On the phonetic interpretation of Yoruba tone

Akin Akinlabi
Rutgers University

Mark Liberman
University of Pennsylvania

The interpretation of Yoruba tone in terms of fundamental frequency of the voice (F0) offers unexpected parallels with accentual languages, and suggests a revised account of some aspects of Yoruba tonal phonology as well.

1 Phonological Background

1.1 Lexical distribution of Yoruba tone

Yoruba has three phonemically distinctive tones—H(igh), M(id), L(ow)—with fairly free distribution.

<table>
<thead>
<tr>
<th>ra H</th>
<th>ra M</th>
<th>ra L</th>
</tr>
</thead>
<tbody>
<tr>
<td>“to disappear”</td>
<td>“to rub”</td>
<td>“to buy”</td>
</tr>
<tr>
<td>oko MH</td>
<td>oko MM</td>
<td>oko ML</td>
</tr>
<tr>
<td>“hoe”</td>
<td>“husband”</td>
<td>“vehicle”</td>
</tr>
<tr>
<td>ilu LH</td>
<td>ilu LM</td>
<td>ilu LL</td>
</tr>
<tr>
<td>“town”</td>
<td>“opener”</td>
<td>“drum”</td>
</tr>
<tr>
<td>pako HH</td>
<td>kese HM</td>
<td>pako HL</td>
</tr>
<tr>
<td>“plank”</td>
<td><em>mythological place-name</em></td>
<td>“chewing stick”</td>
</tr>
</tbody>
</table>

N.B. Vowel-initial nouns (the normal case) must begin with I. or M.

1.2 Mid underlyingly unspecified in Yoruba

There are several reasons to believe that Yoruba “mid” tone is underlyingly just the absence of tonal features (Akinlabi 1985, Pulleyblank 1986).

A. H and L are “stable” under vowel deletion, but M is not.

I. & H generally remain and re-attach when their lexically-associated vowels delete; Lexical M does not. Thus vowel deletion processes treat M as if it were not there.

The examples below arise in the very typical case of a verb followed by its object. Yoruba verbs are all vowel-final, and nouns are usually vowel-initial. One of the two adjacent vowels obligatorily deletes. It is difficult to predict which vowel will delete, but one of them must (Bangbolo 1966, Oyelaran 1970, Akinlabi and Oyebade 1987, Pulleyblank 1988).
In some cases the tone of the deleted vowel remains, and may appear alone or in a glide on one of the remaining vowels; in other cases the tone of the deleted vowel also seems to disappear. The output tone sequence in the following typical examples is straightforward to calculate if M is just the name for the lack of tone. The segmental alignment of the resulting tone sequence requires additional discussion.

\[(2)\]
\[
a. \text{wa (H) } + \text{ọgọ (IH) } \Rightarrow \text{węgọ (H LH)}
\]
\[\text{look (for) education} \Rightarrow \text{look for education}\]

b. \text{mu (H) } + \text{iwe (IH) } \Rightarrow \text{muwe (H LH)}
\[\text{take book} \Rightarrow \text{take a book}\]

c. \text{wa (H) } + \text{ọnọ (IL) } \Rightarrow \text{wonọ (H L)}
\[\text{look (for) way} \Rightarrow \text{look for a way}\]

d. \text{wa (H) } + \text{ọmọ (IL) } \Rightarrow \text{wormọ (H L)}
\[\text{look (for) knowledge} \Rightarrow \text{look for knowledge}\]

e. \text{ji (H) } + \text{ọbẹ (IM)} \Rightarrow \text{jọbẹ (HL M)}
\[\text{steal knife} \Rightarrow \text{steal a knife}\]

f. \text{eẹ (H) } + \text{iwo (IM) } \Rightarrow \text{ẹewọ (HL M)}
\[\text{want horn} \Rightarrow \text{want a horn}\]

g. \text{wa (H) } + \text{ọwọ (MH) } \Rightarrow \text{wọwọ (HH)}
\[\text{look (for) money} \Rightarrow \text{look for money}\]

h. \text{wa (H) } + \text{ile (MH) } \Rightarrow \text{wale (HH)}
\[\text{look (for) house} \Rightarrow \text{look for a house}\]

i. \text{jọ (M) } + \text{aịẹ (IH) } \Rightarrow \text{jạẹ (IH)}
\[\text{resemble witch} \Rightarrow \text{resemble a witch}\]

j. \text{sin (M) } + \text{oku (IH) } \Rightarrow \text{sinku (LH)}
\[\text{bury dead (body) } \Rightarrow \text{bury the dead}\]

N.B. “sin” is Yoruba orthography for [sĩ].

