Production and perception maps of the multidimensional register contrast in Yi

Jianjing Kuang

kuangjianjing@ucla.edu

Abstract

In Yi languages, multidimensional cues are involved in the tense vs. lax contrast: phonation, vowel quality and pitch. The relative contributions of these cues vary across languages and dialects and thus raise questions for production and perception: what is the integrated effect of multidimensional phonetic cues in production? And will these cues have the same contributions in the perception space? The relationship between production and perception maps with multidimensional cues has been a challenge for linguists and psychologists. This paper proposes an approach to generating perception vs. production MultiDimensional Scaling maps of physical measurements as categorized by listeners vs. speakers. In this way, a perception space can be directly compared with a production space. The results from production and perception experiments in Southern Yi show an overall faithful match between production and perception maps, including that the low-vowel contrastive pairs are more different and distinguishable than the high-vowel contrastive pairs. This is due to the significant contribution of vowel quality to the low-vowel contrast.

I. INTRODUCTION

Non-modal phonations are used in most languages as a prosodic cue at the sentence level, but some languages have a phonation contrast at the word level: everything else equal, the contrast between modal and non-modal phonation can distinguish the meanings of words. Languages vary in how they contrast phonations. For example, Hmong (Huffman, 1987; Esposito *et al.*, 2009) and Gujarati (Fischer-Jørgensen, 1967**;** Khan, 2009) contrast breathy and modal phonations, and Mazatec (Kirk *et al*., 1993; Garellek and Keating, 2011) makes a three-way contrast of breathy, modal, and creaky phonations. In Tibeto-Burman languages, phonation contrasts are part of phonological register (tense vs. lax) contrasts. This is different from the type of tense vs. lax contrast found in, e.g., Germanic languages, which does not involve a phonation contrast but instead is primarily a tongue position and duration contrast. The register contrasts in Tibeto-Burman languages commonly can combine multiple dimensions: F0, phonation and even vowel quality. Due to different interactions between these phonetic properties, their contributions to register contrasts seem to vary widely across languages and dialects.

Yi languages, a branch of Tibeto-Burman languages, are a good example of phonation-based register contrasts. Within this branch is a language also called Yi; there are at least six dialects of this language, and their tense vs. lax vowel syllables involve a phonation contrast (e.g. Shi and Zhou, 2005; Edmondson and Esling, 2006; Kuang, 2011), but the contributions of vowel quality and pitch to their tense/lax contrasts vary a lot across dialects. For example, Shi and Zhou (2005) found that in their Yi dialect, a tense syllable (e.g. b<u>e</u>33, where the underscore indicates tenseness and the numbers indicate tone, here mid) has a significantly higher F0 than the corresponding lax syllable

(be33), though the two syllables share the same tonal category. In other dialects, instead, Maddieson and Ladefoged (1985) reported that F1 values of tense syllables tend to be higher than those of lax syllables, indicating that tense syllables may have a lower tongue position than lax syllables. Laryngoscope studies (Esling *et al.*, 2001; Edmondson and Esling, 2006) have revealed a retraction of the tongue root in the tense syllables of Northern Yi. The role of vowel quality was further confirmed in our recent study of Southern Yi (Kuang, 2011): F1 is consistently higher for tense syllables, and makes a statistically significant contribution to producing the tense vs. lax contrast. Therefore, tense vs. lax contrasts in the Yi language, and Yi languages more generally, may involve at least three dimensions of articulation: a phonation difference (characterized by a number of acoustic correlates), a pitch difference (characterized by F0) and vowel quality (characterized by F1). The variation in these multidimensional cues raises a question for both production and perception: What is the integrated effect of these multidimensional phonetic correlates in production, and again how do they integrate in perception? When production of a phonological contrast involves multidimensional phonetic correlates, the weights (or relative contributions) of these cues are not equal in perception (Holt and Lotto, 2006). So a challenge for phoneticians and psychologists is to determine whether perception and production of register contrasts have the same cue weightings.

Moreover, some kinds of tense vs. lax contrasts in Yi languages are apparently easier to keep than others. For example, a tense vs. lax contrast is less frequent in low vowels than higher vowels in general; and as an extreme case of this, the register contrast in low vowels has totally vanished in the Northern Yi dialect. So it must be the case, as suggested in previous studies (Johnson, 2003; Iverson and Kuhl, 1995), that the degrees

of distinctiveness among members of a category are not identical. A perception space can appear warped relative to the corresponding production space by such influences as the perceptual magnet effect (Iverson and Kuhl, 1995); or it can be veridical to the production space, which is unevenly spaced. For example, Pols *et al.* (1969) found that production and perception of vowel spaces are excellently correlated. Jiang *et al*. (2007) found a faithful match between visual speech perception and optical phonetic signals. Which of these production-perception relations, warped or veridical, is more common is not settled yet, since very few studies directly compare native speakers' multidimensional production spaces and perception spaces. Part of the reason is that measures of perception and production are often of different kinds (categorical vs. continuous), so the data are not easy to compare directly. Given that phonological contrasts usually have multidimensional cues, the task is even more challenging. As production and perception of contrasts, including register contrasts, are driven by the same knowledge of native speakers, it is intriguing to know the distribution of contrastive exemplars in both perception and production space.

The goal of this paper is to relate perception spaces to production spaces of native speakers, investigating the distribution of contrastive tense and lax exemplars in both spaces, and comparing the contributions of different dimensions in perception and production. Inspired by Pols *et al.* (1969) and Jiang *et al.* (2007), this paper adopts Multi-Dimensional Scaling and various physical distance calculation methods. The approach presented in this paper will allow us to generate graphical production and perception maps that can be directly compared.

II. PERCEPTION EXPERIMENT

This experiment is aimed to derive a perception map showing the distribution of contrastive exemplars of tense vs. lax vowels in Southern Yi. By comparing this map with a map of the production properties of the same stimuli, we can better understand the relationship between these two aspects of speech.

A. Methods

1. Subjects

All the data in this study were obtained during a trip to Yunnan province of China in the summer of 2009. All the listeners were recruited from Xingping village, a small town in northeastern Yunnan. The main population in the village is Yi people. All the listeners had an education level of elementary school, which made most of them sufficiently comfortable interacting with a computer screen.

Ten listeners, five males and five females (m1-m5, f1-f5), ages 18 to 50, were paid for their participation. Southern Yi is their native language and also their primary language in daily life. All reported no speech, hearing, or language difficulties.

2. Stimuli

The stimuli were natural pronunciations of six native speakers (three males M1-M3 and three females F1-F3) of Xinping village. All are around 40-50 years of age, and use Yi as their primary language in everyday communication. Overall, speakers and listeners have no personnel overlap except for one female subject.

