Original Paper

Phonetica 2019;76:55–81
DOI: 10.1159/000491073

Received: June 14, 2017
Accepted after revision: January 20, 2018
Published online: July 18, 2018

F_0 Timing and Tone Contrasts in Luganda

Scott Myers[a] Saudah Namyalo[b] Anatole Kiriggwajjo[b]

[a]Department of Linguistics, The University of Texas at Austin, Austin, TX, USA;
[b]Department of African Languages, Makerere University, Kampala, Uganda

Abstract

Scholars of Luganda (Bantu, Uganda) have described a contrast between falling and high tone, which is limited to syllables with a long vowel or a coda. The contrast has been represented with H on the first mora of a falling-tone syllable and on both moras in a high-tone syllable. This article explores this contrast through an instrumental study of the timing of f_0 events in Luganda. It was found that the 2-tone classes differed in the timing of both the f_0 rise and the subsequent f_0 fall, supporting a reanalysis of the contrast as one between an early high tone and a late high tone. It was also found that for the speakers in this study, the contrast was limited to syllables with long vowels. The timing of f_0 events was sensitive to the duration of segments in the CVC interval centered on the high-toned syllable, including non-moraic elements and consonants outside the syllable. The association of a tone to a syllable identifies the segments relevant for the timing of the f_0 contour, but that timing is not directly sensitive to the interval of the syllable.

1. Introduction

In a number of languages with lexical tone contrasts, contour tones are limited to heavy syllables, for example, Lithuanian (Kenstowicz, 1970; Blevins, 1993), Northern Tepehuan (Woo, 1972), Hausa (Leben, 1978), Somali (Banti, 1988), Kimatuumbi (Odden, 1996), Tswana (Chebanne et al., 1997), Kinyarwanda (Myers, 2003), Thai (Morén and Zsiga, 2006; Zsiga and Nitisaroj, 2007). One approach to such systems presupposes moras as tone-bearing units (Hyman and Katamba, 1993). Then a falling tone has a high tone on the first mora of a heavy syllable (Figure 1a), a rising tone has a high tone on the second mora (Figure 1b), and a high tone in a heavy syllable has a high tone on both moras (Figure 1c).

The mora has been established as a tone-bearing unit in the phonological literature (McCawley, 1978; Odden, 1995; Yip, 2002). Hyman (1993: 77) has gone further to propose that, at least with regard to phonological patterns, the mora is the only possible tone-bearing unit.

KARGER

E-Mail karger@karger.com
www.karger.com/pho

Scott Myers
Department of Linguistics, The University of Texas at Austin
305 E. 23rd St., B5100
Austin, TX 78712 (USA)
E-Mail s.myers@austin.utexas.edu

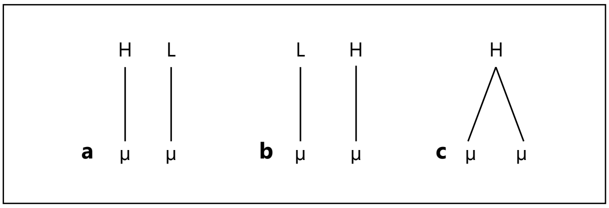


Fig. 1. Moraic representation of (**a**) falling tone, (**b**) rising tone, and (**c**) high tone.

In terms of acoustic realization, all the 3 categories in Figure 1 include a high tone target, and hence would be expected to be associated with a rise and subsequent fall in f_0. But they would differ in timing. Because the falling tone in Figure 1a has a high tone on an earlier mora within the syllable, it would be expected that the corresponding f_0 rise and fall would be earlier in the syllable than in the rising tone in Figure 1b. This is the case in Kinyarwanda (Myers, 2003). The falling tone in Figure 1a and the high tone in Figure 1c both have high tone on the first mora, so they would be expected to have the same kind of f_0 rise, but with a f_0 fall earlier in the syllable in the falling tone than in the high tone. This contrast will be investigated in this study, in Luganda, a Bantu language spoken in Uganda, in which contour tones are restricted to heavy syllables.

The literature on f_0 timing has demonstrated that there is a close temporal coordination between f_0 events and segmental intervals. Silverman and Pierrehumbert (1990) showed that there is a strong correlation in English prenuclear H* between the duration of f_0 rise relative to vowel onset and the duration of segmental rime. Subsequent work has found such regularity in the timing of f_0 events relative to segmental intervals in languages such as Spanish (Prieto et al., 1995), Dutch (Rietveld and Gussenhoven, 1995; Schepman et al., 2006), Mandarin (Xu, 1998; Xu, 2001), Greek (Arvaniti et al., 1998), Chichewa (Myers, 1999), and Kinyarwanda (Myers, 2003).

However, there has been controversy about the segmental units that f_0 contours are timed with respect to. The mora, as in Figure 1, has been argued to be the segmental basis of f_0 timing in Thai (Morén and Zsiga, 2006; Zsiga and Nitisaroj, 2007) and in Yoloxóchitl Mixtec (DiCanio et al., 2014). But Howie (1974) and Silverman and Pierrehumbert (1990) used the rime as their segmental timing reference interval, while Van Santen and Hirschberg (1994), Prieto et al. (1995), Myers (1999), and Xu (2001) argued that f_0 events are timed with respect to syllables. The present paper is intended to contribute to this discussion through a study of f_0 timing in Luganda.

1.1 Background on Luganda

In Luganda, there is a 2-way tone contrast between high tone (H) and low tone (L) in word-medial *short* syllables, that is, open syllables with a short vowel, as in the bold-faced syllable in (1a) and (1b) (syllable boundaries marked here with periods). However, there is a 3-way contrast in what Tucker (1962) refers to as *long* syllables, which are syllables with a coda, as in (1c–e), or a long vowel, as in (1e–h). In a long syllable, the tone can be low, as in (1f), high level, as in (1g), or falling, as in (1c), (1d), (1e), and (1h) (Tucker, 1962; Cole, 1967; Stevick, 1969; Hyman, 1982; Hyman and Katamba, 1993; 2010).

DOI: 10.1159/000491073

(1) (a) [kù.**βá**.là] *kubala* "to produce fruit"
(b) [kù.**βà**.là] *kubala* "to count"
(c) [jà.**láǹ**.nà.mà] *yalannama* "he stretched out"
(d) [ò.kù.**tʃóp̀**.pà] *okucoppa* "to be destitute" (Tucker 1967: xx)
(e) [jà.**lîn.**dà] *yalinda* "he waited for"
(f) [kù.**ɲòò**.là] *kunyoola* " to twist"
(g) [ò.**mwáá**.mì] *omwami* "chief"
(h) [ò.**mwáà**.nà] *omwana* "child"

Tucker (1962) described long syllables in Luganda as having 2 moras, and short syllables as having just one. The contrast between falling tone and high tone in Luganda can then be represented in terms of which mora the high tone is associated with, and this has been how it has been represented in the literature on Luganda tone (Tucker, 1962; Cole, 1967; Stevick, 1969; Hyman, 1992; Hyman and Katamba, 2010). The falling tone has H on the first mora, as in Figure 1a, and the high-level tone has H on both moras, as in Figure 1c. There is no rising tone in Luganda, that is, no syllables with high tone just on the second mora, as in Figure 1b (Tucker, 1962; Cole, 1967; Stevick, 1969; Hyman, 1982; Hyman et al., 1987).

In Luganda, according to these descriptions, the tone-bearing units are moras: a short vowel, either part of a long vowel, or the first (coda) part of a geminate consonant. Sonorant codas are tone-bearing units in Lithuanian (Kenstowicz, 1970), Thai (Morén and Zsiga, 2006), and Roermond Dutch (Gussenhoven, 2000). The role of codas in f_0 timing has been established for English (Van Santen and Hirschberg, 1994), Mandarin (Xu, 1998), and Dutch (Schepman et al., 2006). Luganda is interesting in this regard because the tone-bearing consonant in the geminate case is just part of a long consonant that belongs both to the coda of one syllable and to the onset of the next syllable, with only the coda portion eligible to bear a tone. There is no acoustic boundary between the part of the geminate that belongs to the coda and the part that belongs to the onset. Moreover, a geminate in Luganda is remarkable as a tone-bearing unit as it can belong to any consonant class except the non-nasal sonorants (liquids or glides), so the consonantal tone-bearing units include not only sonorant nasals as in (1c), but voiceless obstruents as in (1d).

