Perception of Syntagmatic Tone Intervals in Ìgbò and Yorùbá

Aaron Carter-Ényì [1], *Quintina Carter-Ényì* [2]

[1]Department of Music, Morehouse College, USA
[2] Faculty of Humanities, Lagos State University, Nigeria

carterenyi@gmail.com

Abstract

This paper revisits the syntagmatic model of tone to see what perception has to say. We asked the question: *Is word identification possible based on syntagmatic (relative) pitch information?* Listeners were asked to identify words from a minimal pair without the context necessary for making paradigmatic (associative) judgments. 1409 Nigerian university students, staff and faculty responded to one of two parallel studies, one for Ìgbò and one for Yorùbá. The results suggest that word identification is closely tied to STI direction (+ 0 –) in Niger-Congo A tone languages. There is also evidence for a cross-language minimum STI distance of +1 or –2 semitones to leave a tone level and enter an adjacent tone level. Word identification based on STI magnitude alone (e.g. HM v. HL) is weaker. Paradigmatic features may be needed to differentiate homophones with HM and HL tone. This is consistent with recent findings by other researchers outside of the context of tone languages that suggest perceiving direction is automatic while judging magnitude is a higher level process [16]. In general, judging unidimensional magnitude is hard [17].

Index Terms: tone, perception, Niger-Congo A languages, Nigeria, Igbo, Yoruba, behavioral study, speech synthesis

1. Introduction

Long dormant, the debate over using syntagmatic or paradigmatic models of tone systems has been reinvigorated in recent years (see [1], [2], [3] and [4]). According to Dilley, “a *syntagmatic* tone interval [STI] relates two sequentially-ordered tones” [1]. This means that only adjacent pairs of syllables have tonemes, and is called an initializing system because the single syllable is unspecified [2]. This is the classical approach proposed by Jakobson, Halle, and Fant, but since the 1970s, this perspective has been displaced by autosegmental theory ([3] and [4]). In autosegmental phonology, a *paradigmatic* tone interval "relates a tone [to] a speaker-specific referent level” [1]. The referent is often conceived as a frequency band within a speaker’s range. Because different people have different ranges the tone levels normalize to speaker range [2]. In this model, each syllable has a toneme based on where it falls within the frequency-banded range. Building on work by Akinlabi, Laniran and Clements ([5] and [6]), in previous work we have shown that if there are frequency bands, they are not static within each phrase or over larger time spans [7]. Lexical tone and paralinguistic intonation coexist within the fundamental frequency (f0) domain of Niger-Congo tone language speech, utilizing different temporal scales.

Because of the faults with a paradigmatic model in terms of phonetic analysis, this paper revisits the syntagmatic model to see what perception has to say. In a true experiment, we asked the question: *Is word identification possible based only on Syntagmatic Tone Interval (STI)?* Listeners were asked to select from two images representing a minimal pair in response to a synthetic audio stimulus, without the context necessary for making associative (paradigmatic) judgments. Experimental work on Yorùbá has not been conducted since Gandour and Harshman in 1978 [8] and, to our knowledge, there is no previous work on the perception of Ìgbò tones. Over 1400 Nigerian university students, staff and faculty responded to one of two parallel studies, one for Ìgbò and one for Yorùbá.

2. Methods

All stimuli were based on disyllables found in scholarly dictionaries ([9] and [10]). These words were treated as two audio segments (e.g. /i.gba/, /m.ma/) each with one tone. This differs from Bamgbose's approach to Yorùbá phonology, which interprets L.H disyllable words with rising contour on the second syllable, as three tones: L.LH [11]. Sloped or contoured tones are not contrastive in Ìgbò [12] and because we are interested in cross-language perceptual schema, stimuli for both languages were prepared using the same process.

2.1. Stimuli

2.1.1. Minimal Pairs

All possible relationships between tone-varied disyllable minimal pairs (differing only in the tone level of one syllable) are illustrated by connecting lines in Figure 1. The inventory of tone combinations for disyllables in Ìgbò (6 types) is considerably smaller than the inventory for Yorùbá (18 types). Ìgbò is a two-level language in which down-stepped high tone maybe a distinct toneme (i.e. HH and H!H may not be the same word). Yorùbá has three tone levels.