The original recording was of a large word list of monosyllable minimal pairs with all possible combinations of tone × register × vowels (details of elicitation

procedures of the fieldwork are in Kuang (2011)). Like most Yi languages, Southern Yi has three contrastive tones, i.e. low (21), mid (33) and high (55) (numbers indicating the level of pitch); as well as two contrastive registers, tense and lax. The register contrast can occur with all the vowels, though the contrast for high vowels is overall more frequent than that for low vowels. The register contrast can also occur with two of the tonal categories, i.e. mid (33) and low (21) tone, but not with the high tone.

Before making the recording, the speakers were asked to go over the word list, checking the contrasts in the minimal pairs. Non-contrastive pairs were excluded. For a total of 12 speakers, simultaneous electroglottograph (EGG) and audio recordings were made. The signals were recorded directly to a computer via its sound card, in stereo, using Audacity, at the sampling rate of 22050 Hz per channel. The audio signal was recorded through a Shure SM10A microphone as the first channel. EGG data were obtained by a two-channel electroglottograph (Model EG2, Glottal Enterprises) and recorded as the second stereo channel. Each word was repeated twice.

Stimuli were selected from these original recordings. The subset selected as stimuli contains syllables with initial [b] and two representative vowels, one high ([u]) and one low ([ɛ], hereafter [e]), with two registers (tense, indicated by underscore; lax, without underscore) and the two tones they can occur with (mid33, low21; high55 is thus excluded as there is no register contrast with this tone). Every test item had pronunciations available from all six speakers. We thus extracted four groups of stimuli with all combinations of tone and vowel height:

A. High vowel with mid tone (bu33/bu̲33)
B. High vowel with low tone (bu21/bu̲21)

C. Low vowel with mid tone (be3/be̲33)

D. Low vowel with low tone (be31/be̲31)

Various acoustic measures of the stimuli were then extracted in order to confirm that this subset is representative of the larger recorded corpus. Measures included in this study were all made by VoiceSauce (Shue *et al.*, 2009) and are: H1*-H2* (corrected version by Iseli *et al.*, 2007); amplitude of H1 relative to the amplitudes of F1, F2, and F3 (H1*-A1*, H1*-A2*, H1*-A3*); Cepstral Peak Prominence (CPP); H2*-H4*. Other acoustic measures include formant frequencies (F1, F2), pitch (F0) and energy. The EGG analysis was done by EggWorks (Tehrani, 2009). Two measures were extracted from the EGG signals: Contact Quotient (CQ) and Peak Increase in Contact (PIC). Refer to Kuang (2011), a previous production study of these recordings, for detailed review and explanations of these measures.

The previous production experiment (Kuang, 2011) showed that a CQ distinction in electroglottographic signals is the statistically most powerful property of the phonation contrast, while the acoustic measures H1*-H2* and H1*-A1*, which are significantly correlated with CQ, are the next best measures for the phonation contrast. The bandwidth of the first formant (B1) and the Cepstral Peak Prominence (CPP) are effective acoustic cues too. In addition, a consistent F1 difference was found. No F0 difference was found, however. Thus in this language, listeners could possibly use phonation correlates or F1, but not F0, to distinguish tense from lax.

Unpaired t-tests on all the measures demonstrated that this subset did not differ significantly from the whole dataset (all p-values > 0.05) and thus can represent it.

3. Procedures - AXB discrimination task for tense vs. lax

These audio stimuli were put into an AXB discrimination task. The pronunciations of M1 and F1 were chosen as the standard for this task, as they maintained a good and typical contrast between tense vs. lax. (i.e. all their tense vs. lax pairs are highly significantly distinctive along the relevant measures (Kuang, 2011), and their productions are near the center of all speakers' productions (Kuang, 2011)). The minimal register contrast pairs which were produced by these two speakers served as the A and B, and the Xs were the pronunciations of all six speakers. For example, here are 4 possible trials:

F1_bu33(A), F2_bu33(X), F1_bu̲33(B) M1_bu33(A), F2_bu33(X), M1_bu̲33(B)

… …

F1_bu33(A), M3_bu33(X), F1_bu̲33(B) M1_bu33(A), M3_bu33(X), M1_bu̲33(B)

Therefore, the listeners heard 20 stimuli in each group (half compared to F1 and half compared to M1), thus 80 stimuli in total. This stimulus set was presented three times to each listener.

The task was run by a Praat script on a computer. Stimuli were played through SONY MDR-NC60 headphones. On the screen, the listeners could see three buttons, labeled as (A), (X) and (B). The buttons of A and B were in yellow and clickable. The listeners heard three stimuli in sequence separated by 0.5 second, and had to decide whether the second (X) is more similar to the first (A) or to the third (B). Listeners had to make a response for every trial by clicking either A or B. They were able to replay the audio as often as necessary before responding, and they also could “regret” and go back to re-listen to the previous sequence. There was an introduction and a practice session

before the formal test. For those who had difficulties operating the computer, the author asked the subjects simply to point on the screen, and assisted them in clicking the mouse. The duration of the experiment by design was under 40 min. Listeners could pause if they felt tired.

B. Results

Listener m2 failed to perceive any differences in the stimuli, and thus is excluded from the data analysis. Thus data from 9 listeners are presented here.

Figure 1 is a set of four-fold displays (Friendly, 1994). A four-fold display shows the frequencies in a 2 x 2 table in a way that depicts the correctness ratio and the distribution of responses. In this display the frequency of responses in each cell is shown by a quarter circle, so each quarter circle represents one of four types of answers, relative to the X stimuli (i.e. stimulus:response = L:L, L:T, T:T and T:L). The radius is proportional to the square root of the count, so the area indicates the proportion. An association between the stimulus and response is shown by the tendency of diagonally opposite cells, with wrong answer types (i.e. stimulus:response= L:T, T:L) in one direction, and correct answer types (i.e. stimulus:response= L:L, T:T) in the other direction. We use color/shading to distinguish the directions: dark shading indicates the correct types whereas light shading represents the wrong types. Confidence rings for the observed data provide a visual test of the null hypothesis of no association.

For example, the bottom left panel shows that for 360 stimulus:response pairs (180 each tense and lax stimuli) under the condition of mid tone (33) + low vowel (e), 30% out of the total are L:L and 41% are T:T, which constitute a total of 71% correct answers; on the other hand, the wrong answers are composed of 20% L:T and 9% T:L.

