MODELING IU MIEN TONE WITH EIGENPITCH REPRESENTATIONS

Ryan Shosted, Marissa Barlaz, Di Wu

University of Illinois at Urbana-Champaign
rshosted, mbarlaz, wudi6@illinois.edu

ABSTRACT

To achieve adequate description of an understudied tone language, we argue for unsupervised computational modeling of lexical tone, even at the earliest stages of documentation. We apply a technique using eigenvalues and discriminant analysis to differentiate lexical tones on monosyllabic items in Iu Mien. The resulting 'eigenpitch' representations can be used to evaluate the differences and similarities between tones, to sharpen the impressions of field linguists, and to contribute to a richer understanding of phonemic tonal contrasts.

Keywords: eigenvalues and eigenfunctions, pitch modeling, lexical tone, Iu Mien

1. INTRODUCTION

1.1. Describing tone languages

Field linguists working on underdescribed languages are confronted with the problem of classifying phonological patterns inductively. As with any phonetic or phonological characteristic of language, lexical tones are produced by multiple speakers, with concomitant variation related to speaker identity, gender, age, body size, mood, and perhaps many other factors [1]. For adequate documentation of a language known to exhibit lexical tone, any analysis of phonetic tone productions should result in two basic outcomes: (1) A definitive statement on the number of phonemic tonal contrasts; and (2) the reliable assignment of lexical tones to individual words in the lexicon of the language. For languages with a non-existent, little-used, or inadequate writing system, much of this analysis traditionally proceeds on the basis of a researcher's auditory impressions and careful phonetic transcriptions (e.g., [2], among many others). However, [3] has argued that "simple observation and introspection" are inconsistent with the goals of a field that is increasingly focused on "appropriate documentation and numbers," concluding that, while phonetic data acquisition is now a more complex enterprise, "[t]he reward of greater complexity is greater certainty."

Higher-level mathematical modeling of tone can be used to better understand lexical tone systems in underdescribed languages where there is an emphasis on acoustic data collection and the number of phonological tones may be unknown or controversial. Advanced modelling has already been applied to lexical tone data. For example, researchers have used polynomial decomposition to model f0 curves for some time [4, 5, 6, 7]. This technique makes it possible to compare the shape of tonal contours in terms of functional coefficients. These coefficients can be studied in order to determine, e.g., whether (and to what extent) a tonal contour is convex or concave.

In this paper, we demonstrate that high-dimensional, eigenvalue-based, unsupervised classification of f0 curves can leverage much of the richness in a recorded database and produce quantitative, falsifiable results that are comparable to existing descriptions of Iu Mien tone. Such an approach may benefit the documentation of underdescribed and endangered languages that exhibit lexical tone.

1.2. The case of Iu Mien

This study deals with Iu Mien (ium: ISO 639-3), an underdescribed Hmong Mien language spoken in China, Laos, Thailand, and Vietnam. Iu Mien is known to have six lexical tones [8]: **low level** (tone-letter 'c'; 11; ˩); **mid falling** ('h'; 31; ˧˩); **high mid level** ('unmarked' or '(u)'; 44; ˦)[1]; **high rise falling** ('v'; 453; ˦˥˧); **low rising** ('x'; 24; ˨˦); **low rising falling** ('z'; 231; ˨˧˩). When high rise falling (v-tone) syllables have a stop coda, a seventh tone occurs: **high level** ('v*'; 55; ˥). Our speaker reported explicit knowledge of only six tones in her Thai dialect, but she did not seem fully aware of which tonal categories were affected. (After our experiment, we learned through further elicitation that the tones affected were x and z, a finding supported by our results; see Figure 5 and discussion in Section 4.)