In this paper, findings are presented for an instrumental study of f_0 timing in Luganda, designed to test the acoustic implications of the mora-based representation in Figure 1, as it plays out over the different syllable types of Luganda. One unexpected finding of this study was that both the f_0 rise and the f_0 fall are significantly earlier in what has been described as the "falling" tone class compared to what has been described as the "high-level" one. This suggests that it is more appropriate to describe the 2 categories as "early" and "late" high tones, respectively, which is how we will refer to them henceforth. Such a contrast has not been described previously for Luganda, but it is parallel to other tone contrasts that lie in the earlier or later timing of the same f_0 event, as in Serbian (Purcell, 1976; Smiljanić, 2006), Swedish (Bruce, 1977), Kinyarwanda (Myers, 2003), Dinka (Remijsen, 2013), and Shilluk (Remijsen and Gwado Ayoker, 2014). There are also such timing contrasts in intonational tones, as in German (Kohler, 1987), English (Ladd, 1983; Pierrehumbert and Steele, 1989), Palermo Italian (D'Imperio and House, 1997; D'Imperio, 2000), European Portuguese (Frota, 2002), and Greek (Arvaniti et al., 2006).

Table 1. Speakers in the study

Speaker	Sex	Hometown (district)	Birth year
S1	Female	Masaka (Masaka)	1972
S2	Male	Ssi (Mukono)	1965
S3	Male	Kiganda (Mubende)	Unknown
S4	Male	Kassanda (Mubende)	1966
S5	Male	Kifamba (Rakai)	1980
S6	Male	Bulamu (Mpigi)	1978
S7	Female	Kayunga (Kayunga)	1980
S8	Female	Gayaza (Wakiso)	1990
S9	Male	Kampala (Kampala)	1987
S10	Male	Kitaama (Mubende)	1972

2. Methods

2.1 Speakers

Ten native speakers of Luganda participated in the study: 3 females and 7 males. All had grown up in the central region of Uganda and were living in the Kampala area at the time of the study. The speakers are described in Table 1.

They ranged in age at the time of the experiment from 24 to 49 years. They came from all over the central region of Uganda, which corresponds to the traditional Buganda kingdom and is homeland of the Baganda, from the Rakai district in the Southwestern corner of the region (S5) to Mukono in the Southeast (S2), Mubende in the Northwest (S3, S4, S10) and Kayunga in the Northeast (S7).

2.2 Materials

All test sentences had low lexical tones throughout except for one test syllable with an early or late high tone in the stem-initial syllable of a sentence-medial verb. The test word was a 3rd person singular verb with low-toned prefixes. The vowel of the test syllable had sonorant consonants on both sides, either a liquid /l/ or a nasal /m, n, ɲ, ŋ/. These consonants were chosen because they were sonorants, which induce minimal local disturbances in the f_0 contour, and have well-defined temporal edges, facilitating segmentation. The set of test words is listed in Table 2.

The test syllable belonged to one of 4 classes: CV, CVC, CVVN, or CVV. A CV syllable was an open syllable with a short vowel. A CVC syllable had a short vowel followed by a geminate nasal stop (the vowel being obligatorily short before a geminate: Tucker, 1962; Clements, 1986). A CVVN syllable had a long vowel followed by a nasal-obstruent sequence (the vowel being obligatorily long before such a consonant sequence: Tucker, 1962; Clements, 1986). A CVV syllable was an open syllable with a long vowel.

We will assume for purposes of discussion that the CVC and CVVN syllable types are closed (C-final), and that the CV and CVV are open (V-final), as proposed by Tucker (1962) and Goldsmith (1990). Some authors have argued that all syllables in Luganda are open in surface representation (Katamba, 1974; Herbert, 1975; Clements, 1986), but one's position on this controversy has no bearing on the particular issues dealt with in this study.

The test verbs belonged to 2-tone classes: early or late high tone. The early-high items in the sample were either present-tense verbs with a high-tone root, or verbs in the recent past. The verb stem is preceded by a low-toned syllable. In items that meet these conditions, there is a high tone located on the stem-initial syllable on the first mora (Hyman, 1992; Hyman and Katamba, 1993). Thus, if the first syllable of the stem is long, that syllable is expected to have falling tone in these tenses. If the first syllable of the stem is short, that syllable is expected to have a high tone.

Table 2. All test words, by syllable type and association class

	Early	Late
CV	[à**nó**nà] "he/she is getting" [à**mé**ɲà] "he/she is cutting" [jà**nó**nà] "he/she got" [jà**lú**mà] "he/she bit" [jà**mí**ɾà] "he/she swallowed" [jà**lé**ɾà] "he/she carried" [jà**lí**mà] "he/she planted" [à**lú**mà] "he/she is biting" [jà**né**ɲà] "he/she blamed" [jà**mé**ɲà] "he/she cut"	
CVV	[jà**nóò**ɲà] "he/she looked for" [à**ɲéè**ɲà] "he/she is shaking" [jà**ɲóò**mà] "he/she despised" [à**nóò**ɲà] "he/she is looking for" [jà**ɲóò**là] "he/she twisted" [jà**ɲéè**ɲà] "he/she shook"	[à**ɲóó**mjè] "he/she has despised" [à**láá**mjè] "he/she has made a will"
CVVN	[jà**lîìn**dà] "he/she waited" [jà**láàm**bà] "he/she labeled" [jà**láà**ŋgà] "he/she announced" [jà**lóòn**dà] "he/she picked" [jà**lîìm**bà] "he/she lied"	[à**líín**zè] "he/she has waited" [à**lúún**zè] "he/she has looked" [à**láán**zè] "he/she has announced" [à**líím**bjè] "he/she has lied" [à**láám**bjè] "he/she has labeled"
CVC	[jà**líɲ̀**ɲà] "he/she climbed" [à**líɲ̀**ɲà] "he/she is climbing" [jà**láǹ**nàmà] "he/she stretched out"	[à**láń**nàmjè] "he/she has stretched out."

The late-high items were present-perfect verbs with low-toned roots and prefixes. In these verbal classes, there is a high tone associated with the second mora of the stem (Hyman, 1992; Hyman and Katamba, 1993). If the first syllable of the stem is long, the high tone goes on the second mora of that syllable, yielding a high-level tone. If the first syllable of the stem is monomoraic CV, the high tone goes on the second syllable of the stem. The test syllables in this study are all stem-initial, so late items with stem-initial CV syllables were not included in the analysis.

Each test word was embedded in test sentences with a preceding subject and a following object or adjunct. Sample test sentences are presented in Table 3, which is organized by syllable type (rows) and tone (columns). The test syllable is bold-faced in the phonetic transcription, and the verb stem is delimited by square brackets in the orthographic representation. The tonal transcriptions represent the falling and high-level tones as they have been transcribed in the literature.

There were 5–10 distinct sentences for each category, repeated enough times to yield 20 tokens in each category. There was a total of 140 tokens per speaker, and 1,400 tokens in the study as a whole.

There was an error in the construction of one of the test sentences in the early/CVVN category such that it had no lexical high tone. Each speaker produced this sentence twice, implying that 20 items had to be excluded from the analysis. Nine further items had to be excluded: 2 tokens because the speaker produced an incorrect word in place of the test word, 1 token because creaky voice interrupted the f_0 rise-fall, 1 token because the syllable-initial consonant was omitted, and 5 tokens because the speaker produced a pause immediately before or after the test word so that it was not phrase-medial. This left a total of 1,371 items for analysis.