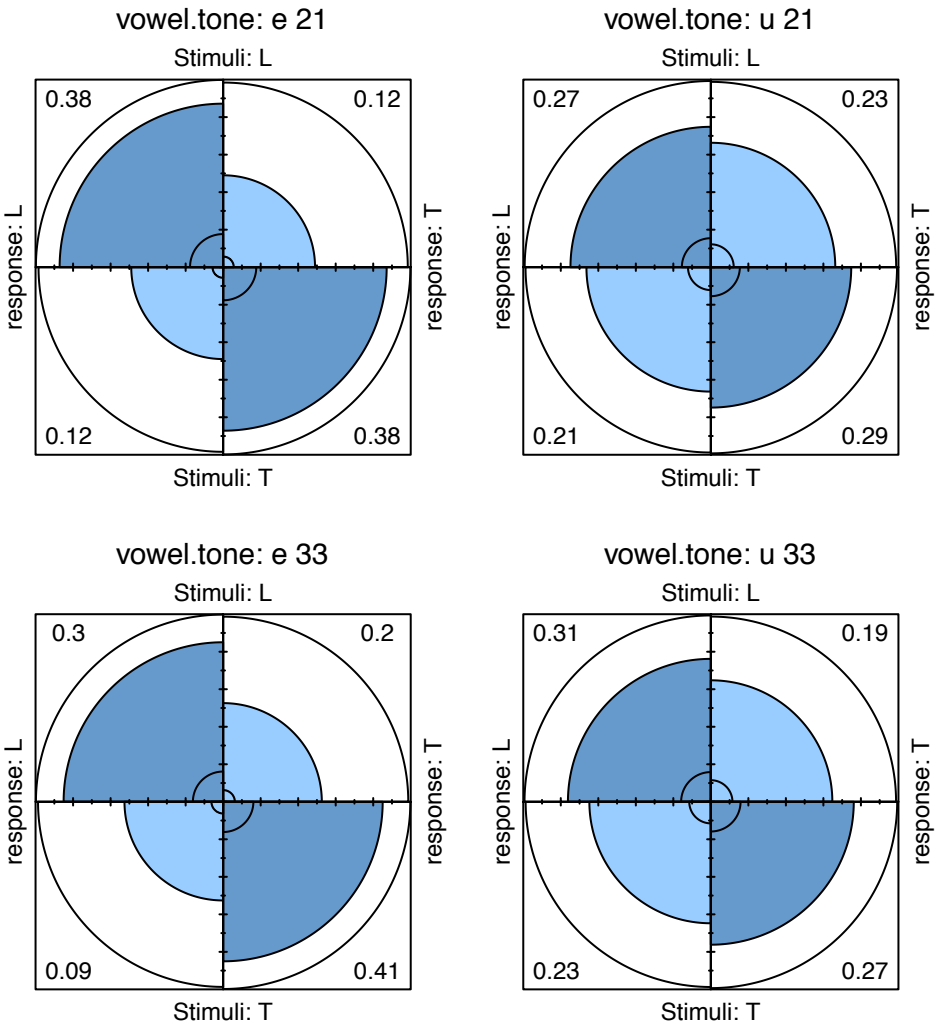

FIG. 1. Four-fold displays for four conditions. The frequency of responses in each cell is shown by a quarter circle, so each quarter circle represents one of four types of answers, relative to the X stimuli (i.e. stimulus: response = L:L, L:T, T:T and T:L). The radius is proportional to the square root of the count, so the area indicates the proportion. The dark shading indicates the correct types (i.e. stimulus:response = L:L, T:T) whereas the light shading represents the wrong types (i.e. stimulus: response=L:T, T:L). Confidence rings for the observed data provide a visual test of the null hypothesis of no association.

According to Figure 1, in the low vowel + low tone condition (e21), the accuracy rates for lax and tense stimuli are both 38%. In the low vowel + mid tone condition (e33), the accuracy rates for lax and tense stimuli are 30% and 41% respectively, slightly better

for the tense condition. Accuracy rates for the high vowel conditions are generally lower. In the high vowel + low tone condition (u21), the accuracy rates are only 27% and 29% for lax and tense stimuli. The high vowel + mid tone condition (u33) are slightly better, 31% and 27% correct for lax and tense stimuli. Counting accuracy by tone and vowel, the average accuracy rate for low vowels is 73.5% compared to 57% for high vowels while accuracy rates for both tone conditions are nearly identical: 66% for the low tone and 64.5% for the mid tone. Comparing the answer rates across panels, it can be concluded that low vowels generally have higher correctness rates than high vowels, and tonal condition does not affect accuracy.

III. PERCEPTION AND PRODUCTION MAPS

A. Perceptual distances of tense vs. lax contrasts from confusion matrix

What does this confusion pattern suggest in the perception space? Here we adopt aspects of Shepard (1972) and Johnson (2003)'s method to generate a perception map via a perceptual confusion matrix. If “A” refers to lax and “B” refers to tense, with capitals used for stimuli and lower case for responses, and if 6 of 10 [bu33] sound like “bu33”; then if P stands for proportion, we can define:

	(A)[bu33]	(B)[bu33]
[a] “bu33”	PAa	PBa
[b] “bu33”	0.6(PAb)	PBb

TABLE I. Similarity matrix (distances of minimal pairs, blank if there is no comparison).

	be21	**be21**	**be33**	**be33**	**bu21**	**bu21**	**bu33**	**bu33**
be21	0	1.159						
be21	1.159	0						
be33			0	0.861				
be33			0.861	0				
bu21					0	0.234		
bu21					0.234	0		
bu33							0	0.336
bu33							0.336	0

Similarly, PBa is the proportion of how many tense "bu33" are heard as lax syllables, etc..With this kind of proportional matrix, we can calculate the similarity of different vowels, given in Table I, by the following equation:

$$S_{ij} = (P_{ij} + P_{ji})/(P_{ii} + P_{jj}) \quad (1)$$

The negative of the natural log of the similarity is used to calculate the perceptual distance (dissimilarity):

$$d_{ij} = - \ln(S_{ij}) \quad (2)$$

Figure 2 shows the resulting distances of the tense vs. lax contrast in four phonological conditions.

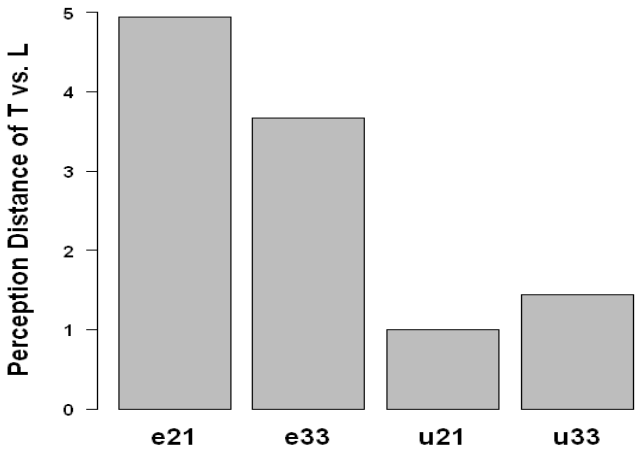

FIG. 2. Perceptual distances of tense vs. lax contrast in four phonological conditions: low vowel + low tone (e21); low vowel + mid tone (e33); high vowel + low tone (u21); high vowel + mid tone (u33).

