THE THEORY OF ADAPTIVE DISPERSION AND ACOUSTIC-PHONETIC PROPERTIES OF CROSS-LANGUAGE LEXICAL-TONE SYSTEMS

Jennifer A. Alexander

Simon Fraser University
jennifer_alexander@sfu.ca

ABSTRACT
This is a cross-language investigation of whether and how lexical-tone-inventory size affects acoustic tone-space size. I extend to tone-systems a model of vowel-system organization, the Theory of Adaptive Dispersion (TAD) (Liljencранt & Lindblom, [1]). I examine five languages with different-sized tone inventories: Cantonese (6 tones), Thai (5), Mandarin (4), Yoruba (3), and Igbo (2). Six native speakers (3 female) of each language produced 18 CV syllables in isolation, with each of his/her language’s tones, 6 times. Tonal F0 in semitones (ST) was measured at 3 equidistant points across the vowel.

Each language’s tone-space was defined in two ways: (1) the F0 difference between its highest and lowest tones; and (2) a two-dimensional, onset F0 x offglide F0, space. Per the TAD, I predicted that languages with larger tone inventories would have larger tone-spaces; this was not supported by either. But the dispersion of tones in (2) supports the TAD hypothesis that sound categories will be generally well-dispersed across the acoustic tone space and will be highly contrastive.

Format: Oral presentation

Keywords: Cross-language, Tone-space, Dispersion

Scientific areas: Speech Perception and Production, Cross-language Studies

1. INTRODUCTION
They are a component of about 42% of the world’s languages (Maddieson, [2]), but lexical-tone systems are under-studied compared to segmental systems. This study aims to increase our understanding of speech-sound inventory organization by extending the Theory of Adaptive Dispersion (TAD) ([1]) to multiple lexical-tone systems. I examine whether and how tone inventory size affects tone-space size.

In lexical-tone languages, pitch-contour and -height contrasts over a tone-bearing unit (TBU) signal word meaning. In Igbo, the syllable /di/ produced with a high-level tone means “husband,” while /di/ means “to exist” when it carries a low-level tone. The languages examined here have different tone inventories: Cantonese has 3 level (high (H), mid (M), low (L)) and 3 contour tones (mid-rising (MR), low-rising (LR), and falling (F)); Thai has 3 level (H, M, L) and 2 contour tones (F, rising (R)); Mandarin has 1 level (H) and 3 contour tones (R, dipping (D), F); Yoruba has 3 level tones (H, M, L); and Igbo has 2 level tones (H, L). Fig. 1 shows pitch tracks of the languages’ tones (F0 in semitones (ST) of citation-form tones in CV syllables by 6 talkers (3 females)).

Figure 1. Pitch tracks of the five languages’ tones.

The main aim of the TAD is to predict the phonetic structure of vowel inventories. A key prediction is that the size of the acoustic vowel space is positively correlated with vowel inventory size. Some studies support this notion. Jongman et al. [3] found that English (11 vowels) and German (14 vowels) have more crowded F1xF2/F3 vowel spaces than Greek (5
vowels). Al-Tamimi & Ferragne [4] reported that the F1xF2 (Bark) vowel spaces of French (11 vowels) > Jordanian Arabic (8 vowels) > Moroccan Arabic (5 vowels). But larger surveys find that vowel space size is not correlated with the vowel inventory size. Livijn [5] surveyed 28 languages and found that only 3 systems with 11+ vowels each had comparatively large (F1xF2, Bark) vowel spaces.

Few studies have quantified an acoustic tone space. Zhao & Jurafsky [6] measured F0 (Hz) of vowels (the proposed vocalic F0 can be perturbed by the preceding consonant; Hombert et al. [7]), F0 values were converted to ST as the ST scale more accurately reflects listeners’ intuitions about intonational equivalence (Nolan [8]). Tone-space dispersion for each talker was the mean Euclidean distance of individual tones from a centroid; the centroid was the mean F0 at each k of a given tone.

Barry & Blamey [9] compared Cantonese citation tones produced by normally-hearing adults and children and cochlear-implanted children. Tones were plotted in an F0 offglide x F0 onset space (Hz). This captured acoustic dimensions that affect listeners’ tone judgments, e.g., mean F0 and F0 direction, endpoint, and slope. The groups of talkers were differentiable in the F0 onset x F0 offglide graph.

This study tests the TAD prediction that the size of the acoustic tone space is positively correlated with tone inventory size. It employs some methods of [6] and [9]. The tone space of each language is defined in two ways: (1) as the difference between its highest and lowest tones’ mean F0s at 3 points along the tonal trajectory (onset, midpoint, and offglide); and (2) in an F0 offglide x F0 onset space.

2. METHOD

2.1. Data collection

2.1.1. Materials

18 CV syllables were used: [ba, bi, bu, da, di, du, ga, gi, gu, la, li, lu, ma, mi, mu, na, ni, nu]. Segments that were identical or similar to those in the languages’ inventories were chosen. Talkers were thus less likely to be distracted by segments when asked to produce the tones.

2.1.2. Talkers

Talkers were 6 (3 female) adult native speakers of each language. Each was literate in both English and his/her native language. Place-of-origin was controlled to minimize dialect variation. Cantonese talkers were from Hong Kong, Thai from Bangkok, Mandarin from Beijing, and Igbo and Yoruba from Nigeria. They left their place of origin at M=21 y.o. All were 18-50 y.o. (M=30.6 y., SD=9.6 y.); due to recruitment difficulties, E. Asians and Nigerians differed in age (E. Asians M=24.4 y.o.; Nigerians M=39.9 y.o.). None reported hearing, speech, or neurological deficits. Language experience was assessed via questionnaire.