Since V-initial nouns cannot start with H tone, no examples of the form X+HX can arise. When a L-tone verb precedes its object, the L tone always deletes even if the vowel is preserved, so the case L+XX offers no evidence in this matter (Akinlabi 1985, Pulleyblank 1986).

**B. Tone glide formation usually acts as if M were not a tone**

Adjacent H and L tones always spread (rightwards) onto each other to create contours; M does not spread onto any adjacent tone.

\[(3)\]
\[
a. \text{ala (IH) } \Rightarrow \text{ala (L LH)} \text{“dream”}\]

b. \text{rara (HL) } \Rightarrow \text{rara (H HL)} \text{“elegy”}\]

c. \text{ole (ML) } \text{“thief”}\]

d. \text{ile (MH) } \text{“house”}\]

In monomorphemic words with initial M, a following H or L can optionally spread backwards onto the first syllable. In this case the initial M is completely supplanted, and the result is homophonous with an underlying HH or LI sequence.
C. Epenthetic vowels are Mid.

Epenthetic vowels appear in front of consonant-initial words in certain cases, for instance when they are in second position in a genitive construction. Such epenthetic vowels assimilate in quality to the vowel that precedes, but emerge as Mid in tone.

2 Summary of our conclusions

We begin with some observations about the phonetic interpretation of Yoruba tone. Many of these have been in the literature for up to 40 years, others are new. Details will be sketched later; the results may be summarized broadly as follows:

1. In HL and LH sequences, a tonal glide is formed on the second syllable, while this does not happen in HM, MH, LM or ML sequences (Ward 1952);

2. Downdrift occurs in HLHL sequences but not in HMHM or MLML (LaVelle 1974);

3. H is raised before L as compared to before M (Akinlabi and Laniran 1987, Connell and Ladd 1990, Laniran 1992); L is lowered before H as compared to before M.

We have verified and quantified these points in a series of phonetic experiments in which we examine the scaling of F0 relationships under variation in pitch range. Data of this kind permit us to model the alignment and scaling of specific features of F0 contours as a function of both phonological variables (such as lexical tone sequence) and performance variables (such as pitch range).

Our results suggest the following conclusions about the phonological surface of Yoruba tone and its phonetic interpretation:

1. Adjacent H and L tones are grouped into phonological units for purposes of F0 alignment and scaling: they become like “derived pitch accents.”

2. M is phonetically as well as phonologically unspecified.

3. Tones associated with multiple tone-bearing-units have just one F0 target; this is a sort of “phonetic OCP.”
Our point (1) is consistent with the phonology and phonetics of pitch accent as treated in Poser 1985, Pierrehumbert and Beckman 1989, Bamba 1991. However, it is a novelty to suggest accent-like phenomena arising at the phonetic level in Yoruba, where the underlying phonology of tones is not at all accentual.

For reasons of time, we will concentrate on the evidence for point (1), simply assuming (2) and (3). However, we will briefly mention (so as to set aside) some ways in which points (2) and (3) are apparently inconsistent with the literature.

Point 2. is contra the argument in Akinlabi 1985 that M is underlyingly underspecified but cyclically filled in, in order to account for M-H glides at the end of “derived” but not simple M-final subject NP’s. We cover this in Appendix 1.

Points (2) and (3) are contra Pierrehumbert and Beckman 1989 on F0 interpolation in Japanese. We discuss this briefly in Appendix 2.

Point 3. is contra Laniran 1992 on tone interpretation in Yoruba, who postulates several F0 targets for a single stretch of like tones in certain cases.

3 Phonetic sketch of Yoruba tone

3.1 Phonetic OCP

See Figure 1, a pitch track for the sentence ó wálníròrun “He drove Alani to heaven.”

3.2 Tone glide formation

All tonal sequences normally require an F0 glide to get from one target to another, in Yoruba as in other languages.

However, in a Yoruba LH sequence, most of the rise takes place on the second syllable, while in a LM or MH sequence, the rise is more evenly split between the syllables. This is true even if the overall amount of rise is comparable, as it is when we compare a LH sequence in a narrow pitch range to LM or MH sequences in a wide pitch range.