As shown in Figure 2, the perceptual distances of tense vs. lax contrast pairs reflect the different perceptibilities of these pairs, consistent with the accuracy pattern in Figure 1: the low-vowel pairs generally have better perceptibility than the high-vowel pairs. The low vowel+low tone pair has the best perceptibility and the high vowel + low tone pair has the least perceptibility.

According to Shepard's/Johnson's approach, a MultiDimensional Scaling (MDS) function can be employed to plot these stimuli in a low dimensional perceptual space, usually two or three dimensions. To generate such an MDS map requires a complete confusion matrix that includes the distances between every possible pair, even if most of them are not essential to the study (e.g. e21 vs. u33 here). However, in a fieldwork experiment like ours, since it is usually not practical have an extremely long experiment to compare the similarity of every pair, the confusion matrix is not complete (shown as the blanks in Table I), and thus these distances cannot be plotted by MDS functions.

The paper will therefore present an alternate approach to generating an MDS perception map when a complete discrimination task is not available. This approach is based on the stimuli rather than on confusions, and can generate a perception map and a production map at the same time.

B. Perception map vs. production map

1. Production map

It is easier to start with the production map. MDS can sum up effects from all the individual production measurements and reveal the overall physical distance between every pair of tokens. The algorithm works as follows (Kruskal and Wish, 1978: 27-28): 1) Use distance functions to compute distances (matrix D) among categories in an original high k-dimensional space (here 12 physical measures are the coordinates of a 12-dimension space); 2) Find a low p-dimension space (p can be any number between 1 to k-1, here 1-11) to best visually present the distances among categories. To do so, 2a) compute the distances among all pairs of points, to form their dissimilarity matrix (d) in this low p-dimension space, and 2b) compare this matrix (d) with the input data matrix (D) by evaluating the stress function. The smaller the stress value, the greater the correspondence between the two. Adjust the coordinates of each point in the direction that best minimizes the stress until the stress won't get any lower. (For our case, a 2-dimension space is adequate to present the data.)

Two popular distance functions can be used to calculate dissimilarity based on physical measurements: Euclidean distance and Manhattan distance. We employ the Manhattan distance in this study since it yields similar results to Euclidean distance, but

with simpler calculations for averaging differences across dimensions. Here is the formula for the distance between p and q over i dimensions:

$$d_1(p, q) = ||p-q||_1 := \sum_{i=1}^{n} |pi - qi| \quad (3)$$

The MDS presented here was performed by using R package Vegan version 1.18-29 (Oksanen *et al*., 2011). Figure 3 is the resulting production map of the tense vs. lax contrast. The relations of the MDS dimensions to the actual physical measures will be discussed below.

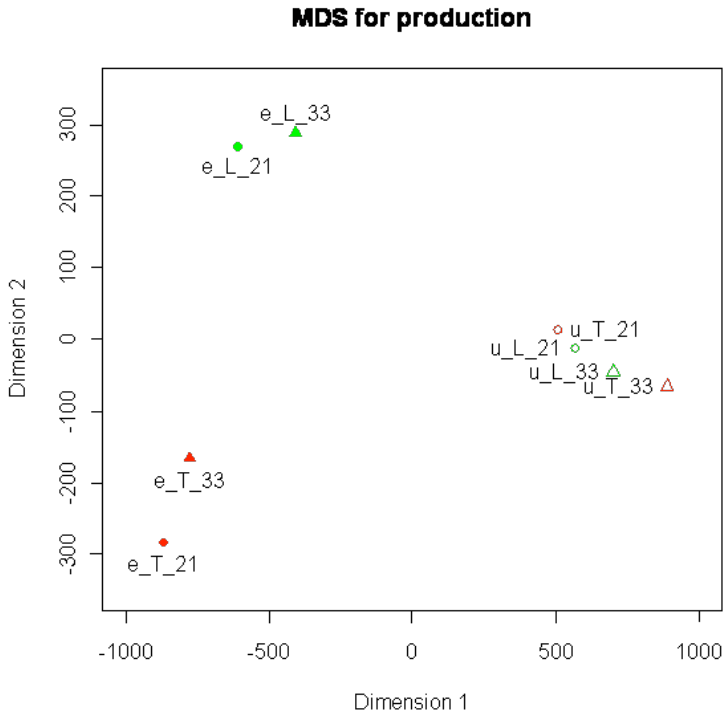

FIG. 3. Production map of the tense vs. lax contrast in a 2-D space, by vowel and tone combinations. The distances are the visual presentation of the similarities in production among the stimuli. The stimuli that are produced similarly by the speakers are likely to cluster together, whereas the ones that are produced distinctively are likely to be far apart.

2. Perception map

The perception map is generated in a similar way, from the same stimulus tokens, but with different labels. In the production experiment, the label of ‘tense’ or ‘lax’ is decided by speakers (based on the recording script); whereas in the perception experiment, the label of 'tense' or 'lax' is decided by listeners. Specifically, in the perception experiment, the AXB task can be treated as an identification task, in which every stimulus was categorized by listeners. Similar calculations as above were carried out, changing only the categorization of the stimuli to correspond to listeners’ responses. The resulting perception map – of exactly the same stimuli as in Figure 3, but spaced and labeled according to listeners’ categorizations of them – is plotted in Figure 4.

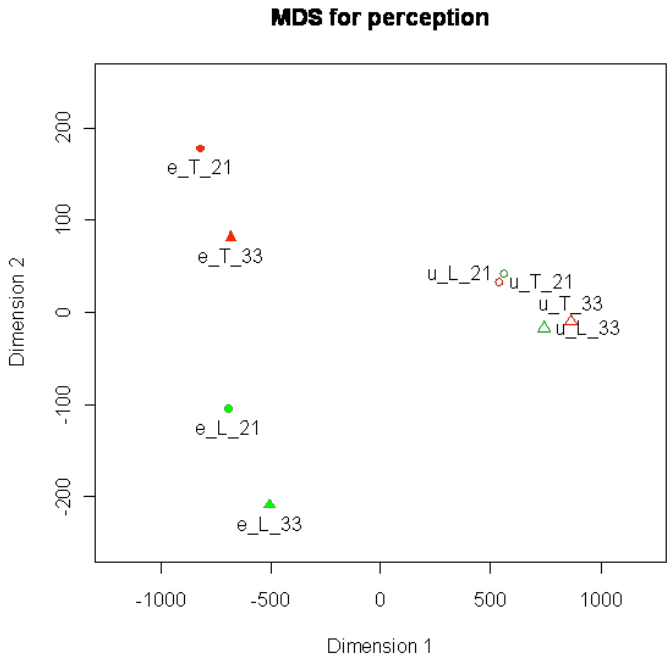

FIG. 4. Perception map of the tense vs. lax contrast in a 2-D space, by vowel and tone combinations. Stimuli categorized by speakers in FIG. 3 are here recategorized (relabeled) by listeners. The distances are the visual presentation of the similarities in perception of the stimuli. The stimuli that sound similar to listeners are likely to cluster together, whereas the ones that sound distinctive are likely to be far apart.