1.3. Purpose

We approach the tone system of Iu Mien inductively, using computational tools to identify tone categories in the language, based on acoustic recordings. Be-

Table 1: Lexical tones in the corpus, including numeric ('T#') and IPA/Chao transcriptions. Median representation $\approx$ 13% ($n = 19$; $n_{exp} = 21$). $\chi^2(6) = 18.67$, $p < 0.01$. This suggests under-representation of h-tone (mid falling, 31) and over-representation of (u)-tone (high mid level, 44). h. = high; l. = low; m. = mid.

Name	Description	T#	IPA	Pctg.	n
c	l.level	11	˩	15%	22
h	m.falling	31	˧˩	6%	9
(u)	h.m.level	44	˦	23%	34
v	h.rise falling	453	˦˥˧	19%	28
x	l.rising	24	˨˦	13%	19
z	l.rising falling	231	˨˧˩	12%	18
v*	h.level	55	˥	12%	17

cause Iu Mien's tonal system is relatively well-documented, our aim is to use Iu Mien as a test case to validate our method. We hope the method may help field linguists characterize complex phonological systems using computational tools, in order to minimize reliance on impressionistic evidence and provide falsifiable quantitative results supporting such evidence. What follows makes use of an admittedly small dataset (245 monosyllabic items), uncontrolled for tonal categories or contextual effects of consonant. The approach could probably be improved with more careful controls of the recorded materials. However, our work is typical of what might be done in a field setting, where all possible contrasts and their distribution throughout the lexicon may not be known *a priori*.

2. METHODS

The database consisted of 245 monosyllabic lexical items produced by one female Iu Mien speaker. Items were of the structure V, CVC, and CVV. We annotated the sonorous portion of each item and extracted f0 at ten time-normalized points using ProsodyPro 5.3.2 [9] (f0 range 30–400 Hz; f0 s.r. 100 Hz; perturbation length 0 s; final offset -0.03 s; smoothing window 0.07 s). Table 1 gives the number of lexical tones represented in the corpus.

2.1. Modeling procedures

The authors, all of whom were unfamiliar with Iu Mien prior to the experiment, listened to all items and classified the lexical tones without supervision, guidance, or collaboration as may sometimes occur in a field setting. (The third author is a native speaker of Mandarin Chinese, an unrelated tone language.) Later, the canonical tone categories assigned to each word by Purnell [8] were noted independently. We then performed principal component analysis (PCA) of the 10-point normalized pitch contours [10]. The number of PCs was optimized by calculating the proportion of successful reassignment, corrected for the number of retained PCs [11] . We performed discriminant analysis on the resulting PCs using two techniques.

2.1.1. Discriminant analysis 1

We randomly divided the corpus into training (80% of the items) and testing (20%). k-means clustering was used to identify and classify the number of tones, k, present in the corpus. Clusters can be compared post-hoc with the categories of tones assigned by a researcher (including Purnell, the dictionary author [8]). This method allowed us to explore outcomes when the total number of lexical tones is posited, but the assignment of individual items to categories is unknown. In this paper, we present $k = 7$, the number of tones posited in [8].

2.1.2. Discriminant analysis 2

We randomly divided the corpus into training (60%), testing (20%), and cross-validation (20%) sets. Purnell's tone categories, assigned to each item, were used as priors [8]. However, these priors could be easily exchanged for the tonal categories assigned by other investigators, presenting useful opportunities for evaluation of consistency across and within researchers. The priors were used as the basis of a (training-based) model, which was then applied to supplementary observations (the testing and cross-validation sets).

2.2. Interpretation of model results

Discrepancies between the model's predictions and the tonal categories assigned by a researcher may help to clarify the speaker's idiosyncratic production of some tone categories and/or clear up potential misconceptions on the part of a researcher. Instability of group membership may lead to the conclusion that some tonal categories have merged (for this speaker) or that the researcher has misidentified some crucial data in the corpus. Visualizations of tonal patterns and hierarchical clustering of the same are helpful for understanding variation in production as well as the predictive capacity of the various researcher-generated models under comparison.