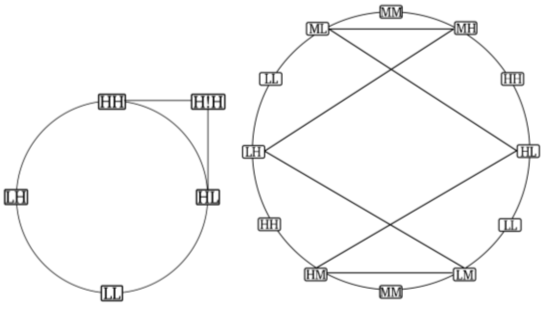


Figure 1: *Tone-varied disyllable minimal pairs in Ìgbò (6 types, left) and Yorùbá (18 types, right).*

All possible minimal pair types were included in the Igbo study, but only nine types (out of 18) were included in the Yoruba list because not all are common in the lexicon (unique tone contrasts are italicized in Table 1). The types can be divided into two broader categories: STI direction (+, 0, –) and STI magnitude, the latter reflecting the scalar

http://dx.doi.org/10.21437/TAL.2016-23

distance between an adjacent (e.g. HM) or a non-adjacent (HL) tone level. Both types are well-represented in the experimental stimuli.

Igbo Homophones	Tone Contrast	Yoruba Homophones	Tone Contrast
a.fa	*HH-HL*	a.ra	*MH-MM*
a.ka	HH-HL	a.ro	*MH-ML*
a.kwa	HH-HL	ba.ta	*LH-LL*
a.kwa	*HH-LH*	e.ru	MH-ML
a.lu	*HH-H!H*	i.la	MH-ML
a.wo	HH-HL	i.re	MH-MM
e.gbe	HH-HL	i.she	*MH-LH*
e.nyi	HH-HL	jo.ko	*HM-MM*
e.ze	*H!H-HL*	mi.mọ	*HH-HM*
i.gwe	H!H-HL	mi.mọ	*HM-HL*
i.ke	HH-HL	mi.mu	HH-HM
i.si	HH-HL	o.do	MH-ML
m.ma	H!H-HL	o.gun	*MH-LH*
o.du	HH-H!H	o.ko	MH-MM
o.gu	HL-LL	o.kun	MM-ML
o.ha	*LH-LL*	o.ri	MH-LH
o.ke	HH-LH	ọ.kọ	*MM-ML*
o.kpa	*HL-LL*	pi.pa	HH-HM
o.nya	HH-HL	shi.shu	HH-HM
u.nyi	HH-HL	si.sun	HM-HL

Table 1: *Minimal pairs used in the study*

2.1.2. Synthetic Experimental Stimuli

A female and a male speaker recited the list of homophones with both tonal variations and on neutral tone. For each minimal pair, one of two syllables was modulated at semitone increments from one pitch target (based on natural speech) to another using *Melodyne* software. Celemony's *Melodyne* has the ability to "modify the pitch center" of segments without altering timing or formants. An example is ákwá (crying) modulated to ákwà (cloth) (shown in Figures 2 and 3). All modulations of a homophone (both male and female voice) formed the stimuli set for that minimal pair. In all, 452 experimental stimuli were created for Ìgbò and 405 for Yorùbá. As noted before, the presence of different pitch (f0) trajectories in Yorùbá was not controlled in this study, a uniform method was used for both languages. Signal processing tools implemented in MATLAB ([13] and [14]) were used to verify that only pitch varied across stimuli in a set, and there was little or no effect on vowel quality or other features in the re-synthesis.

2.2. GUI Design

2.2.1. Task and Modules

The task was the same in the primer (5 iterations) and experimental module (20 iterations): (1) hear an audio stimulus (single word), (2) see two images representing a tone-varied minimal pair, (3) hear the audio stimulus again, and (4) make a selection from the two images. The audio stimulus was natural for the primer and synthetic during the experimental module. The study used a forced-choice response format, participants chose from one of two images (representing words of the minimal pair) with no alternative.Each participant heard only one version of each homophone (unless the homophone composed multiple minimal pairs, see "a.kwa"). Many participants were needed to gather sufficient data.

Figure 2: *images for /akwa/ (HH-HL) minimal pair*

2.2.2. Randomization

To avoid inter-stimulus effects and image-side bias, a complex randomization was used. The following parameters were randomized: order of minimal pairs ($2.43 * 10^{18}$); image side; selection of stimulus from each set (7.16 * 2).

2.3. Participants

A pilot study was completed at Lagos State University in May of 2014. From May through July the experiment was conducted in locations throughout southern Nigeria, including University of Nigeria-Nsukka, Imo State University, University of Port Harcourt, University of Ilọrin and University of Lagos. Two to four laptops with the GUI were set-up and monitored in public areas and volunteer participation was solicited. Participants received no compensation but generally reported enjoying the task, which took 2–3 minutes. Often, participants returned with friends who also wished to perform the task. When groups of friends participated in succession, there was a tendency to crowd around the laptop and to instruct each other what the "correct" response was based on the images, not having heard the audio and not knowing that each participant hears a unique set of stimuli. In these situations we asked bystanders to let the current participant complete the task on her/his own and did not exclude data. Reponses from the primer (which did have "correct" values) were used to cull data from participants that either did not understand the task or were not fluent. Our hope is to develop the GUI model into a tone language learning software. Based on verbal feedback received during the experiment, this is a promising and useful application.