3. Comparison between these two maps

Since the two maps are generated based on the same dataset, they can be compared directly. As shown in the two maps, the perception map is generally faithful to the production map, though due to perceptual confusions, the scale of the perception map shrinks a little. The low vowel + low tone pair has the best distinctiveness in production and also the best perceptibility in perception. This pattern is the same as for the perceptual distances seen in Figure 2.

It is interesting that the tonal effect (differences between pairs with the same vowel and register but different tones) is more salient in the perception map than in the production map, as the cut-off between low tone and high tone is clearer in perception. Since they are based on a complete matrix, more information can be read from the two maps. For example, dimension 1 of the spaces mostly distinguishes low vowels from high vowels; and dimension 2 is more about phonation and tone. Interestingly, the distinctive cues for register in low vowels vs. high vowels are different. Dimension 2 mostly distinguishes low-vowel clusters by phonation (tense vowels on one side and lax vowels on the other side), but for the high vowels, they are mostly distinguished by tone. These contributions of measures to dimensions of the perception and production maps can be quantitatively extracted by the metaMDS function in the R vegan package.

TABLE II. Weights of phonetic measures in each dimension V of the production map.

	V1	V2	V3
H1*-H2*	1.223	0.803	0.408
H2*-H4*	1.011	0.168	0.334
H1*-A1*	0.671	0.991	0.183
H1*-A2*	0.516	0.530	0.789
H1*-A3*	1.060	0.458	0.190
CPP	0.604	0.280	0.790
PIC	1.468	0.278	0.200
CQ	0.568	1.451	0.411
F0	0.352	0.237	0.671
F1	2.414	0.953	0.567
F2	1.375	0.409	0.219
F3	1.355	0.012	0.290

TABLE III. Weights of phonetic measures in each dimension V of the perception map.

	V1	V2	V3
H1*-H2*	1.132	0.323	0.087
H2*-H4*	0.996	0.206	0.224
H1*-A1*	0.718	0.813	0.228
H1*-A2*	0.493	0.362	0.280
H1*-A3*	1.055	0.176	0.031
CPP	0.523	0.606	0.316
PIC	1.425	0.125	0.122
CQ	0.619	0.329	0.359
F0	0.377	0.568	0.138
F1	2.269	0.782	0.388
F2	1.374	0.143	0.087
F3	1.271	0.427	0.217

The bigger the absolute values, the bigger the contributions of the weight in tables II and III. So perceptually, vowel-quality related measures (F1, F2 and F3) contribute most to dimension 1. For dimension 2, the phonation-related measure H1*-A1* becomes the dominant perceptual cue. Similarly, dimension 1 in the production map is mostly about vowel quality and dimension 2 mostly distinguishes the phonations (as measured by CQ). Notice that tone-related cue in the perception map (F0) are weighted more than they are in the production map.

In terms of acoustic cues that cause confusion in perception, comparing the weights between production and perception, we can conclude that the weights of the phonation cues are generally smaller in the perception space than they are in the production space. It can also be seen that the distance between tense and lax phonation becomes less in the perception space.

IV. FURTHER ANALYSIS AND DISCUSSION

In the previous section, we employed different approaches to generate perceptual distances from a confusion matrix, a perception map, and a production map of the tense vs. lax contrast in Yi. The remarkable result is that perception matches fairly well with production. The general pattern is that the tense vs. lax pairs in different phonological conditions are not equally contrastive, and the low-vowel pairs are more distinguishable in general than the high-vowel pairs. Figure 5 is the generalization of the distances of contrastive pairs in production and perception (both Shepard/Johnson and new approaches), clearly exhibiting a similar pattern.

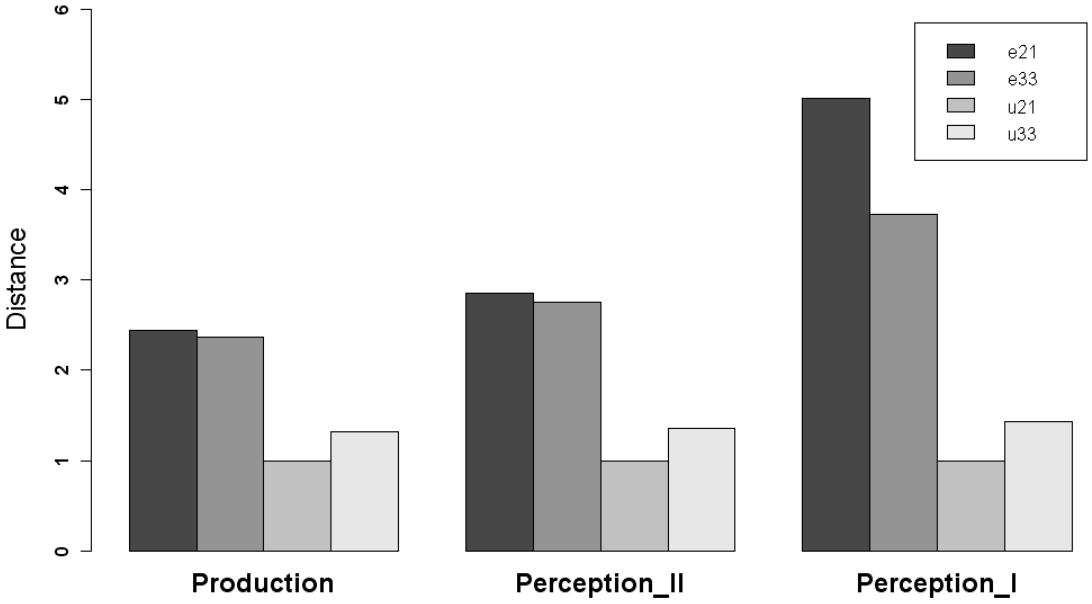

FIG. 5. Normalized distances of tense vs. lax contrast in four phonological conditions. Distances are clustered by type of maps. 'Production' refers to production map; 'perception-I' refers to the perceptual distances from perceptual confusion matrix; and 'perception-II' refers to the perception map from perceptually recategorized physical measurements.