Table 3. Example test sentences, by syllable type and association class

	Early	Late
CV	[òmùlèènzà**nó**nèɲ̀ɲàmà] Omulenzi a[nona] ennyama boy get-present meat "The boy is getting the meat."	–
CVC	[òmùzììmbìjà**líɲ̀**ɲòòmùgùwà] Omuzimbi ya[linnya] omuguwa builder climb-past rope "The builder climbed the rope."	[òmùsùùmbàà**láń**nàmjèmùd̀dùùndìɾò] Omusumba a[lannamye] mu ddundiro shepherd spread-perfect in farm "The shepherd has stretched out in the farm."
CVVN	[òmùlwààɲìjà**líìn**dòòmùlààŋgìɾà] Omulwanyi ya[linda] omulangira fighter wait-past prince "The fighter waited for the prince."	[òmùlwààɲàà**líín**zòòmùlààŋgìɾà] Omulwanyi a[linze] omulangira fighter wait-perfect prince "The fighter has waited for the prince."
CVV	[òmùlààŋgìɾàjà**nóò**ɲòòmùsààŋgò] Omulangira ya[noonya] omusango prince look-for-past case "The prince looked for the law case."	[òmùlèènzà**ɲóó**mjòòmùlààŋgìɾà] Omulenzi a[nyoomye] omulangira boy despise-perfect prince "The boy has despised the prince."

2.3 Procedures

Recordings were made in a computer room on the campus of Makerere University in Kampala, Uganda, using a Shure SM10A head-mounted microphone and a Marantz PMD670 solid-state recorder, with a sampling rate of 44.1 kHz and 16-bit amplitude resolution.

The sentences were presented to the speakers in a PowerPoint slideshow presented on a laptop computer, with each sentence on a separate slide. The slides were shuffled into a quasi-random order. Speakers were instructed to read each sentence to themselves first, and then to produce it without internal pauses, as a statement and a separate utterance (rather than as a member of a list). They were told that if they were not satisfied with their initial production, they could keep saying the sentence until they felt they had it right. They proceeded at their own pace through the sentences, but were instructed to finish saying a sentence before pressing the key to switch to the next one.

The recording for each speaker was transferred to a computer, and broken down into separate sound files for each utterance. Where the speaker had produced a sentence more than once, the last one was used, unless it had a clear internal pause or slip.

2.4 Measurements

The following segmental landmarks were marked in annotations using Praat (Boersma and Weenink, 2013): the onset of C_1 (the consonant at the beginning of the test syllable), the onset of V_1 (the vowel of that syllable), the onset of C_2 (the consonant interval following that vowel), and the onset of V_2 (the vowel following the test syllable). The onset and offset of consonant intervals were the onset and offset, respectively, of the local minimum of intensity and wave complexity in the waveform. The vowel intervals were the intervals between consonant intervals.

In CV or CVV syllables, the C_2 interval includes one singleton sonorant consonant, which is the onset to the following syllable. In CVC syllables, the C_2 interval includes the whole long nasal consonant, since there is no phonetic boundary within the extended nasal stop closure of the geminate. In CVVN syllables, there is a nasal interval followed by an oral obstruent interval. The nasal interval is the coda in the assumed syllabification, while the whole nasal + consonant sequence is a heterosyllabic consonant cluster parallel to the geminate. It was unclear which interval would be the best predictor of f_0 timing, so both were measured, and the 2 were incorporated into competing models of f_0 timing.

A PitchTier file was made for each token using the autocorrelation method in Praat, with 5 Hz smoothing. The following f_0 landmarks were identified: the rise onset, the rise offset, the fall onset, and the fall offset. The f_0 rise was defined as the monotonically rising interval of f_0 points that overlapped in time with the test syllable. The rise onset was the first f_0 point in the sequence, that is, the first point that was preceded by a point with a lesser or equal numerical f_0 value, and followed by a point with a greater value. The rise offset was the last point in the sequence, that is, the first subsequent point that was preceded by a lesser f_0 value and followed by an equal or lesser value. The f_0 fall was defined in a parallel manner as the monotonically falling interval after the f_0 rise.

The f_0 rise and fall were defined and measured independently to address the hypothesis that the f_0 fall in late high tones is relative to the f_0 rise than in early high tones. As it turned out, the rise offset was quite often at the same temporal point as the fall onset, which was then the f_0 maximum of the rise-fall pattern. But the rise offset was also often followed by an interval in which f_0 moved slightly up or down without establishing a monotonic trend. This f_0 plateau interval between rise offset and fall onset was generally quite short (mean = 4.9 ms).

Using Praat scripts, the time points of the segmental and f_0 landmarks were collected, together with the f_0 values at the f_0 landmarks. From the time points, the durations of the following intervals were calculated in ms:

(2) Segmental intervals
- (a) C_1 interval: from the onset of C_1 to the onset of V_1
- (b) V_1 interval: from the onset of V_1 to the onset of C_2
- (c) C_1-V_1 interval: from the onset of C_1 to the onset of C_2
- (d) C_2 interval: from the onset of C_2 to the onset of V_2

(3) F_0 timing intervals
- (a) Rise lag: from the rise onset to the C_1 onset
- (b) Peak delay: from the C_1 onset to the rise offset
- (c) Fall delay: from the C_1 onset to the fall onset
- (d) Relative peak delay: the rise onset interval divided by the C_1-V_1 duration
- (e) Relative fall delay: the fall onset interval divided by the C_1-V_1 duration.

The segmental reference point for f_0 timing interval measurements was C_1 onset, that is, the beginning of the syllable, which has been established in cross-linguistic work as a stable reference point for the onset of f_0 movement (Van Santen and Hirschberg, 1994; Prieto et al., 1995; Xu, 1998). Relative peak and fall delay, as in (3e) and (3f), express peak and fall delay as a proportion of the C_1-V_1 interval.

The key intervals in (2) and (3) are illustrated in Figures 2–5, which give the annotated spectrogram and pitch-track displays for sample sentences in different syllable and tone classes. In the spectrogram displays, *C1* marks the C_1 interval defined in (2a), *V1* marks the V_1 interval defined in (2b), *C2* marks the C_2 interval defined in (2c), *rise* marks the f_0 rise from rise onset to rise offset, and *fall* marks the f_0 fall from fall onset to fall offset. The (b) display in each figure gives the corresponding PitchTier display, with red points indicating the f_0 rise interval.

Hypotheses

Based on the previous literature on f_0 timing, we would expect the duration of C_1, V_1, and C_2 intervals to be significant predictors of the f_0 timing intervals, with longer segmental intervals associated with later f_0 events, for example, greater peak delay and fall delay. Following Van Santen and Hirschberg (1994) and Xu (1998), we expect both the onset C_1 and the coda C_2 to play a role in predicting the location of the f_0 peak.

On the contrary, work on the timing of f_0 events with respect to moras (Myers, 2003; Morén and Zsiga, 2006; Zsiga and Nitisaroj, 2007; DiCanio et al., 2014) would suggest more specifically that it is the duration of moraic segments that would be the basis for predicting the timing of f_0 timing events. The moraic segment intervals are V_1 and C_2 in the case of closed syllables (CVC/CVVN). The non-moraic segments are C_1 and C_2 in open syllables (CV/CVV). If the timing of f_0 events depends on moras, it would be expected that peak delay and fall delay would be sensitive to the moraic segments but not to the non-moraic ones.

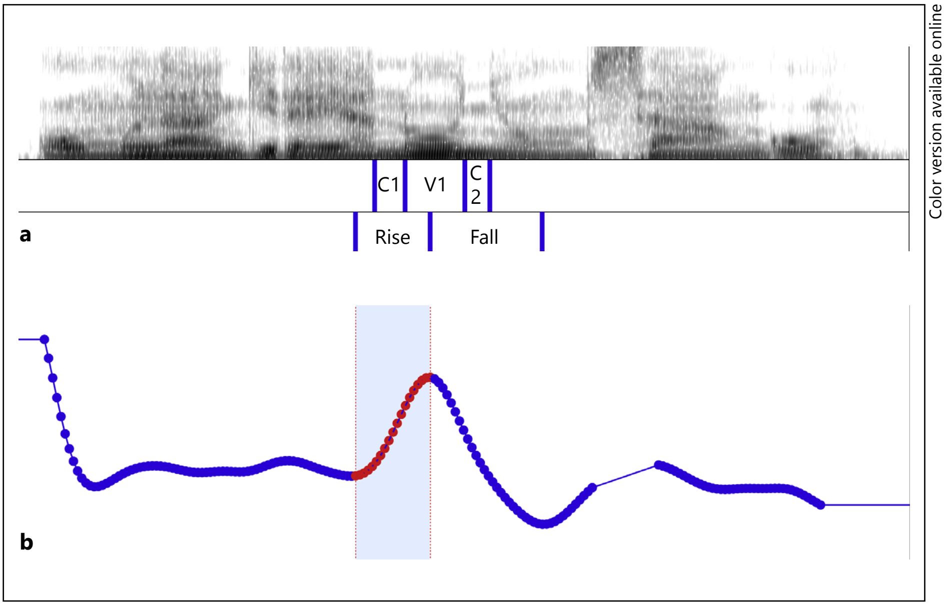


Fig. 2. Sample spectrogram (**a**) and pitch track (**b**) displays for the CVV/early sentence: *Omulangira yanoonya omusango*. "The prince looked for the law case." The test interval is *noo* [nô:].

