ON THE PHONETIC INTERPRETATION OF THE YORUBA TONAL SYSTEM

Akin Akinlabi
Rutgers University

Mark Liberman
University of Pennsylvania

ABSTRACT

Yoruba is a tone language, with three lexically contrastive levels H(igh), M(id), and L(low). Various phonological and phonetic properties of these tones are explained in terms of the view that M is phonologically unspecified, and that adjacent H and L join to form “derived pitch accents.”

BACKGROUND

Yoruba has three phonemically distinctive tones—H(igh), M(id), and L(low). H occurs in word-initial position only in (marked) consonant-initial words, which reveal an implicit initial vowel when preceded by another word in a genitive construction. Most nouns and adjectives start with a vowel, which is genitive construction. Most nouns and adjectives start with a vowel, which is

Thus Yoruba presents itself as a fundamentally tonal language, in which tonal features have a lexical distribution about as free as that of any other phonological features.

There are several reasons to believe that Yoruba M(id) tone is underlyingly just the absence of tonal features ([1], [12]). We will mention just one of these: tones L and H remain when their lexically-associated vowels delete, but M does not. Thus in the case of a verb followed by a vowel-initial object, one of the two adjacent vowels obligatorily deletes. The tonal consequences are simple to calculate if we assume that M is just the name for lack of tone—then all “real” tones remain stable under vowel deletion.

(1) a. wa (H) + ẹkọ (LH) \(\rightarrow\) look (for) education
   \(\rightarrow\) education

b. mu (H) + iwe (LH) \(\rightarrow\) take book
   \(\rightarrow\) a book

c. ṣe (M) + aje (LH) \(\rightarrow\) resemble a witch
   \(\rightarrow\) resemble a witch

d. sin (M) + oka (LH) \(\rightarrow\) bury dead (body)
   \(\rightarrow\) bury the dead

SOURCE OF DATA

There have been several earlier instrumental studies of Yoruba tone (e.g. [3], [14], [6], [4], [5]). In order to apply to Yoruba the scaling technique previously applied to English in [8] and to Igbo in [9], we devised 78 Yoruba phrases exhibiting an appropriate range of tone sequences, with texts that avoid consonant immediately likely to interrupt or strongly affect F0.

These phrases were read a total of 18 times each, six in each of three pitch ranges, by three native Yoruba speakers. Pitch range was varied by instructing the speaker, in each utterance, to address one of three (imaginary) interlocutors, placed immediately adjacent to the speaker, across the room, or out the door and down the hall. Within each recording session, the list of phrase/pitch-range combinations was randomized.

In this paper, space does not permit us to report fully on this experiment or to discuss its relation to previous work. Instead, we will focus on three key points and a general conclusion that is suggested by them.

TIMING OF TONE GLIDES

As was first noted in [15], Yoruba HL and LH sequences postpone the falling or rising F0 glide to the second syllable. By comparison, the transitional glide for sequences involving M (HM, ML, MH, LM) occurs significantly earlier.

Figure 1 shows the F0 tracks for the initial LH sequences in the six narrow-pitch-range repetitions of ṣe (LH) “I am Orunlami”1 as produced by one male speaker.

Table 1 exemplifies the raising of H before L, by showing the means and standard errors of peak F0 measurements in each of the three pitch range conditions, for one of three subjects.

Table 1

<table>
<thead>
<tr>
<th>Phrase</th>
<th>narrow</th>
<th>middle</th>
<th>wide</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean H/LM</td>
<td>125</td>
<td>154</td>
<td>250</td>
</tr>
<tr>
<td>standard error</td>
<td>3.2</td>
<td>8.8</td>
<td>2.6</td>
</tr>
<tr>
<td>mean H/LM</td>
<td>143</td>
<td>182</td>
<td>287</td>
</tr>
<tr>
<td>standard error</td>
<td>3.3</td>
<td>7.0</td>
<td>2.9</td>
</tr>
</tbody>
</table>

This postponement of the HL and LH glides in Yoruba is a well-established fact. Our contribution is to compare it systematically to HM, ML, MH, LM glides across pitch ranges, and to verify the stability of the difference under this manipulation.

TONE DISSIMILATION

In Yoruba, H is raised before L (as opposed to before M), and L is lowered before H (as opposed to before M). The raising of H before L has been cited in [2], [4], [5]. The lowering of L before H is (we believe) a new observation.

In Yoruba, H is raised before L (as opposed to before M), and L is lowered before H (as opposed to before M). The raising of H before L has been cited in [2], [4], [5]. The lowering of L before H is (we believe) a new observation.

Table 1 exemplifies the raising of H before L, by showing the means and standard errors of peak F0 measurements in each of the three pitch range conditions, for one of three subjects.

1All examples in this paper use the standard Yoruba tone marking convention, according to which H is marked with an acute accent, L with a grave accent, and M with no accent. Vowels with a dot underneath are non-ATR.
lower than the first. This is consistent with the general expectation of downshift in such sequences; dissimilatory lowering applies to both L tones in this case. When the first L is followed by an H, while the second L is followed by an M, the expected downshift effect is almost completely nullified. This is because the first L is lowered because it is in an LH sequence, while the second L does not experience this effect. Thus dissimilatory lowering of the first L, and downshift lowering of the second L, leave them at about the same level.

[Graph showing Exemplification of L-before-H lowering]

**Figure 3**

**DOWNSHIFT**

Since [16] and [13] it has been understood that the tendency of pitch to fall in the course of phrases in tone language like Yoruba is not a sort of phrasal wave on which tonal ripples ride, but rather is connected specifically with alternating high and low tones. Sequences of like tones, especially H tone sequences, remain more or less level.

Since [2], it has been known that in Yoruba, this downshift does not extend to sequences in which H or L alternate with M (HMHM...or MLML...). At least, the amount of downshift is much lower in these latter cases.

For a quantitative picture of the difference between the amount of lowering in HLHL vs. HMHM or MLML, see Figure 4. Here we show the relationship between adjacent F0 maxima in the sequences HLHL (plotted with squares), HMHM (plotted with pluses) and MLML (plotted with triangles). The x-axis gives the average height of the two F0 maxima, while the y-axis shows the difference between them. In the case of HLHL, this difference is about 15-35 Hz., showing a healthy amount of downshift, the tonally-conditioned effect that has been called "catathesis." In the case of HMHM and MLML, the difference is about 5-15 Hz., perhaps reflecting the more general downshift sometimes distinguished as "declination."

Japanese accent is interpreted (e.g. by [10] as a lexically-specified H sequence that functions as a unit. Catathesis is triggered only by accents in Japanese according to [11]. Each accentless 'minor phrase' has an (ungrouped) H and L tone pair, but accentless sequences rise and fall with only a small amount of downshift: Japanese accentual H is higher than non-accentual H ("accentual boost") according to [7], even though Japanese accent is not stress-like, does not cause greater segment durations, and is not considered a strong position for alignment with music.

**REFERENCES**