Nonetheless, Shepard/Johnson's perceptual distances (perception-I in Figure 5) and the perception map (perception-II in Figure 5) have slightly different meanings. Perceptual distances are based on the confusion matrix, so they are purely from perception data. These distances can tell us which contrast is more perceptible than the others. On the other hand, the perception map is the perceptual categorization of the physical stimuli, and it tells us what people hear given the stimuli. Owing to the advantage of the MDS method, which can convert different types of data into comparable pattern plots, the perception map can then link to the production map of the stimuli. These maps provide us a whole picture of native speakers' and listeners' knowledge about their tense vs. lax contrast. Their match is evidence that perception and production of the

phonological contrast is driven by the same knowledge and that listeners are good at picking up the acoustic cues in the speech signals.

Though the phonation contrast is the distinctive feature for the tense vs. lax contrast in Southern Yi, vowel quality and tone also play some role in perception. A good model should be able to capture all the useful cues that listeners and speakers use. The approach in this paper is robust in preserving fine phonetic details for each exemplar, and can naturally deal with multidimensional cues in phonological contrast, which is meaningful since multidimensional is more common than unidimensional in contrasts. The approach is also very practical for any fieldwork perception experiments, since no synthesized signals are required and only an identification task is needed, which is simpler than a complete discrimination task.

What do these production and perception maps imply for linguistic study? The question why a phonological process or sound change is more likely to happen to some exemplars than to others is one of the central topics in linguistics. Previous studies have found that a phonological process is more likely to take place between exemplars that are close in the perceptual map. For example, Huang (2004) has shown that Mandarin Tone 214 and Tone 35 are closer in the perceptual map of Mandarin tones than any other tones are, and has suggested that that is why the well-known Mandarin tone sandhi happens between these two tones. We can expand this conclusion by considering a corresponding tone production map. Figure 6 is a production map of Mandarin tones, produced by 10 Mandarin speakers (gender balanced, age 20 to 25) from Beijing. It is clear that this map matches Huang's perception map in terms of tonal similarities. That is, Tones 214 and 35 are not just heard as more similar than other tone pairs, but they are actually produced

more similarly, so that the perception simply tracks the production differences, which could equally well be related to the tone sandhi effect.

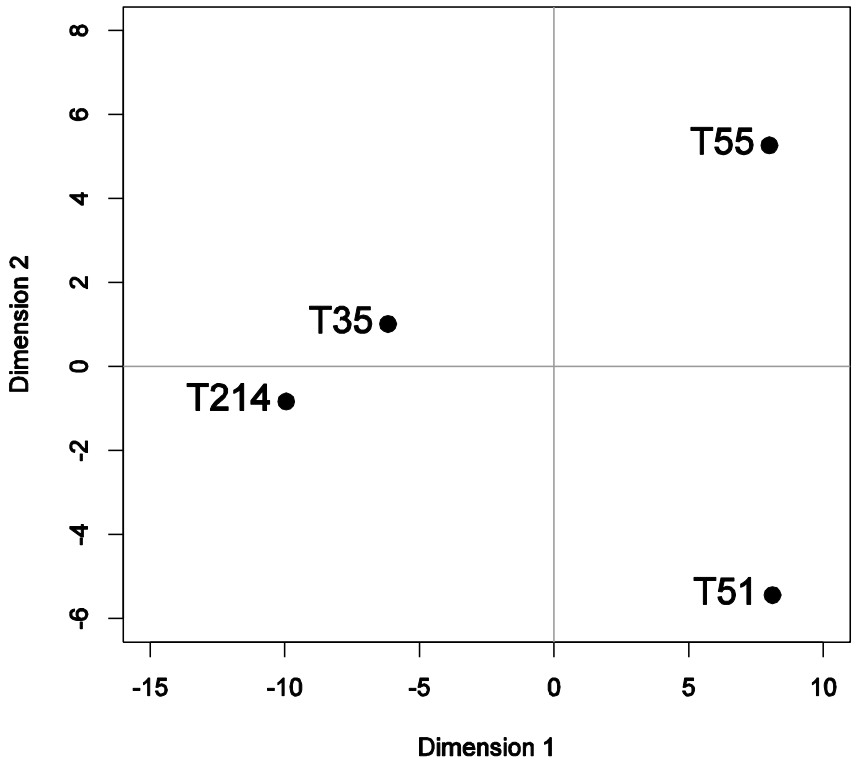

FIG. 6. Production map of Mandarin tones. Distances in this map indicate the similarities in producing Mandarin tonal categories. This production map can be compared with the perception map by Huang (2004:48). It can be seen that T214 and T35 are the most similar categories not only in the perception map but also in the production map.

In our case of the Yi languages, we are more interested in diachronic sound change. Low-vowel pairs are more distinguishable than high-vowel pairs in both production and perception; what do these maps indicate for future sound change? Given the fact mentioned at the beginning of this paper, that low-vowel pairs preserve fewer phonation contrasts than do high-vowel pairs, this is a little surprising at first glance. But this could

happen if the dominant cue for distinguishing tense vs. lax contrast were no longer phonation; and this cue is only robust in the low-vowel pairs but not the high-vowel pairs.

A logistic regression model was run for the stimuli's production data to evaluate the contributions of phonation, tone and vowel quality. To simplify the statistical model, only three measures are presented here, each taken to be the best representative of a phonological dimension: CQ as the phonation measure, F0 as the tone measure, and F1 as the vowel quality measure. Since the various spectral measures that contribute to the phonation contrast essentially reflect CQ, and CQ can account for the most variability, we use CQ instead of any acoustic measure to represent phonation. As indicated in Figure 7, indeed, vowel quality (i.e. F1 here) makes a significant contribution to the low-vowel contrast pairs but little contribution to the high-vowel pairs. A subsequent paired t-test was conducted to examine the degree of distinction of the phonation in different vowel quality situations. It turns out that the tense vs. lax contrast has a much more distinctive CQ in the high vowel pairs ($t(54)= -5.78$, $p=3.894e-07$, $p<0.01$), compared to that in the low vowel pairs ($t(54)=2.23$, $p=0.03$). This is just as Pierrehumbert (2003) notes: When a phonemic contrast is carried by a single phonetic dimension, the difference along that dimension is larger than when the contrast is carried by multiple dimensions. Therefore, the tense vs. lax contrast in low-vowel pairs benefits a lot from the vowel quality difference. Perhaps vowel quality is an easier and more salient cue than phonation.