3. RESULTS

3.1. Discriminant analysis 1: *k*-means clustering

Using the training data and $k = 7$ to arrive at seven clusters, the optimized number of PCs was two (94.58% variance explained). Figure 1 shows the seven resulting f0 curves. Figure 2 shows a scatter plot of the seven learned categories and their membership from the seven dictionary categories. Table 2 presents the seven resulting, machine-learned tone categories, along with Purnell's tonal classification of the items belonging to each category [8].

Figure 1: Machine-learned f0 curves (1–7) for a $k = 7$ discriminant model.

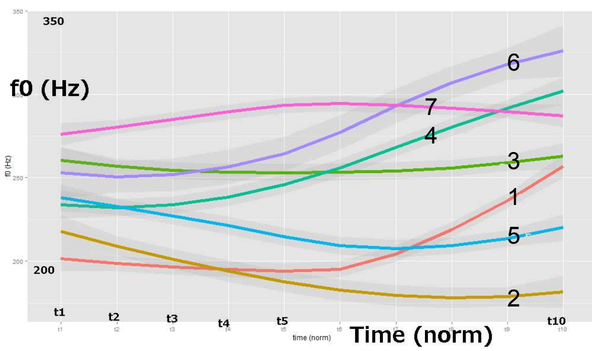
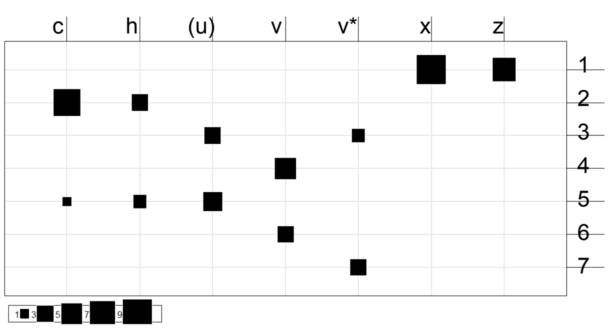


Figure 2: Machine-learned tones (1–7) and their correspondence to Purnell's dictionary tones. The largest black square in the key represents $n = 9$; for the smallest black square, $n = 1$.



As shown in Table 2, three groups emerged from the seven-tone model: rising, falling, and level tones. The rising group accurately corresponds to all dictionary rising tones, but failed to capture the fall at the end of the z-tone (low rising *falling*) and v-tone (high rise *falling*) tones. The falling group accurately corresponds to the h-tone (mid falling) but inaccurately corresponds to the c-tone (low level) and (u)-tone (high mid level). The level group accurately corresponds to both level tones ((u)-tone and v*-tone).

Table 2: Results of discriminant analysis 1 with researchers' IPA/Chao transcriptions of machine-learned f0 curves (*k*-tones). The traditional tone letters discovered for each group, post-hoc, are also presented (cf. Figure 2).

k-tone	Transcription		Members (post hoc)	
1	l.rising	˩˧	x, z	˧˥, ˧˩˥
4	m.rising	˧˥	v	˦˥˧
6	m.rising	˧˥	v	˦˥˧
2	l.m.falling	˨˩	c, h	˩˩, ˧˩
5	m.falling	˧˨	c, h, (u)	˩˩, ˧˩, ˦˦
3	m.level	˧˧	(u), v*	˦˦, ˥˥
7	h.m.level	˦˦	v*	˥˥

3.2. Discriminant analysis 2: Clustering with dictionary tones as priors

For this procedure, the optimized number of PCs was seven, accounting for 99.51% of the variance in the training data. The first two PCs accounted for 94.6% of the variance (cumulative). PC1 appears strongly correlated with a point about 70% through the f0 curve and PC2 appears strongly correlated with the endpoint of the f0 curve. The resulting model, based on machine learning of the f0 contour of Purnell's tones [8], was used to correctly classify 61.23% of supplementary (testing) observations. In addition, leave-one-out cross-validation of the remaining 20% of the corpus generated tones broadly similar in shape to those produced during the testing phase. The authors each independently transcribed the f0 curves generated by training, testing, and cross-validation, using IPA/Chao conventions (Table 3). The f0 curves based on the training set are shown in Figure 3; the testing-based curves, in Figure 4. In addition, hierarchical clustering analysis was performed on the testing model, with results presented in Figure 5.