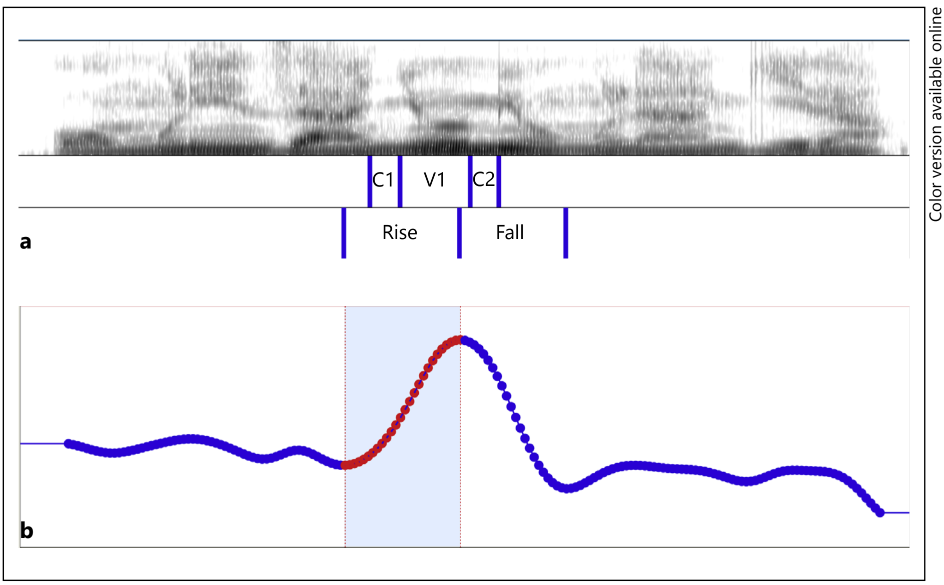


Fig. 3. Sample spectrogram (**a**) and pitch track (**b**) displays for the CVV/late sentence: *Omulenzi anyoomye omulangira*. "The boy has despised the prince." The test interval is *nyoo* [ɲó:].

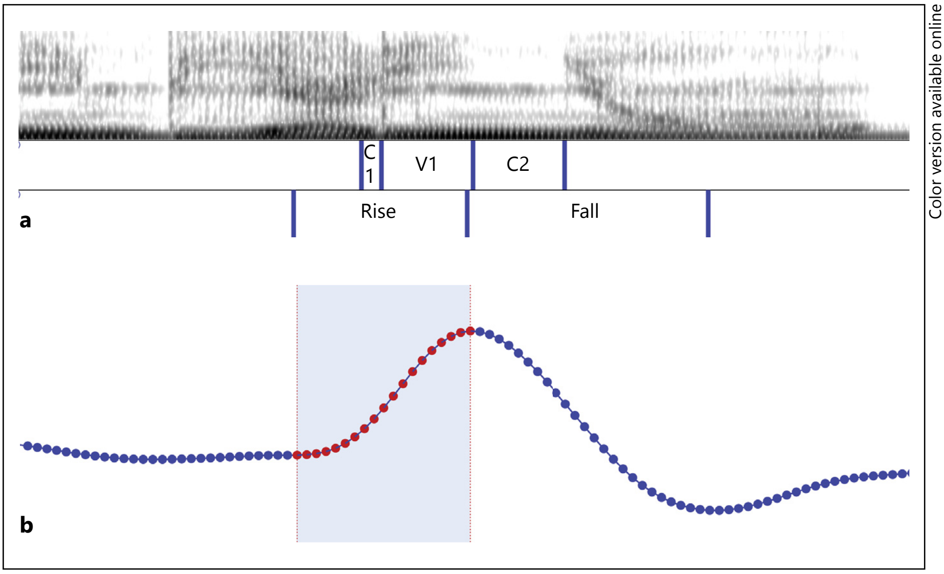


Fig. 4. Sample spectrogram (**a**) and pitch track (**b**) displays for the CVC/early sentence: *Omuzimbi yalinnya omuguwa*. "The builder climbed the rope." The test interval is *linny* [lî̞ɲ:].

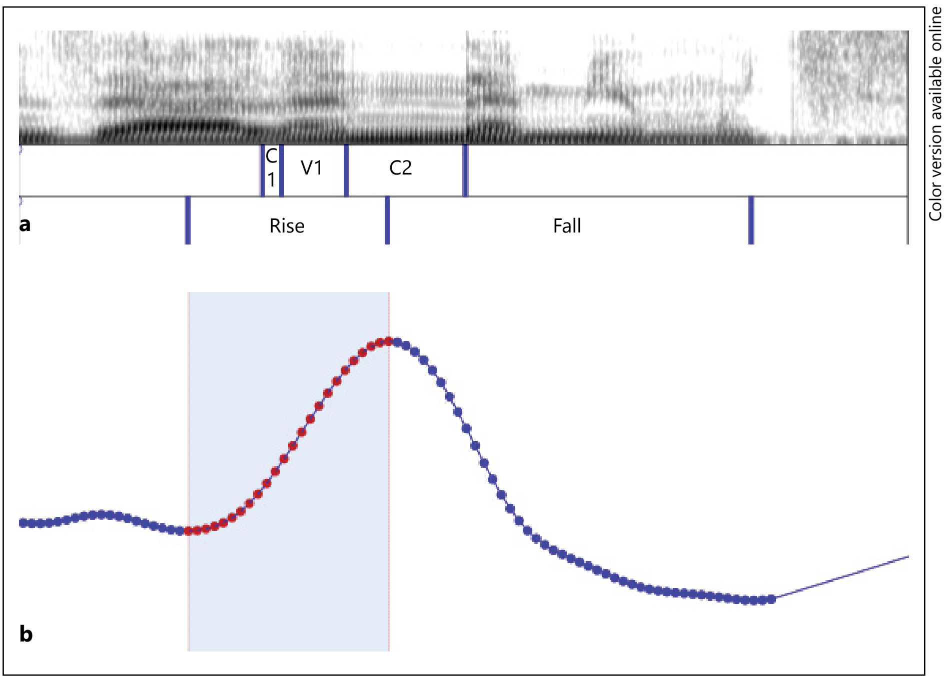


Fig. 5. Sample spectrogram (**a**) and pitch track (**b**) displays for the CVC/late sentence: *Omulenzi alannamye mu kisenge*. "The boy has stretched out in the room." The test interval is *lann* [lán:].