See Figures 7, 8, 9.

Falling sequences work similarly: in a HL sequence, most of the fall takes place in the second syllable, while in HM or ML, the fall is split between the two syllables.

See Figures 4, 5, 6.
3.3 Tone dissimilation

As Table 1 shows, the value of H is significantly higher when L follows compared to when M follows.

<table>
<thead>
<tr>
<th></th>
<th>narrow pitch range</th>
<th>middle pitch range</th>
<th>wide pitch range</th>
</tr>
</thead>
<tbody>
<tr>
<td>H in HML (mean)</td>
<td>125</td>
<td>154</td>
<td>250</td>
</tr>
<tr>
<td>H in HML (s.e.)</td>
<td>3.2</td>
<td>8.8</td>
<td>2.6</td>
</tr>
<tr>
<td>H in HLM (mean)</td>
<td>143</td>
<td>182</td>
<td>287</td>
</tr>
<tr>
<td>H in HLM (s.e.)</td>
<td>3.3</td>
<td>7.0</td>
<td>2.9</td>
</tr>
</tbody>
</table>

For evidence of the lowering of L before H as opposed to before M, see Figure 2.

3.4 Downdrift

Since Welmers 1959 and Stewart 1971, it has been understood that the tendency of pitch to fall in the course of phrases in tone language like Yoruba 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.

At least since Akinlabi and Laniran 1987, it has been known that in Yoruba, downdrift does not extend to sequences in which H or L alternate with M (HMHM... or MLML...). At least, the amount of downtrend 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 3.

4 Discussion

In Ida Ward’s 1952 *An Introduction to the Yoruba Language*, she remarks (p. 34) that “The juxtaposition of high and low tones, either high-low or low-high, needs some comment.” Citing an example of the form HL, she observes that there is a fall on the second, syllable that is “heard as a more or less deliberate glide,” and warns that “Unless the English speaker makes it, he is apt to give the impression of gliding down on the first syllable ..., which does not satisfy the Yoruba.”

In a footnote, she suggest that “The best way for the Englishman to acquire this glide is to try to put a slight stress on the low syllable; it is not actually more heavily stressed, but the habit of using a fall in a stressed syllable in English will help him to reach the correct tone here.”

This seems to be both the earliest observation in the scholarly literature of the Yoruba tone glide formation phenomenon, and also a sort of preconfiguration of our suggest that what is involved is a sort of derived pitch accent.

See 10 for the pitch track of the sentence òrùnlámì lèmi “I am Orunlami.” To an native English ear, this does sound as if there were pitch accents (of the early fall H+L* type) on the L-toned syllables la and le.
Of course no such thing is true, from the point of view of the Yoruba lexicon. Our suggestion, however, is that at the phonological surface, each of these syllables is associated with a (HL) unit that functions phonetically much like a pitch accent.

Comparing Yoruba to Japanese, the canonical pitch accent language, we can list some striking similarities:

1. Japanese accent is interpreted (Pierrehumbert and Beckman 1989) as a metrical HL group that functions as a unit (including at the level of lexical specification). Yoruba H and L are independently specified in the lexicon, but appear to join together when they happen to be adjacent, and to function as a unit for purposes of F0 alignment and scaling.

2. *Catathesis* is triggered only by accents in Japanese (Poser 1985); each accentless “minor phrase” has an (ungrouped) H and L tone pair, but accentless sequences rise and fall with only a small amount of declination. Yoruba *downdrift* (= catathesis) occurs strongly when H and L are adjacent; HMH or MLM sequences show only a very small amount of declination if any; HMLMH sequences show an intermediate amount.

3. Accentual H’s are higher than non-accentual H’s (“accentual boost,” Kubozono 1993), 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. In Yoruba, H is considerably raised when it occurs before L as opposed to before M.

There are also important differences: pitch accents in Japanese are specified in the lexicon, rather than arising by combination of underlying separate tones; in Japanese the H of an accent aligns with the end of the accented syllable, and the following L may not be realized if that syllable is final; in Yoruba, we must apparently also postulate a special treatment of LH sequences, whereas all Japanese pitch accents are of the HL type.

There are also similarities to the characteristics of tonal accent noted for the Manding language Mawukakan (Bamba 1991): accentual H is raised compared to non-accentual H; downdrift is only caused by accentual HL, not other (non-metrically joined) H L sequences.