4. DISCUSSION

The results of *k*-means clustering (Section 3.1) may be inadequate to the task of describing Iu Mien's tones. Perhaps this is because the speaker did not differentiate all tones (as self-reported). Also, the model may be inadequate at capturing the distinctions (poor data representation, suboptimal number of PCs, etc.). The results of Section 3.2 were more encouraging, however. The tone cluster dendrogram presented in Figure 5 reveals great similarity be-

Figure 3: f0 curves based on training data (60% of the recorded corpus) with Purnell's dictionary tones as priors.

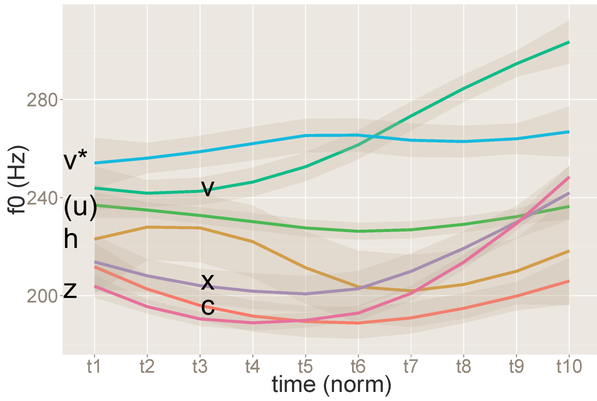


Figure 4: Automatic classification of f0 curves based on the training model and further testing data (20% of the recorded corpus).

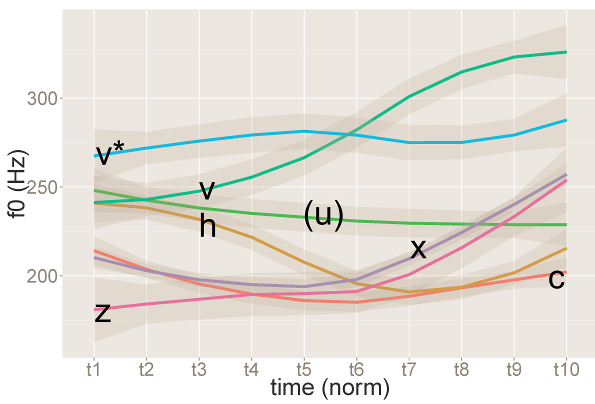
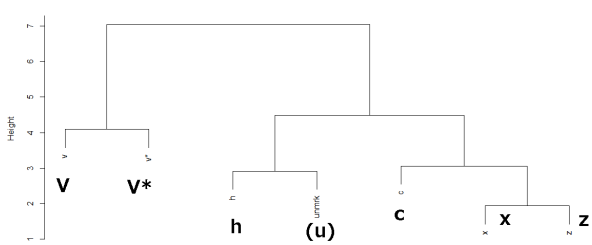


Figure 5: Hierarchical clustering of Iu Mien tones based on the testing model (see Figure 4).



tween x- and z-tones (phonologically low rising ˨˦ and low rising falling ˨˦˨). This suggests that tones x and z have merged for this speaker, a fact we were able to verify independently with further elicitation.

Table 3 points out one of the special considerations of combining computational modeling and traditional fieldwork. While the algorithms we use here are able to produce stable groupings of f0 contours, there is still variation in the way they may be transcribed by a linguist. In only one instance did the transcription of a machine-learned f0 contour conform to the tone description given by [8] (c-tone). We are encouraged, however, by the similarity of the machine-learned tones and their dictionary descriptions. It seems that variation across items is accounted for in the relatively abstract eigenpitch representations.

