Tone, Accent and Relative Prominence

Mark Liberman
University of Pennsylvania

1 Yoruba

The phonetic interpretation of Yoruba tone reveals some unexpectedly “accentual” aspects. To see why they are unexpected, we start by looking at Yoruba’s essentially tonal core.

1.1 Lexical distribution of Yoruba tone is free

Yoruba has three phonemically distinctive tones—H(igh), M(id), L(ow). 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 words start with a vowel, which is L or M but not H. Except for this minor tonotactic restriction, tones occur freely in lexical representations, without apparent restriction on word melodies.

<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>ọkọ MH</td>
<td>ọkọ MM</td>
<td>ọkọ 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>

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.

1.2 Mid is 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.

L & 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 (Barngbose 1966, Oyelaran 1970, Akinlabi and Oyebade 1987, Pulleyblank 1988).

---

1 Joint work with Akin Akinlabi (Rutgers) and Moussa Bamba (UPenn)
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, which will not be given here.

(2)  

a. wa (H) + ẹkọ (LH) ⇒ ṟẹkọ (H LH)
look (for)educationlook for education
b. mu (H) + iwe (LH ⇒ muwe (H LH)
takebooktake a book
c. wa (H) + ọnọ (LL) ⇒ ọnọ (H L)
look (for)waylook for a way
d. wa (H) + imọ (LL) ⇒ ọmọ (H L)
look (for)knowledgelook for knowledge
e. ji (H) + ọ̀bẹ̀ (LM) ⇒ ọ̀bẹ̀ (HL M)
stealknifesteal a knife
f. ẹ̀ (H) + iwo (LM) ⇒ ẹ̀wo (HL M)
wanthornwant a horn
g. wa (H) + ọ̀ wọ̀ (MH) ⇒ ọ̀ wọ̀ (HH)
look (for)moneymoneylook for money
h. wa (H) + ile (MH) ⇒ ọ̀ lẹ̀ (HH)
look (for)housetheshed
i. jọ̀ (M) + ọ́jẹ̀ (LH) ⇒ jọ́jẹ̀ (LH)
resemblewitchresemble a witch
j. sin (M) + ọku (LH) ⇒ ọ̀kọ̀ (LH)
burydead (body)bury the dead

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

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).

Thus in all the cases that can arise, and whose output is not obscured by the deletion of verbal L, we can say that H and L always remain when their lexically-associated vowel deletes, while M never does.

N.B. The examples above whose output is specified as (HL M) are pronounced exactly as this notation implies in some dialects, but in standard Yoruba, they are pronounced as H followed by a a tone described as being between M and L, a sort of “downstepped mid.” The proper treatment of this case is not clear, but it is clear enough that the L tone is in some sense preserved.

B. 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.

(3)  

a. ile (MH) + ẹ̀bọ̀ (HL) ⇒ ẹ̀bọ̀ (MH M HL)
house Tàyo Tàyo’s house
b. ọ́kọ̀ (ML) + ọ́dọ̀n (LM) ⇒ ọ́kọ̀ ọ̀dọ̀n (ML M LM)
car DotunDotun’s car
C. Tonal Spreading (mostly) treats Mid as nonexistent

Adjacent H and L tones always spread (rightwards) onto each other to create contours, whose phonetic characteristics we will examine in more detail shortly. They normally do not spread onto M, a curious fact to be examined more closely later.

When H or L spread onto M, M is (usually) completely erased.

Thus 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 said to be homophonous with an underlying HH or LL sequence (although as we will see, such sequences are not very distinct phonetically in any case).

M never spreads onto any adjacent tone—as indeed it could not if it does not exist.

A more striking example results from the existence of a purely tonal morpheme, an H tone, that marks the end of (most) subject NPs in Yoruba. This H always forms a rising glide with a preceding L, but it displaces a preceding M if the subject is in some sense morphosyntactically simplex:

The displacement of M in (a) and (b) above, in contrast to the preservation of L in (c), is consistent with the interpretation of M as lack of tonal specification.

There is a small difficulty to be discussed. As pointed out in Akinlabi 1985, a subject-final M tone is NOT displaced when the subject is morphologically or phrasally complex, instead forming a MH glide.