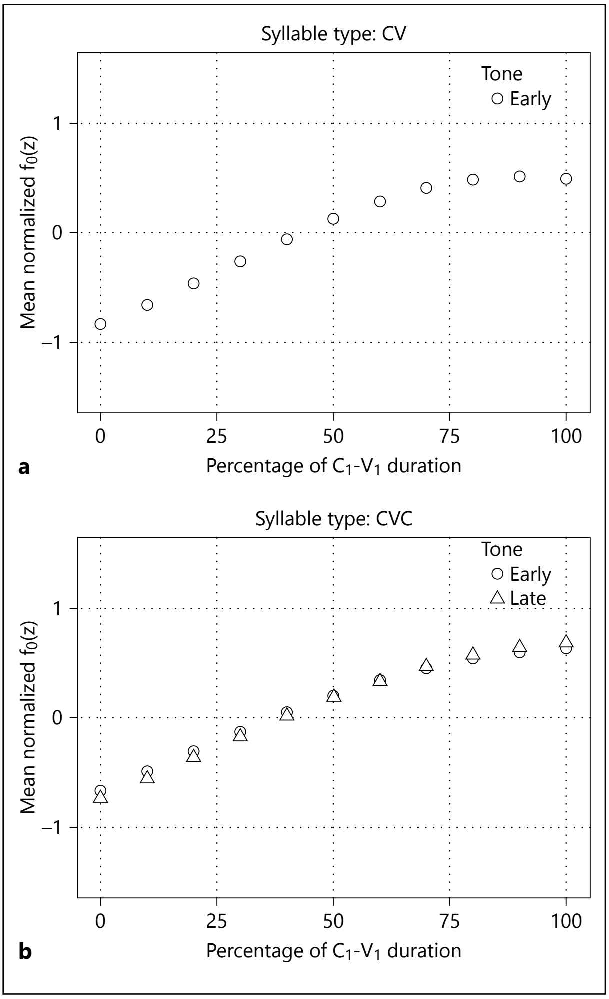


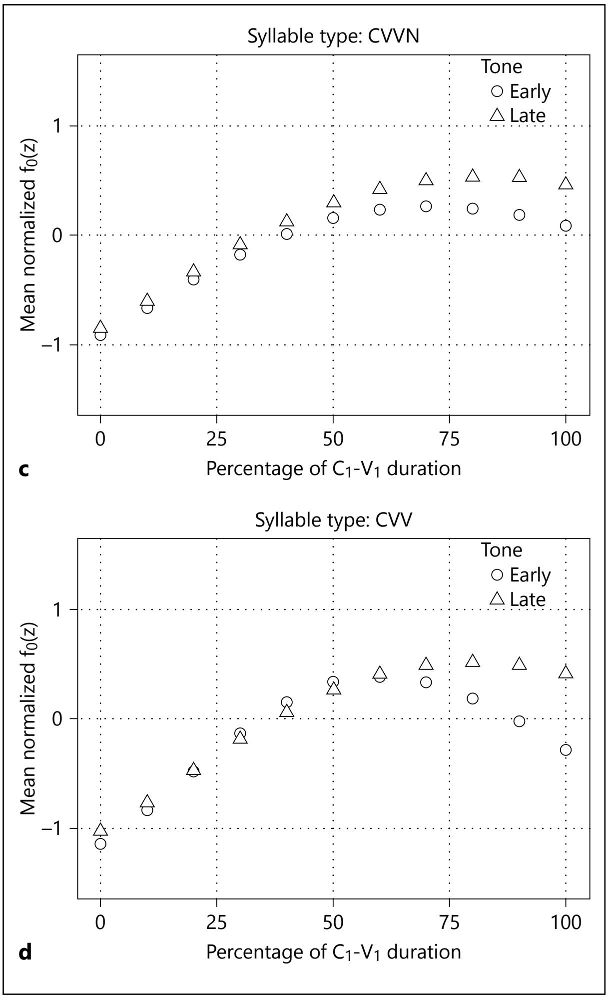
Fig. 6. Mean normalized f0 values (z) as a function of the proportion of the C_1-V_1 duration, pooled across all speakers, for each tone class and syllable type.
(*For figure 6c,d see next page.*)

Given the previous, transcription-based descriptions of Luganda tone cited above, we would expect the early high-tone class in long syllables to be similar to the late high-tone class in rise onset interval and in peak delay, since both tones begin with a high tone on the first mora (as in Fig. 1). But we would expect fall delay to be greater in the late high-tone class than in the early high-tone class, reflecting the claim that the high tone in the late high-tone class extends to the second mora of the syllable, while that in the early high-tone class is limited to the first mora. If, on the contrary, both peak delay and fall delay are greater in the late high-tone class than in the early high-tone class, which would support a re-analysis of the contrast as a difference in timing of the whole f_0 rise-fall.

3. Results

Figures 6a–d give normalized time plots of the f_0 contour for each tone and syllable class in the study. For these graphs, f_0 was sampled at 10 equally-spaced intervals within the C_1-V_1 interval. Each f_0, value was converted to a z-score relative to the

DOI: 10.1159/000491073



speaker's range, to facilitate comparison across speakers with different pitch ranges. Each point in the graph represents the mean z-score for a given proportional time-point pooled across speakers, for early and late high-tone classes (except for CV, which only has an early tone).

It can be seen in all these graphs that f_0 rises from the syllable onset (0 on the x-axis) and hits a peak late in the C_1-V_1 interval (after 50% on the x-axis). In the CVC syllable in Figure 6b, the early and late lines are close to identical, but in CVVN and CVV (in Fig. 6c, d, respectively), the 2 lines diverge toward the end of the C_1-V_1 interval, with an earlier peak and an earlier decline in the early high-tone class (marked with circles) than in the late high-tone class (marked with triangles).

Mixed model analyses of the f_0 measurements were conducted, using the packages *lme4* (Bates et al., 2014) and *lmerTest* (Kuznetsova et al., 2014) in R (R Core Team, 2014). The alpha level was 0.05. The focus of this study is the contrast between early and late tones, which is limited to long syllables; so the main comparison was within the long syllables (CVC/CVVN/CVV), a subset including 1,173 items. The dependent variables in these comparisons were the f_0 rise lag, peak delay and fall delay, and the f_0 levels at each of the 4 landmarks in f_0 rise-fall contour. The fixed effects were C_1 duration, V_1 duration, C_2 duration, Tone (early/late), Syllable (CVC/CVVN/CVV), and the interaction of Tone and Syllable. In models of f_0 level, the duration of C_1, V_1, and C_2 were not expected to have an effect, so they were omitted from the fixed effects. Speaker and Item (sentence) were included as random intercepts, and the interactions of Speaker with the categorical fixed effects Tone and Syllable were included as random slopes.

In these tests, peak delay and fall delay were the only measurements of f_0 timing or level to show significant effects of tone class, that is, a significant difference in the measurement due to Tone or an interaction of Tone with Syllable. The measurements of f_0 at various f_0 landmarks were not significantly affected by tone class: f_0 at the rise onset, the rise offset, the fall onset, and the fall offset. There is no evidence, then, that the 2 tone categories differ in f_0 level at any landmark in the f_0 contour. Moreover, rise delay was not significantly affected by tone class; the only significant factor being the duration of C_1. We will, therefore, focus in the next 2 sections on the 2 measurements that do reflect the tone contrast: peak delay and fall delay.

3.1 Peak Delay

Figures 7a–d present for each syllable type the peak delay plotted against the C_1-V_1 interval, pooled across all speakers. The plotting characters (blue triangles for early high tone and red crosses for late high tone) for each token in the sample indicate where in the test syllable the f_0 rise ended. The blue line is the linear regression line fit to the early high-tone tokens, and the red line is the regression line for the late high-tone tokens. In each scatterplot, the solid black diagonal line marks x = y, so that points on that line represent tokens in which the rise offset was exactly at the end of V_1. Points below that line represent rise offset values before the end of V_1, and points above it represent tokens in which the rise continued into the following C_2.

One clear trend in all these graphs is that greater C_1-V_1 duration values tend to be associated with greater peak delay values. In all 4 syllable types, the plotting characters are densest near the x = y trend line, indicating that the rise offset tend to be close to the V_1-C_2 boundary, on one side or the other.

Long syllables (CVC, CVVN, CVV) are the only syllables with a contrast between early and late tones, so Figure 7b–d are the only ones with both early and late high-tone values. There is considerable overlap between the 2 classes in Figure 7b (CVC), but in Figure 7c (CVVN) and to a greater extent in Figure 7d (CVV), there is a discernible tendency for the peak delay values for the late class to be greater (i.e., higher on the plot) than those for the early class.

These facts deviate in 2 ways from the expectations, founded on previous literature. First, it would be expected from previous descriptions that the 2 tone classes would differ in fall delay, but not peak delay. Second, the contrast between the 2 tone classes has always been described as holding for all long syllable types, including CVC, CVVN, and CVV.