3. Results and Discussion

For the /a.kwa/ HH-HL stimuli set, we found an inflection point of 4.2 semitones. This means that the fifth version in Figure 3 (each version is separated by dotted red lines) is ambiguous, those to the left were interpreted as *ákwá* (crying) and those to the right *ákwà* (cloth).

Figure 3: *Fundamental frequency plot (using YIN [12]) of the entire stimuli set for /akwa/ modulated from HH to HL*

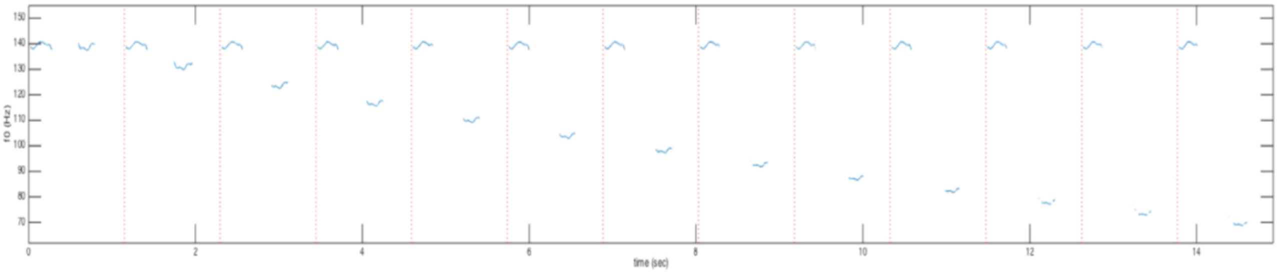
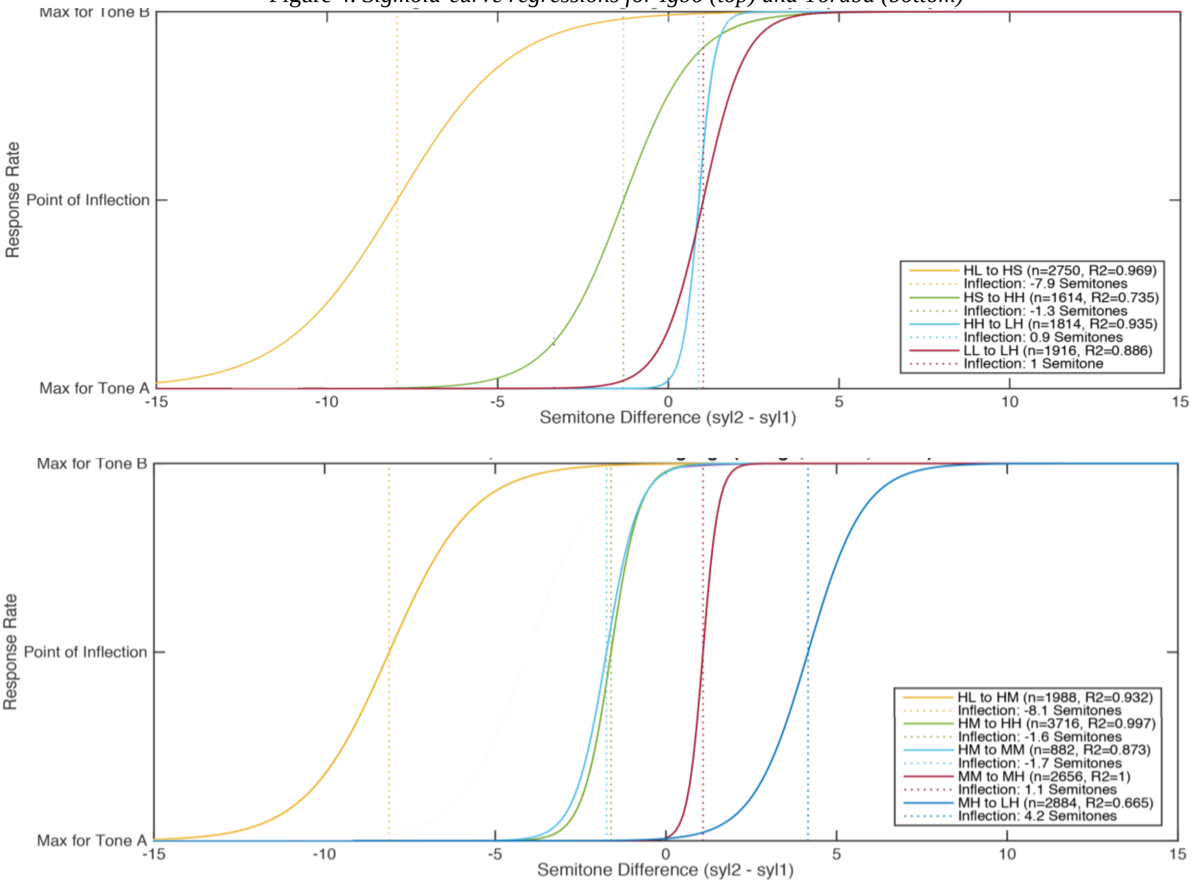