2.1.3. Recording procedure

Syllables were presented one at a time via Dell PC and E-PRIME [Schneider et al., 10]. Each was presented in native orthography and Roman letters; with tone numbers, letters, and diacritics; and in context (real words only). Fig. 2 shows a Mandarin trial.

Figure 2. Mandarin trial in tone-production task.

In a quiet room, the talkers read each syllable aloud. Syllables were randomized (Rand) in 3, and sequential (Seq) in 3, of the 6 blocks. Each syllable was presented once per block. All Seq blocks were presented first and all Rand blocks second, or vice-versa. Each block was recorded as a .wav file at 44.1 kHz with a Marantz Professional Solid State Recorder and a Shure WH20XLR Dynamic Headset Microphone.

2.1.4. Data processing

Following [6], the vowel of each syllable was demarcated. The start of vocalic modality marked its onset; vocalic offset was the final glottal pulse. Vocalic F0 (Hz) at 10 equidistant
timepoints $k$ was measured via Praat script; $k_1$ was then excluded. A Perl script converted Hz values to ST with the equation

$$F_{0ST} = \frac{(12 \log(F_{0Hz}/100 \text{ Hz}))}{\log(2)}.$$  

F0 outliers, defined as F0 values more than 2.5 SD from the mean for that tone for each talker, were omitted. Only 0.025% were outliers.

### 2.2. Tone-spaces as extreme-tone F0 differences at onset, midpoint, and offglide

Linear mixed-effects regressions (lmer), fitted using the software package R [see Baayen, 11], were used to investigate whether tone space (TS) size, defined as the F0 difference between the highest (top, $T$) and lowest (bottom, $B$) tones of a language, differs as a function of tone inventory size. Three sets of models compared the F0 distances between the $T$ and $B$ tones.

Lmers incorporate fixed and random variables. Here, Language and Tone were fixed, and Talker and Item random, variables. Item was the Block + Repetition + Syllable (e.g., Rand1Bi for [bi] in the first Rand block).

#### 2.2.1. TSOntset

TSOnset compared F0 differences between the following tones at tonal onset: Cantonese H & L; Thai F & R; Mandarin F & FR; Yoruba H & L; and Igbo H & L. See Fig. 3, below.

The tone x language interaction was significant (p<0.001, $\alpha$=0.0125, corrected). Thus, Yoruba < Igbo < Cantonese < Thai < Mandarin in tone-space size at tone onset.

#### 2.2.2. TSMidpoint

TSMidpoint compared F0 differences between these tones at tonal midpoint: Cantonese H & LR; Thai H & R; Mandarin H & FR; Yoruba H & L; and (e) Igbo H & L (Fig. 4).

The tone x language interaction was significant in comparisons of Thai, Mandarin, and Cantonese (p=0.0001, corrected), but was N.S. otherwise. Thus, Mandarin > Cantonese > Thai = Yoruba = Igbo in tone-space size at midpoint.

#### 2.2.3. TSOffglide

TSOffglide compared F0 differences between the following tones at tonal offglide: Cantonese MR & LR; Thai R & L; Mandarin R & F; Yoruba H & L; and Igbo H & L. See Fig. 5.

The tone x language interaction was significant in comparisons involving Thai only (p<0.01, corrected). Thai > Igbo = Yoruba = Mandarin = Cantonese in tone-space size at tonal offglide.

### 2.3. Tone-spaces as offglide x onset F0 plots

Fig. 6 shows tone offglide F0 x onset F0 spaces. Dashed lines define the tone-space periphery.
Save for the Cantonese LR and LF tones, the languages’ tone categories are dispersed through the tone space. This supports the TAD hypothesis that sound categories will be dispersed and highly contrastive. But these results do not support the TAD prediction that tone-space size is positively correlated with tone-inventory size. Mandarin > Cantonese & Thai, and Cantonese = Thai, in tone-space size.

3. RESULTS AND DISCUSSION

Each of 5 languages’ tone-spaces was defined as (1) the F0 difference between its highest and lowest tones at $k_1$, $k_5$, and $k_9$; and (2) within an onset F0 x offglide F0 space. Per the TAD, I predicted that languages with larger tone inventories would have larger tone-spaces. This was not supported by either (1) or (2). Indeed, the results of (1) indicated that tone-space size was sometimes fixed (e.g., Yoruba and Igbo at $k_5$ and $k_9$). Looking at (1), we might surmise that other acoustic cues (e.g., duration) might be key to tone differences. However, (2) suggests that tone categories are generally dispersed across an onset F0 x offglide F0 space, indicating that F0 trajectory may indeed define tone spaces. Time-variant aspects of tonal contours are omitted in this 2D model; addition of a tonal-midpoint 3rd dimension might disambiguate, e.g., the crowded Cantonese LR and LF tones.

We might consider the possibility that some tone languages’ phonetic structures are qualitatively different than others. Level-tone languages might be better-predicted by a model wherein the tone space is roughly fixed in size until a certain tone-inventory size is reached. This tone-space might be anchored with “point tones”: basic tones such as H and L around which other tones align in the acoustic space (analogous to “point vowels” [i-a-u], around which other vowels fit within the vowel space). More work is needed to address such questions.

4. ACKNOWLEDGMENTS

I thank A. Bradlow, P.C.M. Wong, M. Goldrick, C. Chan, D. Kiefer, and two anonymous reviewers for their advice.

5. REFERENCES