Figure 8 provides bar-plots of the mean peak delay, vowel duration, and relative peak delay by syllable type and tone class. The error bars represent the SD.

In Figure 8a, mean peak delay is greater in CVV syllables (167 ms) than in CVVN syllables (134 ms), and greater in turn in CVVN syllables than in CVC syllables (118 ms). This corresponds to the mean vowel duration in the 3 syllable types, as shown

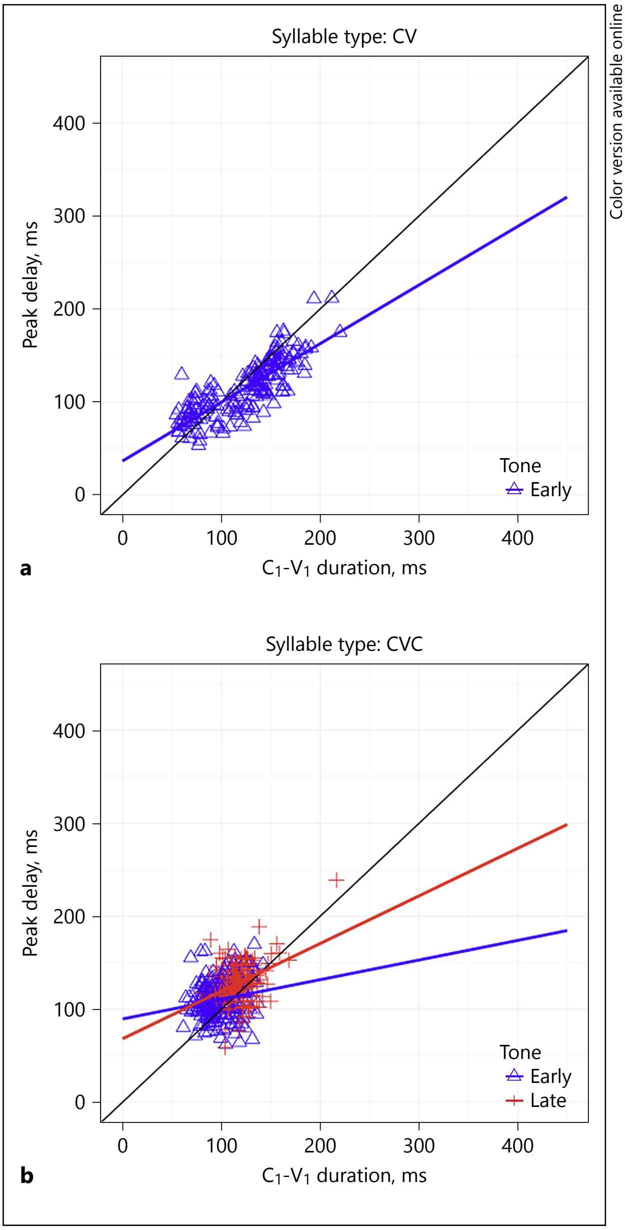
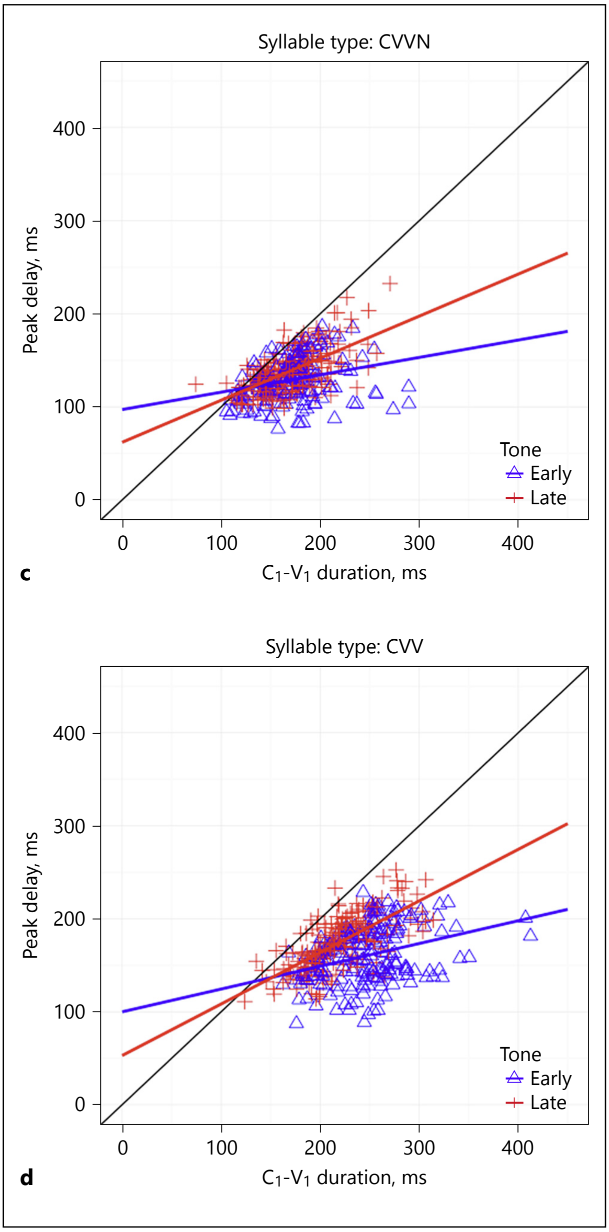


Fig. 7. Peak delay (ms) as a function of C_1-V_1 duration (ms) for each tone class and syllable type.
(*For figure 7c,d see next page.*)

in Figure 8b: CVV (174 ms) > CVVN (147 ms) > CVC (79 ms). The long vowel in an open syllable (CVV) is longer than the long vowel in a closed syllable (CVVN), and the long vowels in these 2 conditions are longer than the short vowel in CVC (Maddieson and Ladefoged, 1993; Hubbard, 1994; Vogel and Spinu, 2013). The syllable types thus show the same progression of mean values in peak delay as in vowel duration: CVV > CVVN > CVC.



Color version available online

The relation between segmental duration and peak delay can also be seen in the mean relative peak delay (peak delay divided by C_1-V_1 interval duration), shown in Figure 8c. CVC has the highest mean value (1.14), with a mean over 1 indicating that the rise offset is attained only in the C_2 interval. The mean is less in CVVN (0.80) and CVV (0.73), indicating that in these syllable types the rise offset is towards the end of the C_1-V_1 interval.