Figure 4: *Sigmoid-curve regressions for Ìgbò (top) and Yorùbá (bottom)*



The p-values for individual stimuli are significant not because of huge numbers of responses, but a modest number of responses for each stimulus (between 10 and 40) and a very large effect size, which we measured through a coefficient of determination (r^2). The *R-squared* value was 0.988 for the sigmoid-curve fitting to responses for /a.kwa/ HH-HL. Other stimuli sets produced similarly significant results, with the exception of /o.gun/ in Yorùbá and /o.ha/ in Ìgbò.

Sigmoid curves were also fit to aggregate data for minimal pairs of the same type and these results are presented in Figure 4. The HH-HL minimal pair (as in /a.kwa/ in Figure 2 and 3) is the most common in Ìgbò and provides a very clear contrast between homophones. The difference in STI direction (0 v. –) is reinforced with a decisive drop from the high (neutral) tone (*H*) past what could be interpreted as a down-step (*!H*) to the low-tone level (*L*). This is a stronger contrast than found in HH-H!H or H!H-HL minimal pairs because utilizes both contrastive direction and magnitude. Because HH-HL crosses an adjacent tone-level to a non-adjacent tone-level it is excluded in Figure 4, which only displays results for minimal pairs that are contrasted by a tone alternation between adjacent levels. In Figure 4, the sigmoid curves for distinct syntagmatic direction (e.g. HM to HH), have a sharp slope, but the sigmoid curves for distinct syntagmatic magnitude (e.g. HL to HM), do not. This indicates the threshold between the perceptual categories HH and HM is quite crisp, while the threshold between HM and HL is fuzzier. Generally, minimal pairs that differ in direction are more distinct that minimal pairs that differ in magnitude of change in the same direction.

The most interesting result is the striking similarity between the location of the inflection points for Ìgbò (on the top) and Yorùbá (on the bottom) in Figure 4. Ìgbò is a terraced two-tone language with down-step, while Yorùbá has three discrete tone levels. Hence, there is no corresponding inflection point in Ìgbò for Yorùbá's MH-LH. A worthy post-hoc criticism of the experimental design by D. Robert Ladd is that it discounts the effect of segment slope (contoured tones) in Yorùbá. This is a weakness with regard to understanding Yorùbá tonology specifically. However, the approach of synthesizing stimuli with the same process for both languages strengthened the findings of cross-linguistic perceptual schema. Despite very different accepted tonological models and very distinct phonetic implementation of tones by Ìgbò and Yorùbá speakers, the findings indicate similar perceptual categories of syntagmatic tone intervals (STIs) exist in both Ìgbò and Yorùbá. Further details of findings and more discussion is available in [15].

4. Conclusions

In both Ìgbò and Yorùbá, word identification is closely tied to the direction between adjacent tones (+, 0, –). There is also evidence for a cross-language minimum magnitude of change of +2 or –3 semitones to confidently leave a tone level, forming an asymmetrical which allows for some fluctuation in pitch while remaining at the same tone level. Because this window is larger than the range of f0 perturbations attributed to intrinsic pitch of vowel (IPV) effects, it follows that IPV is not constrained in Niger-Congo A tone languages. Figure 5 displays directional contrasts in green and magnitude-only contrasts in yellow. Results showed the latter is a weaker form of contrast.