Table 3: Summary of researcher-transcribed f0 from training, testing, and cross-validation phases of discriminant analysis 2.

	Dictionary Description		Transcriptions Train	Test	Xval
c	l.level	˩	˨˩˨	˩ ˨˩˨ ˨˩	˨˩
h	m.falling	˧˩	˨˩˨ ˧˩˨ ˧˨	˧˩˨	˨˩˨
u	h.m.level	˦	˧	˧	˧
v	h.rise falling	˦˥˦	˧˦	˧˦	˧˦
x	l.rising	˨˦	˨˩˧	˨˩˧	˨˩˧
z	l.rising falling	˨˦˨	˩˩˦ ˨˩˦ ˩˧	˩˧	˩˩˦ ˨˦ ˨˩˦
v*	h.level	˥	˦	˦	˥˦

Results were consistent across training, testing, and cross-validation for (u)-tone (high mid level ˦): it may be better described as mid level ˧. Similarly, results for v-tone suggest that the speaker produced a mid rising tone ˧˦ instead of the high rise falling tone ˦˥˦ described by [8]. The high level v*-tone ˥ is a bit lower, but still level ˦, according to both the training and testing results. The greatest discrepancy in our results is perhaps that of the x-tone, which is described as low rising ˨˦ by [8]. Our results suggest that it has merged with the z-tone. In fact, we conclude that z → x for the present speaker, since our results point to a low falling or low rising tone, most similar to the canonical low rising x-tone. There is no evidence in our data for a low rising falling contour ˨˦˨, i.e., a canonical z-tone.

We believe our method may be of considerable value to field linguists working on lexical tone. It can serve as both a check on the auditory impressions of investigators and can help substantiate similarity and difference between f0 contours in a rigorous way. Other factors known to be relevant to lexical tone, including voice quality and length, could also be incorporated into the models presented here.

5. REFERENCES

[1] Yip, M. 2002. *Tone*. Cambridge: Cambridge University Press.

[2] Hyman, L. M., Tadadjeu, M. 1976. Floating tones in Mbam-Nkam. In: Hyman, L. M., (ed), *Studies in Bantu tonology*. Los Angeles, CA: University of Southern California 57–111.

[3] Maddieson, I. 2001. Phonetic fieldwork. In: Newman, P., Ratliff, M., (eds), *Linguistic Fieldwork*. Cambridge: Cambridge University Press 211–229.

[4] Chen, S. H., Wang, Y. R. 1990. Vector quantization of pitch information in Mandarin speech. *IEEE Transactions on Communications* volume 38 1317–1320.

[5] Andruski, J. E., Costello, J. 2004. Using polynominal equations to model pitch contour shape in lexical tones: An example from Green Mong. *Journal of the International Phonetic Association* 34, 125–140.

[6] Grabe, E., Kochanski, G. 2007. Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and Speech* 50, 281–310.

[7] Roengpitya, R. 2007. The variations, quantification, and generalizations of Standard Thai Tones. In: Ohala, M., Beddor, P. S., Solé, M.-J., (eds), *Experimental Approaches to Phonology in Honor of John Ohala*. Oxford: Oxford University Press.

[8] Purnell, H. C. 2012. *An Iu-Mienh–English Dictionary*. San Francisco, CA: Silkworm Books / Center for Lao Studies.

[9] Xu, Y. 2013. ProsodyPro: A tool for large-scale systematic prosody analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)* Aix-en-Provence, France. 7–10.

[10] Tian, J., Nurminen, J. 2004. On analysis of eigenpitch in Mandarin Chinese. *2004 International Symposium on Chinese Spoken Language Processing* 89–92.

[11] Jombart, T. 2008. adegenet: a r package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405.

---

[1] [8] calls this tone 'unmarked'; we abbreviate it and use parentheses to alert the reader that it is our own non-standard convention, hence '(u)-tone' or '(u)' throughout.