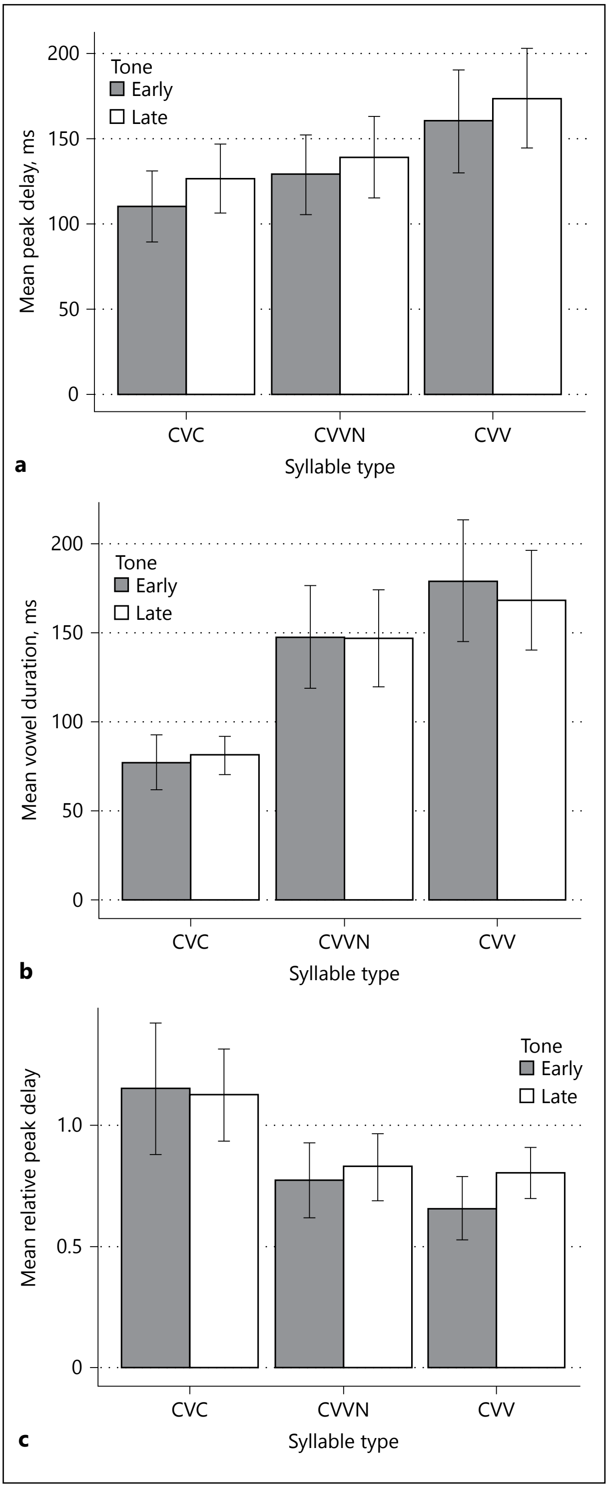


Fig. 8. a Mean peak delay (ms), (**b**) mean vowel duration (ms), and (**c**) mean relative peak delay, by syllable and tone class.

Table 4. Difference in mean relative peak delay between early and late tone classes in CVV syllables, by speaker

S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
0.22	0.08	0.18	0.16	0.19	0.17	0.14	0.08	0.11	0.11

Table 5. Fixed effects in a mixed-model analysis of peak delay (long syllables only)

Factor	β	df	*t* value	*p* value
Intercept	**39.23**	**62.5**	**5.9**	**<0.001**
C_1 duration, ms	**0.72**	**553.6**	**16.8**	**<0.001**
V_1 duration, ms	**0.37**	**1,131.0**	**15.7**	**<0.001**
C_2 duration, ms	**0.22**	**1,099.7**	**7.8**	**<0.001**
Syllable: CVV	–6.87	104.5	–1.1	0.29
Syllable: CVVN	4.5	71.3	0.8	0.45
Tone: Late	7.69	55.4	1.2	0.22
Syllable: CVV * Tone: Late	**21.89**	**54.5**	**2.8**	**0.007**
Syllable: CVVN * Tone: Late	3.12	52.0	0.4	0.67

The f_0 rise offset was later in the syllable in late high tone than in early high tone. In Figure 8a, the mean peak delay is greater in late high tone than in early high tone in each syllable type: CVC (early: 110 ms; late: 127 ms), CVVN (early: 129 ms; late: 139 ms), and CVV (early: 160 ms; late: 174 ms). However, we have seen in Figure 7 that the peak delay varies with C_1-V_1 interval duration, so it is important to also consider the relative peak delay. The mean relative peak delay is greater in late than in early high tones in CVVN syllables (early: 0.77; late: 0.83) and CVV syllables (early: 0.66; late: 0.80), but not in CVC syllables (early: 1.15; late: 1.14).

The greatest difference between early and late high tones is thus in CVV syllables, where the difference in mean relative peak delay is 14% of C_1-V_1 interval duration. This is a small difference, but it was consistent across speakers, as illustrated in Table 4, which gives the difference for each speaker in relative peak delay between early and late high tone in CVV syllables.

Each speaker had greater relative peak delay in late high tone than in early high tone in CVV syllables.

The results of the mixed-model analysis are presented in Table 5. Significant effects ($p < 0.05$) are indicated by bold face.

There are significant main effects for all the segmental duration factors: C_1, V_1 and C_2. The positive coefficients (β) indicate that a longer duration of any of these segmental intervals is associated with a greater peak delay. There is no major effect for either Tone or Syllable, but there is a significant interaction of the 2 (Syllable: CVV * Tone: Late). This reflects the fact, noted above, that the effect of Tone on peak delay is greater in CVV syllables than in other syllable types. This suggests that the difference between early and late high tone in the position of the rise offset might be limited to CVV syllables.

Table 6. Fixed effects in mixed-model analysis of peak delay in syllable subsets

	Intercept	C_1	V_1	C_2	Tone: Late
CVC	**50.03**	**0.84**	**0.20**	**0.21**	7.64
CVVN	12.16	**0.74**	**0.48**	**0.41**	**12.72**
CVV	**41.32**	**0.69**	**0.37**	0.10	**30.98**

To test this interpretation, we examined each syllable type separately, with mixed models in which the only fixed effects were the segmental duration factors (C_1, V_1, C_2) and the factor Tone. Table 6 provides the coefficient for each factor in the model of each subset, with significant factors indicated by bold face.

In each syllable subset, the duration of C_1 and V_1 were significant predictors of peak delay. C_2 was a significant factor in CVC and CVVN, but not in CVV. Tone was a significant factor only in the syllables with long vowels (CVV and CVVN). Peak delay in the CVC syllable type (modulo segment duration) was also greater in the late high-tone class than in the early high-tone class, as indicated by the positive coefficient, but the difference was not enough to reach the level of significance ($t_{18.5} = 1.9$, $p = 0.08$).

In the models in Tables 5 and 6, the C_2 interval in CVVN syllables was the interval of nasal murmur, corresponding to the syllable coda. An alternative annotation in which the whole nasal + consonant interval was included in C_2 was also tried as a predictor variable, but the alternative models with the nasal + consonant interval were found to have significantly inferior fit. The Akaike Information Criterion (AIC) for the model given in Table 5 (9,833.1) was lower than that of the alternative (9,842.0), and this difference between the models was significant in an ANOVA model comparison (Baayen et al., 2008): $\chi^2 = 8.9$, df = 0, $p < 0.001$.

In the model in Table 5, C_1, V_1, and C_2 are separate factors with separate weights. A simpler alternative model would replace these duration factors with one: the duration of the whole C_1-V_1-C_2 interval. But this simpler model had a higher AIC value, as well as a significantly worse fit in the ANOVA comparison. The same is true for alternative models with C_1-V_1 or V_1-C_2 as predictor variables. We conclude that the different segment intervals should be treated as separate factors, as in Table 5.

In the model in Table 5, the dependent variable was peak delay relative to the syllable onset. Alternative models (Schepman et al., 2006) in which the dependent variable was the interval between rise offset and V_1 onset, or the interval between rise offset and V_1 offset, were also considered. These alternative models were found to yield exactly the same pattern of results (in terms of which factors are significant) as the model in Table 5. This is perhaps unsurprising, given the strong correlation among the different versions of the peak delay measurement.

The preceding comparisons were among long syllables, the ones that have the contrast between early and late high tone within a syllable. To see how the timing of f_0 events compares between long and short syllables, we have to focus on the early high-tone class, since it is the only one in which the high tone is in the stem-initial syllable in all syllable types. There were 774 items in this subset of the data. The mean peak delay in CV syllables in this subset was 114 ms, and mean relative peak delay was 0.97. There are 4 syllable types included in the analysis (CV/CVC/CVVN/CVV), rather than just 3 as in the preceding analyses, so Syllable is broken down into 2

Table 7. Fixed effects in a mixed-model analysis of peak delay (early tones only)

Factor	β	df	*t* value	*p* value
Intercept	**46.05**	**87.3**	**7.6**	**<0.001**
C_1 duration	**0.78**	**473.9**	**15.9**	**<0.001**
V_1 duration	**0.24**	**753.0**	**8.4**	**<0.001**
C_2 duration	**0.21**	**749.2**	**6.6**	**<0.001**
Closed	1.91	59.6	0.3	0.77
Long vowel	4.94	54.3	0.7	0.52
Closed * Long vowel	8.19	44.9	1.0	0.33

dichotomous factors: Closed (in which CVC and CVVN are closed, and CV and CVV are open) and Long vowel (in which CVVN and CVV have long vowels, and CV and CVC do not). Such a breakdown was not possible in the previous comparison because there were no CV syllables (open syllables with a short vowel). Thus, the fixed effects in the model were C_1 duration, V_1 duration, C_2 duration, Closed, and Long Vowel. As in the previous models, Speaker and Item (sentence) were included as random intercepts, and the interactions of Speaker with the categorical fixed effects Closed and Long Vowel were included as random slopes. Table 7 presents the results of this comparison, with significant effects ($p < 0.05$) highlighted in boldface.