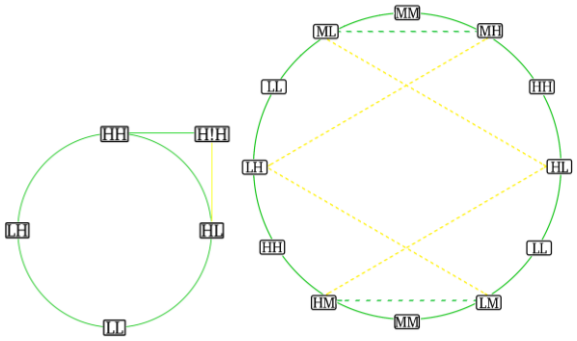


Figure 5: *Tone-varied minimal pairs in Ìgbò (6 types, left) and Yorùbá (18 types, right).*

Weaker results for distinguishing homophones that only differ in magnitude, not direction, is consistent with the findings of Demany *et al* [16]. In a psycho-acoustic study (not specific to tone language perception), they found the ability to judge change of direction is nearly automatic and is an ability most people have, however, judging magnitude of pitch change has a greater cognitive burden and is generally not as accurate. In general, making perceptual judgments of magnitude is difficult in all domains, not just pitch [17]. Ìgbò and Yorùbá speakers can judge magnitude but it may take longer and the threshold between small and large magnitude perceptual categories varies from person to person. Some paradigmatic cues, in pitch (f0) trajectory or non-formant timbre (such as aperiodicity in the low range) may help listeners to distinguish homophones that only differ in magnitude of change, such as HM from HL. Additionally, contour theory (in music), which compares relative pitch height beyond immediate adjacency, may provide an alternative to a strictly paradigmatic or syntagmatic model of tone perception.

5. Acknowledgements

Many people deserve thanks for their role in an undertaking spanning several years and two continents, notably: mentors and peers at Ohio State University, including Nathaniel Condit-Schultz, Johanna Devaney, David Huron, David Odden, and Shari Speer; linguists at-large who have offered insight over the years, namely Roger Blench, D. Robert Ladd and Will Leben; Michael MacAskill for helping figuring out the complex randomization for the GUI using PsychoPy; and most of all our colleagues in Nigeria, including David Aina, Sam Chukwu, Austin Emielu, Anthony Mereni and Anthony Okoro.

6. References

[1] Dilley, L., "The Phonetics and Phonology of Tonal Systems," Ph.D., Speech and Hearing Bioscience and Technology, Massachusetts Institute of Technology, Cambridge, 2005.

[2] Ladd, D. R., *Intonational phonology*. New York: Cambridge University Press, 2008.

[3] Leben, W. R., "Rethinking autosegmental phonology," *Selected Proceedings of ACAL 35*, pp. 1-9, 2006.

[4] Clements, N., *et al.*, "Do we need tone features?," in *Tones and Features*, E. Hume *et al*., Eds.: De Gruyter Mouton, 2011, pp. 3-24.

[5] Akinlabi, A. and Laniran, Y., "Tone and intonation in Yoruba declarative sentences," in *19th Annual Conference on African Linguistics, Boston University*, 1988.

[6] Laniran, Y. O., and Clements, G. N., "Downstep and High Raising: Interacting Factors in Yoruba Tone Production," *Journal of Phonetics*, vol. 31, pp. 203-50, 2003.

[7] Carter-Cohn, A., "The multiple roles of pitch in Yorùbá: Evidence from Yorùbá sermons," in *WALC2013: West African Languages Congress*, University of Ibadan, Nigeria, 2013.

[8] Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and speech*, *21*(1), 1-33.

[9] Williamson, K., *Igbo English Dictionary*. Benin City: Ethiope Publishing Corporation, 1972.

[10] Abraham, R. C., "Dictionary of Modern Yoruba," in *Dictionary of Modern Yoruba*. London: Hodder and Stoughton, 1962.

[11] Bamgbose, A., *A Grammar of Yoruba*. New York: Cambridge University Press, 2010.

[12] Clark, M. M., *The tonal system of Igbo*. Dordrecht, Holland; Providence, RI, U.S.A.: Foris Publications, 1990.

[13] de Cheveigne, A. and Kawahara, H., "YIN, a fundamental frequency estimator for speech and music," *Journal of the Acoustical Society of America*, vol. 111, pp. 1917-30, 2002.

[14] Slaney, M., "Auditory Toolbox, Version 2," Interval Research Corporation, 1998.

[15] Carter-Enyi, A., "Contour Levels: An Abstraction of Pitch Space based on African Tone Systems," PhD, School of Music, Ohio State University, Columbus, 2016.

[16] Demany, L., *et al.*, "Implicit versus explicit frequency comparisons: two mechanisms of auditory change detection," *Journal of experimental psychology. Human perception and performance*, vol. 37, pp. 597-605, 2011.

[17] Donkin, C., *et al.*, "Why is accurately labelling simple magnitudes so hard? A past, present and future look at simple perceptual judgment," *The Oxford Handbook of Computational and Mathematical Psychology*, p. 121, 2015.