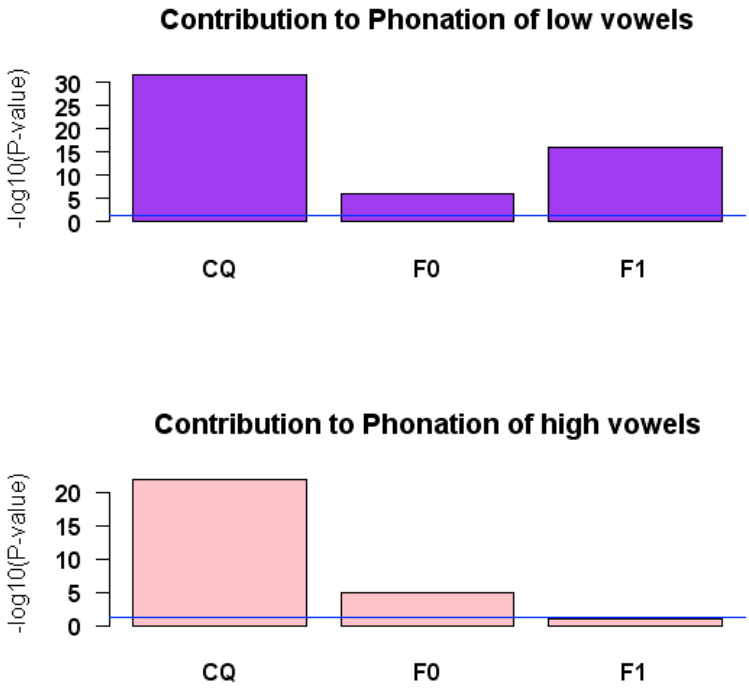

FIG. 7. Contribution to the tense vs. lax contrast for different vowel qualities. P-values of the logistic regression models are converted into –log10 values, so the more significant a cue the bigger the converted –log10(p-value). These numbers are visually presented as bars, the heights of which indicate the importance of the cues. Here, phonation is the most important cue for both high and low vowels as indicated by the height of CQ. F1 makes a significant contribution to the low-vowel pairs but not to the high-vowel pairs.

Furthermore, individual perception maps can reveal different preferences between vowel quality and phonation in perception, as seen in Figure 8 and Figure 9.

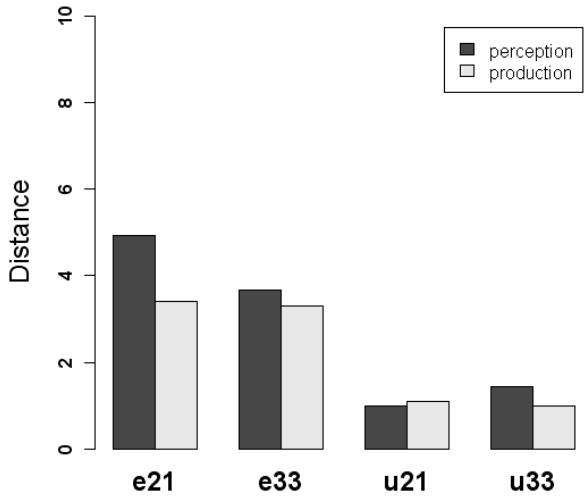

FIG. 8 Overall distances of tense vs. lax in four phonological conditions: Darker bars represent the perception distances (all listeners), while production distances (for all stimuli) are lighter bars.

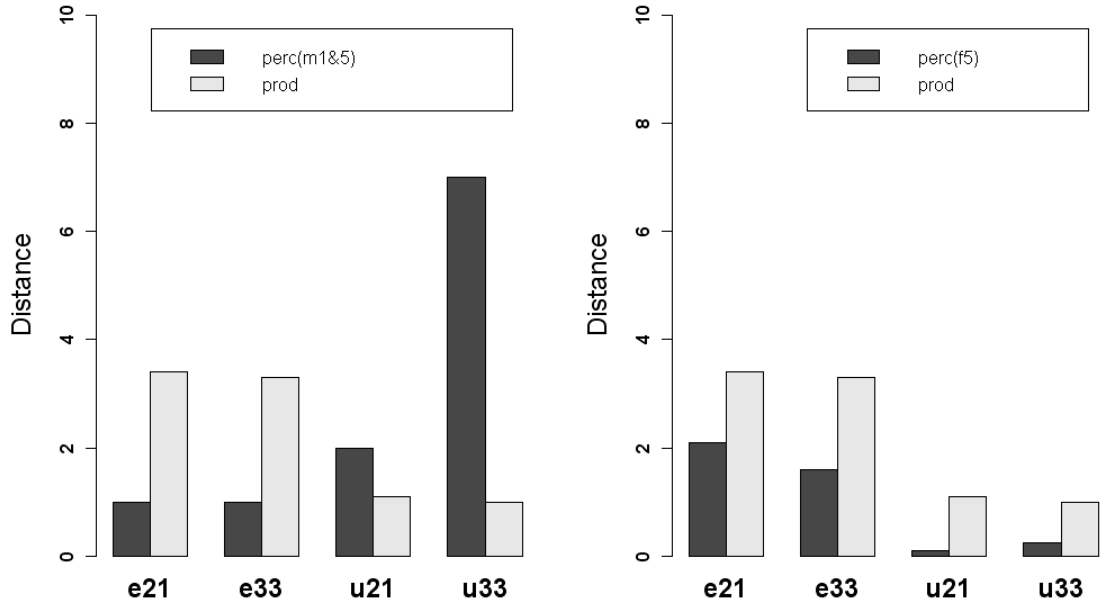

FIG. 9 Variation in perception maps: Darker bars represent the perception distances (left: combination of m1 and m5; right: f5), while production distances (for all stimuli) are lighter bars.

As can be seen, all the listeners are able to hear the difference between tense and lax (the overall correct ratio is bigger than 0.5). Most listeners hear the difference much better in the low vowel pair, when phonation has a smaller difference but vowel quality has a significant contribution; however they perform less well on the high vowel pairs, when the vowel quality difference is absent though the phonation difference is much bigger. This suggests that these listeners are more sensitive to vowel quality differences than to phonation. With the absence of vowel quality differences, f5 fails to perceive the different phonation categories for [u] at all. Therefore, vowel quality is such a robust cue that some (but not all) listeners do not even pay attention to the phonation difference. This fact could suggest that the sound change that has previously happened in the Northern dialect (low vowel pairs contrast by vowel quality) is now beginning to happen in this Southern dialect.