5 Appendix1: Yoruba Mid specified on the surface?

Akinlabi 1985 argued that Yoruba mid tones are underlingly unspecified, but become specified at the end of the phonology. We sketch this argument below, along with our re-interpretation of the facts.

Ordinarily, when a Yoruba mid-toned syllable interacts with an adjacent high or low, the mid is completely displaced. For instance, there is a purely tonal morpheme, an H tone, that marks the end of subject NPs in Yoruba. This H will form a rising glide with a preceding L, but it displaces a preceding M:

\[
\begin{align*}
(6) & \quad \text{a. } o\text{m}o \quad (M) + (H) \quad + \text{le} \quad (M) \quad \Rightarrow \quad o\text{m}o \text{ le} \quad (MH \ M) \\
& \quad \text{child} \quad \text{subject marker} \quad \text{stubborn} \quad \text{"the child is stubborn"} \\
& \quad \text{b. } ag\text{ba} \quad (M) + (H) \quad + \text{le} \quad (M) \quad \Rightarrow \quad ag\text{ba} \text{ le} \quad (MLH \ M) \\
& \quad \text{elder} \quad \text{subject marker} \quad \text{stubborn} \quad \text{"the elder is stubborn"}
\end{align*}
\]
As pointed out in Akinlabi 1985, the displacement of M in simple nouns, as in (a) above, contrasts with the treatment of final Mid tone when the subject is morphologically or phrasally complex. For example, the phrase *omóle* meaning "the child is stubborn" is also the name of a prominent family in western Nigeria. When this name, which ends in a mid tone, is the subject of a sentence, the subject-marking H does not displace its final mid tone, but rather forms a M-H glide. The same is true for any subject including more than one closed-class word.

\[(7) \quad \text{o}m\text{óle} \quad \text{subject marker} \quad \text{come} \quad \text{"Omole went"} \]

\[(8) \quad \text{nominalizing prefix} \quad \text{to command} \quad \text{subject marker} \quad \text{to be numerous} \quad \text{"orders abounded"} \]

In Akinlabi 1985, this was taken as evidence that Yoruba has a surface Mid tone, which is filled in by default for all unspecified TBUs at the end of each phonological cycle.

In our present treatment, the subject-marking H is in effect a junctural or boundary tone, which docks on a preceding unspecified syllable at the end of a simple subject, but is prevented from doing so by the "extra right brackets" at the end of a complex subject.

6 Appendix 2: Interpolation or relaxation?

Pierrehumbert and Beckman 1989 argue for straight-line interpolation across tonally unspecified syllables between F0 targets associated with tonally specified syllables. This view is inconsistent in some cases with the view that Yoruba M is unspecified phonetically. Consider for instance a Yoruba LMLM sequences, which is a rise-fall-rise, but would be low level or falling if the M syllables were treated by interpolation between L targets. If we wish to maintain Yoruba Mid tones as unspecified on the surface, we must conclude that in such cases, "unspecified" means to adopt a neutral F0 level.

However, P&B's examples do not speak very clearly to this question, since in their analysis of Japanese, the tonally unspecified moras are always between H and L. Therefore there is little empirical difference in prediction between a theory that postulates interpolation and one that postulates relaxation. An example is given below:

\[(9) \quad \text{moriya-no mawari-no} \quad \text{omawarisan} \]

Moriya’s neighborhood’s policemen

In P&B's analysis, there is an H associated with the second syllable of *moriya*, and an L associated with the first syllable of *omawarisan*. In between these two targets, their analysis has all the moras unspecified, with the realization being a linear interpolation.
Figure 1:
Exemplification of L-before-H lowering

Figure 2:
Downdrift (HLHL) and Declination (HMHM, MLML)

Figure 3:
Akin HL (narrow pitch range): Time 0 at end of medial C

Akin ML (wide pitch range): Time 0 at end of medial C

Akin HM (wide pitch range): Time 0 at end of medial C

Figure 4:

Figure 5:

Figure 6:
Akin LH (narrow pitch range):  Time 0 at end of medial C

Figure 7:

Akin LM (wide pitch range): Time 0 at end of medial C

Figure 8:

Akin MH (wide pitch range): Time 0 at end of medial C

Figure 9:
References