For example, the phrase Omole meaning “the child is stubborn” also happens to be 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. This remains true even for common nicknames derived by truncation, such as Olórùm tó mí “God is enough for me” Rightarrow tó mí. More generally, the subject-marking H does not displace a final M on any subject including more than one closed-class word, whether as part of a compound word, or a phrasal structure, or a name derived from a phrase.
b. Tomi (HM) + (H) + 10 M ⇒ Tomi 10 (HM+H M)
   Tomi subj. go “Tomi went”

c. o - li - ran (M H MM) + (H) + 10 (M) ⇒ oleran 10 (MHM+H M)
   prefix - to have - beef subj. go “the butcher went”

d. 1omọ (MM) + ọkunrin (MHM) + (H) + 10 (M) ⇒ 1omọ ọkunrin 10 (MM MHM+H M)
   child male subj. go “the boy went”
e. 1omọ (MM) + ti o ri mi (H H M) + (H) + 10 (M) ⇒ 1omọ ti o ri milo (MM H H M+H M)
   child that he see me subj. go “the child that saw me went”

In Akinlabi 1985, this difference 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. Subject-marking H is presumed to be a tonal morpheme arising in INFL. The phonological cycle is supposed to work so that M will have been filled for complex subjects, but not for simplex ones, at the point in the derivation when the subject-marking H is available for linking.

It seems more promising, in the current context, to rely on the “extra right brackets” at the end of a complex subject to prevent incorporation of the junctural H into the last syllable of the last word of the subject, forcing it to remain as an extrasyllabic tonal target, as junctural tones often do.

The technical details of such a treatment remain to be supplied. The point relevant to the current discussion is the complete displacement of final M (as opposed to L) in simplex subjects, which again argues for M being a lack of tonal specification. The fact that such displacement does not occur with complex subjects can be seen as providing proof that displacement is genuinely occurring in the simplex case, since it shows that the alternative outcome is phonetically distinct.

1.3 The paradox of Yoruba tone (non-)spreading

In configurations (a) and (b) below, where each syllable has its own tone, the first syllable’s tone insists on crowding onto the second syllable. In configurations (c) and (d), where the second syllable is unspecified for tone, the first syllable’s tone stays at home, leaving its neighbor tonally empty (i.e. Mid).

a. σ σ
   H L
b. σ σ
   L H
c. σ σ
   H
d. σ σ
   L

This is the Williams/Goldsmith tonal well-formedness condition turned on its head: the rule seems to be “a tone spreads iff the target syllable already has a tone.”

Proposed analysis: Yoruba tones do not spread in order to fill in tonally-unspecified vowels; rather, they spread in order to join with other tones to form tonal complexes, HL or LH units.

These tonal complexes have some phonetic similarities to (underlying) pitch accents in other languages.

1.4 Excursus: phonetic aspects of Yoruba tone

In this discussion, we include phonetic aspects of the tone-spreading phenomenon.
1. Tones associated with multiple tone-bearing-units have just one F0 target; this is a sort of “phonetic OCP.”

2. 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);

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

4. 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).

1.4.1 Phonetic OCP

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

Here there are two L syllables in a row, /wa/la/. In such cases, it is typical in Yoruba to have just one tonal target, placed at the end of the domain of like tones. Instead, we might have imagined that the H-to-L fall would mainly take place on the first syllable, with the second syllable level low. The necessary rate of F0 change can certainly be found in other Yoruba examples, so the systematic failure to see a falling-plus-level pattern cannot be due to implementation constraints.

This perspective is contra Laniran 1992 on tone interpretation in Yoruba, who postulates several F0 targets for a single stretch of like tones in certain cases. In some of her examples, there are additional M tones not marked in her analysis (arising due to the epenthesis process mentioned earlier). In other cases, it is possible that a long stretch of like L or H tones may be broken up into more than one domain. However, the typical case (and pretty much the only case for sequences of two or three H or L tones) is the one exemplified in Figure 1.
Notice that we seem to have a sequence of tones (here a HL glide) that is aligned with a sequence of syllables (here /wa la/). We will propose a general notation for this state of affairs later on.

### 1.4.2 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 2, 3, 4.

![Figure 2](image1.png)

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

![Figure 3](image2.png)

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

![Figure 4](image3.png)

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

Falling sequences work similarly: in a HI 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 5, 6, 7.

![Graph](image1.png)

**Figure 5**

![Graph](image2.png)

**Figure 6**

![Graph](image3.png)

**Figure 7**

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 phe-
nomenon, and also a sort of prefiguration of our suggest that what is involved is a sort of derived pitch accent.

See Figure 8 for the pitch track of the sentence Ṣórúnámí lẹ́mi “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.

![Figure 8](attachment:figure8.png)

1.4.3 Tone dissimilation

As Table 1 shows, the value of H is significantly higher when L follows compared to when M follows. This exemplifies H-before-L raising, which accompanies H-onto-L spreading in Yoruba.

