAKERS. There are many individual and contextual influences on F_0 (Connell and Ladd 1990, Snider 2014, Yu 2014). Most obviously, F_0 contours are scaled relative to a speaker's pitch range. The contour for a word is sensitive to its segmental, phrasal, and prosodic context. Even when uttered in isolation, a word carries phrase- and utterance-level prosodic information such as a final fall to the bottom of the speaker's pitch range. It is usual to elicit tone data from multiple speakers, and to vary the target syllable or word within sentence frames so that we can identify linguistically salient aspects of the tonal melodies (Pike 1948). It is best to use a variety of frames, controlling for phonological and morphosyntactic context. An example of the frame annotation is shown in Figure 6. The frame is labelled F2 and corresponds to the examples we saw earlier in (3).

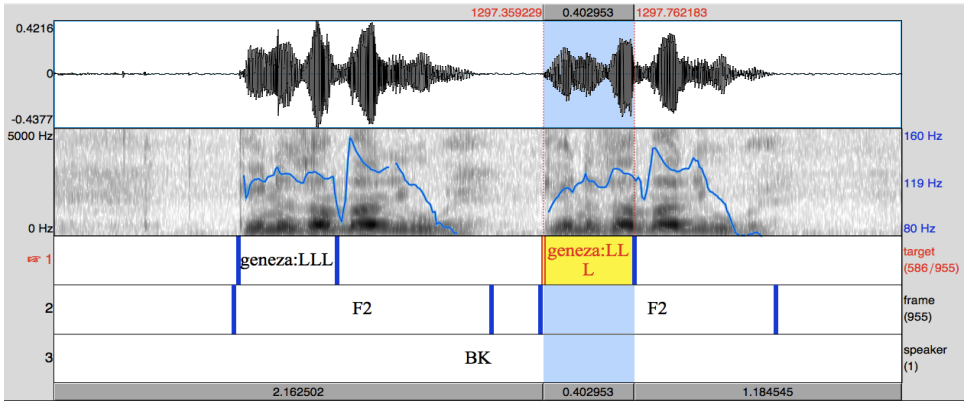

FIGURE 6: Alekano target words showing transcriptions (LLL) and a sentence frame (F2)

Toney supports playback of target words (by clicking) and whole frames (by right-clicking). Figure 7 shows a partial screenshot once all the items have been classified. We can listen to all forms in the LLH column to verify that the second L tone is raised, and confirm that these are distinct from the LHH column. We can observe that *gaha-F3* appears in both LH and LL columns, something which will need to be verified, and considered in light of the recordings of the other two speakers.

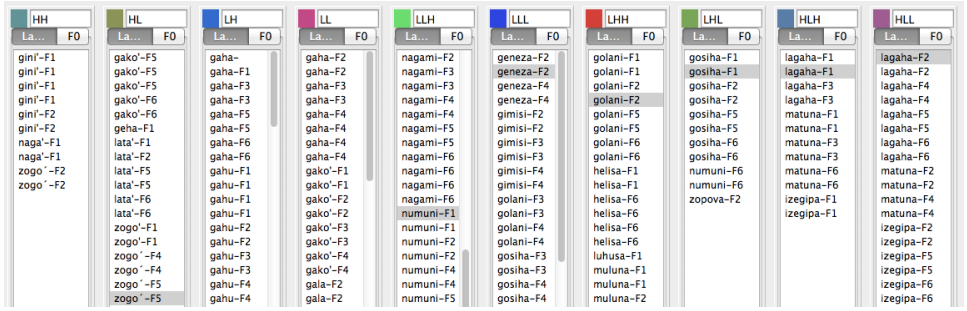

FIGURE 7: Classification of Alekano words from six sentence frames, controlling for the tone on preceding and following syllables, with frame identifiers displayed

Just as we classified the forms according to sentence frames, we can classify them according to speakers. As we have seen, the third Praat tier contains a speaker identifier. Such labels should span every frame that was produced by a given speaker. In our Alekano example, the speaker identifier spans the entire file, and we have three separate files, one per speaker. These can all be loaded at once, and the speaker id (e.g. *BK*) can be displayed alongside each form (e.g. As *gosiha-BK* or *gosiha-F4-BK*). Any systematic difference between speakers should manifest itself in the pattern of speaker identifiers and column labels.

To the system, the frame and speaker labels are just arbitrary categories that can be used for dividing up the data. We can add more dimensions to our labels. For example, we could cross-classify all forms for syllable weight, vowel height, and onset laryngealization. The Praat label would consist of colon-separated fields, e.g. *gahu:LH:light:low:no* (and so tone is in position 1, syllable weight is in position 2, and so forth). These extra fields can be created inside Toney by selecting a new "value position", and then populating clusters and labelling them.