It is also interesting to notice that m1 and m5 perform differently from the other listeners. These two listeners hear the contrast better in the high vowel pair but totally fail to distinguish tense vs. lax categories in the low vowel pair, even when the vowel quality difference (F1) is a very strong cue for the low vowel pair. That means these listeners do not pay attention to the vowel quality distinction. They use only phonation as the cue for the tense vs. lax contrast. Given that CQ is more saliently different in the high vowel pair than in the low vowel pair, these listeners could hear a difference better in the high vowels than in the low vowels.

In summary, the present study adopted the MDS method to generate production and perception maps directly from categorizations of the stimuli by either speakers or listeners. The new perception map is consistent with the perceptual distances generated

via a confusion matrix, which validates the feasibility of the new approach. These maps clearly visualize the mental representations of native Yi speakers' knowledge about their tense vs. lax contrast, and the unbalanced distribution of contrastive exemplars matches between the production and perception maps. These maps can preserve the fine phonetic details from multiple phonological dimensions, and changes in cue weightings can indicate possible future sound change.

ACKNOWLEDGMENTS

This paper is based on part of my MA thesis. This work was supported by NSF grant BCS-0720304 to Patricia Keating and a 2009 summer research reward from the UCLA Linguistics Department. I am heartily thankful to Professor Patricia Keating for her encouragement, guidance and support throughout this work. I also would like to thank Yen-Liang Shue for VoiceSauce and Henry Tehrani for Eggworks. I am grateful for the invaluable comments from Professors Jody Kreiman, Sun-Ah Jun and Bruce Hayes. The fieldwork of this study was largely helped by Professors Jiangping Kong, Feng Wang and Baoya Chen at Peking University, Haichao Yang at Yunan University and Yan Lu at Yunnan Nationalities University.

Reference:

Edmondson, J. A., Esling, J.,Harris J. G., Li, S. and Ziwo, L. (**2001**). "The aryepiglottic folds and voice quality in the Yi and Bai languages: laryngoscipic case studies," Mon-Khmer Studies **31**, 83-100.

Edmondson, J. A., and Esling, J. H. (**2006**). "The valves of the throat and their functioning in tone, vocal register, and stress: laryngoscopic case studies," Phonology **23**, 157-191.

Esposito, C. M., Ptacek, J. and Yang, S. (**2009**). "An acoustic and electroglottographic study of White Hmong phonation," J. Acoust. Soc. Am. **126**, 2223(A).

Fischer-Jørgensen, E. (**1967**). "Phonetic analysis of breathy (murmured) vowels in Gujarati," Indian Ling. **28**, 71–139.

Friendly, M. (**1994**). "A fourfold display for 2 by 2 by k tables," Technical Report **217**, York University, Psychology Department. This is available online at < http://www.math.yorku.ca/SCS/Papers/4fold/4fold.ps.gz> (last viewed 05/20/2011).

Garellek, M., and Keating, P. (**2011**). "The acoustic consequences of phonation and tone interactions in Jalapa Mazatec," Journal of the IPA. **41,** doi:101017/S0025100311000193

Huffman, M. (**1987**). "Measures of phonation type in Hmong," J. Acoust. Soc. Am. **81**, 495-504.

Holt, L. L., and Lotto, A. J. (**2006**). "Cue weighting in auditory categorization:

Implications for first and second language acquisition," J. Acoust. Soc. Am. **119**, 3059-3071.

Huang, T. (**2004**). *Language-specificity in auditory perception of Chinese tones* (Ph.D. dissertation, Ohio State University, Columbus, OH), Chap.2, pp.48.

Iseli, M., Shue, Y.-L. and Alwan, A. (**2007**). "Age, sex, and vowel dependencies of acoustic measures related to the voice source," J. Acoust. Soc. Am. **121**, 2283-2295.

Jiang, J., Alwan, A., Keating P., Auer, E. and Bernstein, L. (**2007**). "Similarity structure in visual speech perception and optical phonetic signals," Perception and Psychophysics **69**, 1070-1083.

Johnson, K. (**2003**). *Acoustic and auditory phonetics* (Blackwell Publishing, Oxford), Chap. 4, pp. 70-73.

Khan, S. D. (**2010**). "Breathy phonation in Gujarati: an acoustic and electroglottographic study," J. Acoust. Soc. Am. **127**, 2021(A).

Kirk, P. L., Ladefoged, J. and Ladefoged P. (**1993**). "Quantifying acoustic properties of modal, breathy, and creaky vowels in Jalapa Mazatec," in *American Indian linguistics and ethnography in honor of Lawrence C. Thompson*, edited by A. Mattina and T. Montler (University of Michigan, Ann Arbor, MI), pp. 435–450.

Kruskal, J. B., and Wish, M. (**1978**). *Multidimensional Scaling* (Sage Publications), pp. 27-28.

Kuang, J. J. (**2011**). *Production and Perception of the Phonation Contrast in Yi* (M.A. thesis, UCLA, Los Angeles, CA), Chap. 2, section 2.5.

Iverson, P., and Kuhl, P. K. (**1995**). "Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling," J. Acoust. Soc. Am.

97, 553–562.

Maddieson, I., and Ladefoged, P. (**1985**). ""Tense" and "lax" in four minority languages of China," J. Phonetics **13**, 433-454.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L.,Solymos, P. M., Stevens H. H. and Wagner, H. (**2011**). "vegan: Community Ecology Package," R package version 1.18-29/r1589. This is available online at <http://R-Forge.R-project.org/projects/vegan/> (last viewed 05/20/2011).

Pierrehumbert, J. (**2003**). "Probabilistic Phonology: Discrimination and Robustness," in *Probability Theory in Linguistics*, edited by R. Bod, J. Hay and S. Jannedy (The MIT Press, Cambridge, MA), pp. 177-228.

Pols, L. C. W., van der Kamp, L. J. Th., Plomp, R. (**1969**). "Perceptual and physical space of vowel sounds," J. Acoust. Soc. Am. **46**, 458–467.

Shepard, R. N. (**1972**). "Psychological representation of speech sounds," in *Human Communication: A Unified View*, edited by E. David and P. Denes (McGraw-Hill, New York), pp. 67-113.

Shi, F., and Zhou, D. (**2005**). "An acoustic study of tense and lax vowels in Southern Yi" (in Chinese), Yuyan Yanjiu **25**, 19-23.

Shue, YL, Keating, P. and Vicenik, C. (**2009**) "VoiceSauce: A program for voice analysis," J. Acoust. Soc. Am. **126**, 2221(A). This is available online at <http://www.ee.ucla.edu/~spapl/voicesauce/> (last viewed 05/20/2011).

Tehrani H. (**2009**). EggWorks: A program for EGG analysis. This is available online at <http://www.linguistics.ucla.edu/faciliti/facilities/physiology/egg.htm> (last viewed 05/20/2011).