<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 9, which plots the relationship of successive L tones in the sequences HLHLM (plotted with squares) and HLHLH (plotted with pluses). When both L tones are followed by H, the second L tone is considerably lower than the first. This is consistent with the general expectation of downdrift 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 downdrift 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 downdrift lowering of the second L, leave them at about the same level.

This exemplifies L-before-H lowering, which accompanies L-onto-H spreading.
1.4.4 OCP, spreading and dissimilation: problem and possible solution

The OCP business and the tone spreading business, taken together, pose a problem for the conventional autosegmental notation. We propose a solution that can also contribute to a speculative connection between tone spreading and tonal dissimilation.

We observed that sequences of like L or H tones in Yoruba are (typically) treated phonetically as a single tonal domain, contributing a tonal target at the end of the sequence, just as if there were only a single tone and a single syllable.

The autosegmental notation for this state of affairs is not very helpful. Neither the formation of the HL and LH units, nor the essential nature of the tonal domains, is perspicuous.

\[
(9) \quad C \quad V \quad C \quad V \quad C \quad V \quad C \quad V \quad C \quad \downarrow \quad H \quad \downarrow \quad L \quad \downarrow \quad H
\]

It is more helpful to extend the notation we used earlier, indicating domains with aligned parentheses, and placing the parentheses and tones so as to indicate the F0 target locations:

\[
(10) \quad C \quad V \quad C \quad V \quad C \quad V \quad C \quad V \quad C \quad V \quad ( \quad H \quad ) \quad ( \quad L \quad ) \quad ( \quad H \quad )
\]

So far this is just a variant notational for the picture given in (9), before tone spreading. The spreading in effect allows the parentheses to cross, creating overlapping constituents of a type that a bracket-based notation does not
represent easily. In a case like this, an alternative way to encode the same information is to suppose that each domain has two anchor points—the one at the end that we have already used, and another one at the beginning. In this notation, tone spreading copies a tone across the sequence, creating something like:

\[
(11) \quad \text{C V C V C V C V C V (H) (H L) (L H)}
\]

This notation suggests that Yoruba HL and LH tone spreading creates an extra pair of tonal targets, one at the end of the original tonal domain, and another at the start of the following tonal domain. This is actually pretty much what we want to do in order to create F0 time functions for Yoruba phrases, as we will soon see.

This move may give us an explanation of the dissimilation phenomenon. It is plausible to assume (Öhman 1967, Fujisaki 1981) that (log) F0 varies with displacement of (some point of) the laryngeal structure, while displacement of that point involves a mass element with stiffness and viscous resistance elements. This will produce a smoothing (lowpass) effect; in everyday language, the larynx cannot vary F0 at an arbitrary rate, instead having a maximum rate of change. An attempt to produce an abrupt change will be “smoothed out” in a characteristic way, as shown in Figure 11.

![System step response](image)

Figure 11

If we assume that this smoothing process is linear (it probably is not, but a linear approximation to the system may still show us something of interest), then we can predict its output for an arbitrary input by a simple numerical calculation. It is characteristic of such systems (as has often been pointed out) that they may produce “undershoot.” Spreading an H or L target out in time, as is accomplished by our version of tone spreading, will obviously reduce or eliminate such undershoot, as exemplified in Figures 12 and 13. This reproduces a tone dissimilation effect, caused not by actually raising or lowering the H or L target, but rather by failing to undershoot the target when it is spread out in time due to being duplicated at the start of the following tonal domain.
This is probably not the whole story of tonal dissimilation— but it may help us to understand where the phenomenon comes from, and suggests an interesting way of thinking of tonal domains, motivating us to look closely at the detailed time functions of F0 for clues about the phonological analysis of tone.

1.4.5 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.

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 10. 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 downdrift, the tonologically-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 downtrend sometimes distinguished as “declination.”

1.5 Comparison to a pitch accent system

Comparing Yoruba to Japanese, a 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 into a sort of compound tonal domain when they happen to be adjacent.
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. Acccentual 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.

### 1.6 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:

(12) moriya-no mawari-no 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.

In addition, P&B’s argument for sparse tonal association with interpolation (as opposed to the tone spreading analysis assumed by previous authors) is only that the intervening material shows “sloping” F0 values, interpolated from one target to another, rather than “level” F0 values that would represent a naive phonetic interpretation of the spreading analysis. In Yoruba, we see a similar “sloping” interpretation over a string of like tones derived from the lexicon, suggesting that the naive interpretation is wrong, and a string of like tones are treated phonetically as a single tonal domain with a single (new) F0 target.

### References