The same method can be used to break a melody into its components. Thus, instead of classifying bisyllabic forms into one of *LL, LH, HL, HH,* we could establish two orthogonal categories, one for the first syllable and one for the second. Now, the user's classification task would consist of making two independent judgements, one per syllable. For best results, the data should be relabelled in Praat, using a new segmentation, one for each syllable. Alternatively, we can perform tonal classification based on the initial or final part of the bisyllable.

5. DISCUSSION AND CONCLUSION. In their presentation at the Berkeley Tone Workshop in February 2011, Woodbury and Cruz demonstrated an approach to elicitation and transcription based on spreadsheets (Cruz and Woodbury 2014). Participants were given a handout with one segmentally transcribed lexeme per row. Each row was numbered. Participants listened while the Chatino scholar (Cruz) produced each form and independently wrote down their tone transcriptions. After all forms were transcribed, we spent the bulk of the time asking Cruz to produce items in succession, by calling out pairs of row numbers, usually non-adjacent in the spreadsheet. After each such test, individuals would decide whether the pair should be grouped together, as having the "same" F_0 contour. A key insight to emerge from this activity was that early tone transcription is a clustering task, where we only need to decide if two instances are the same or different. This is the insight that underpins the work that has been reported here.

As already noted, early elicitation of tone data is often difficult, thanks to a large number of influences on the F_0 contour that work to obscure the underlying tonal classes. We have developed software that is designed to fit into this niche of early elicitation, and help a linguist to identify the linguistically salient contrasts and annotate them consistently in extended recordings. Key features of this approach are as follows:

> **Collocation**: Items that were widely separated in a field recording can be brought together, making it easy to check that they are transcribed appropriately. All items with the same transcription can be cross-checked, and any items that do not follow the pattern stand out and can be corrected right away.

Ear training: The software makes it easy for users to listen through lists of items having the same or similar tone melody. Similarly, a non-linguist native speaker can be alerted to the tonal contrasts of his/her language and will hopefully learn to alert the linguist to new contrasts, or to non-contrasts. Unlike a native speaker, the software does not tire of repeating the same set of forms over and over again.

Progressive elicitation: files from a series of elicitation sessions can be loaded at the same time, facilitating the growth of the collection over time without any need to re-record items. When an item is reclassified, its label is saved back to the file it came from. When a class label is modified, all items in that class are relabelled in the corresponding files.

Primary documentation: The source audio could contain primary documentation (such as a recorded narrative) instead of controlled elicitation. The only requirement is for individual forms to be annotated, with an optional context window (the frame).

Words and frames: We can listen to words with or without the surrounding sentence frame.

Audio annotation: The tone labels are stored as annotations of one or more original recordings, rather than a separate collection of extracted audio clips which would lose the connection with the source recording. If the segmental transcription of a word needs to be changed, this only needs to be done once, at the place where the word is located in the file. Similarly, if the tonal transcription of the word is changed in the Praat file, the word will be assigned to this new tonal category inside Toney. The Praat file can have any number of extra tiers, corresponding to other kinds of annotations required by the user, and these are left untouched when Toney saves tone labels back to the Praat file.

ACKNOWLEDGEMENTS

We are grateful to Emiliana Cruz and Tony Woodbury for providing the Chatino data, and to two anonymous reviewers for detailed feedback on a previous draft. The first prototype was developed by Anaïs Poruk at the University of Melbourne. The software and data samples are available for download from http://lp20.org/toney/. The open source project is hosted on GitHub, https://langtech.github.io/toney. This research has been supported by NSF award 0951651 *Prosodic Systems in New Guinea: Integrating computational and typological approaches to linguistic analysis*.

References

Boersma, Paul and Weenink, David (2014). *Praat: Doing phonetics by computer* [Computer program]. Version 5.3.63, retrieved 15 April 2014 from http://www.praat.org/

Connell, Bruce and D. Robert Ladd (1990). Aspects of pitch realisation in Yoruba. *Phonology* 7, 1–29.

Cruz, Emiliana and Tony Woodbury (2014). Finding a way into a family of tone languages: The story and methods of the Chatino Language Documentation Project. This volume.

Hyman, Larry M (2001). Fieldwork as a state of mind. In Paul Newman and Martha Ratliff (eds). *Linguistic Fieldwork*. Cambridge University Press.

Pike, Kenneth (1948). *Tone Languages*. University of Michigan Press.

Snider, Keith (2014). On establishing underlying tonal contrast. This volume.

Yu, Kristine (2014). The experimental state of mind in elicitation: illustrations from tonal fieldwork. *Language Documentation and Conservation* 8: 738–777

Steven Bird
stevenbird1@gmail.com

Haejoong Lee
haejoong@ldc.upenn.edu