In this comparison, all the segmental duration variables (C_1, V_1, C_2) make a significant contribution to the model of f_0 peak timing, but neither of the syllable type categories (Closed, Long Vowel) do so. In a follow-up mixed-model analysis focusing just on the subset of CV syllables, all 3 segmental duration factors had significant effects: C_1, V_1, and C_2.

3.2 Fall Delay

In general, the fall onset was only slightly after the rise onset, with a mean difference of 4.9 ms between the 2 time-points. Quite often, there was no difference at all, because the rise offset coincided with the fall onset in the same time-point. It is not surprising, then, that the means for fall delay and relative fall delay, represented in Figure 9, echo exactly the patterns for peak delay given in Figure 8.

As in the case of peak delay, Figure 9a shows that fall delay was greater for syllables with greater vowel duration (CVV > CVVN > CVC), and was generally greater for late high-tone syllables than for early high-tone syllables. Figure 9b shows that, as with relative peak delay, relative fall delay is greatest for CVC, and it is greater in late high tone than in early high tone only in CVVN and CVV syllables.

Table 8 presents the results of a mixed-model analysis of fall delay. Boldface marks significant effects ($p < 0.05$).

As one would expect from the pattern of group means, these results are entirely parallel to those in Table 5 for peak delay. There are significant effects for the segmental duration variables C_1, V_1, and C_2, with the positive coefficients indicating that greater fall delay was associated with greater segment duration. There were no significant main effects of Syllable or Tone, but a significant interaction of Tone: Late with Syllable: CVV. This reflects greater difference between early and late tones in the CVV syllables compared to other syllable types.

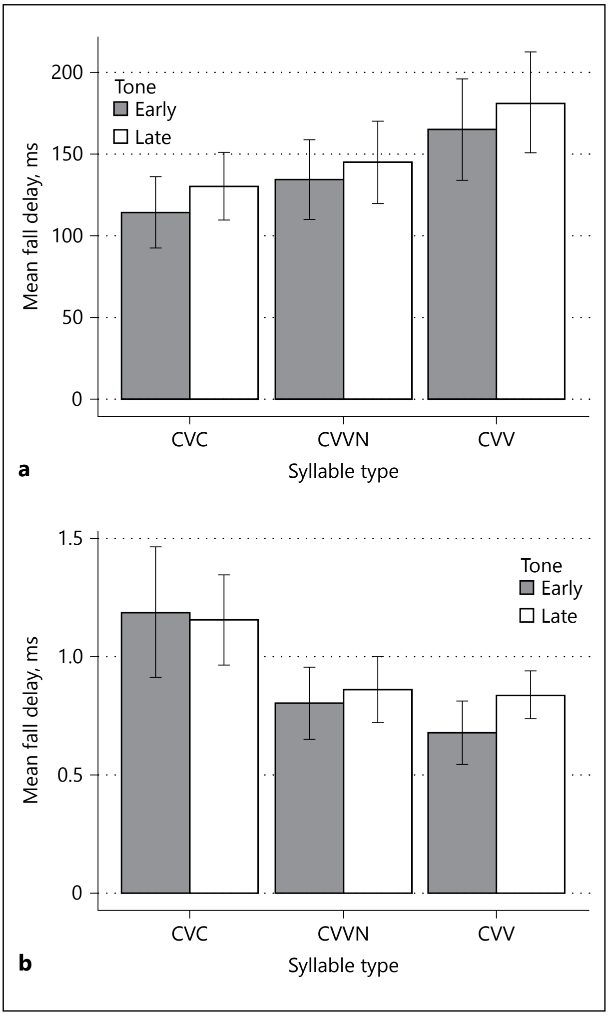


Fig. 9. a Mean fall delay (ms) and (**b**) mean relative fall delay, by syllable and tone class.

Table 8. Fixed effects in a mixed-model analysis of fall delay (long syllables only)

Factor	β	df	*t* value	*p* value
Intercept	**36.04**	**56.0**	**5.2**	**<0.001**
C_1 duration, ms	**0.73**	**473.5**	**17.0**	**<0.001**
V_1 duration, ms	**0.42**	**1,099.7**	**18.0**	**<0.001**
C_2 duration, ms	**0.24**	**1,096.9**	**8.5**	**<0.001**
Syllable: CVV	–11.44	102.1	–1.7	0.08
Syllable: CVVN	2.53	72.7	0.4	0.68
Tone: Late	6.58	51.4	1.0	0.33
Syllable: CVV * Tone: Late	**26.29**	**54.5**	**3.2**	**0.002**
Syllable: CVVN * Tone: Late	5.03	52.2	0.7	0.52

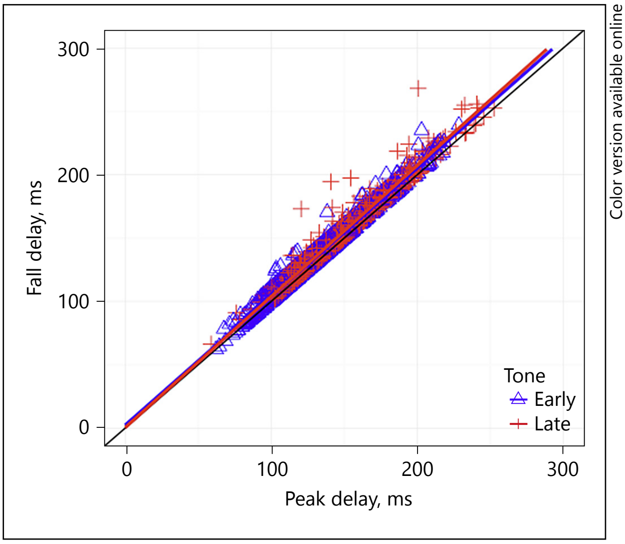


Fig. 10. Fall delay (ms) as a function of peak delay (ms) for each tone class (long syllables only).

Table 9. Alternative mixed-model regression analysis of fall delay (long syllables only)

Factor	β	df	*t* value	*p* value
Intercept	**3.02**	**139.5**	**2.9**	**0.004**
Peak delay, ms	**1.01**	**372.6**	**124.8**	**<0.001**
Syllable: CVV	0.48	43.1	0.5	0.59
Syllable: CVVN	1.21	28.4	1.4	0.18
Tone: Late	–0.22	24.3	–0.3	0.80
Syllable: CVV * Tone: Late	**3.06**	**29.8**	**2.9**	**0.006**
Syllable: CVVN * Tone: Late	0.61	37.5	0.6	0.56

If the rise offset and the fall onset are just different stages of the same coordinated rise-fall gesture, it would be expected that they would correlate with each other. In Figure 10, fall delay in long syllables is plotted as a function of peak delay, and it is evident that the 2 are closely associated in a linear manner. Here, as in Figure 6, the blue plotting characters and regression line mark the early high-tone class, and red plotting characters and regression line mark the late high-tone class. The black diagonal line marks x = y.

To test whether this relation accounts for the variation in fall delay, an alternative model was constructed in which fall delay was predicted on the basis of peak delay, rather than the segmental variables C_1, V_1, and C_2. The results for this model are presented in Table 9. Boldface marks significant effects ($p < 0.05$).

There is a significant effect of peak delay. There is also a significant interaction of Syllable: CVV and Tone: Late. The positive coefficient for this interaction indicates that the fall delay is greater in late-high CVV syllables by an average of 2.45 ms compared to other combinations of tone and syllable type, over and above the effect of peak delay. The AIC for the model in Table 9 (7,632.8) is lower than that for Table 8 (9,856.7), although the former has fewer predictors. The model in Table 9 is therefore more informative. The fall delay depends on the peak delay.

4 Discussion

4.1 Tone Alignment

This study has found that the timing of the f_0 peak (rise offset and fall onset) in Luganda depends on the duration of segments in the whole C_1-V_1-C_2 sequence in the syllable with a high tone. The peak delay was greater when vowel duration was greater, as has been established repeatedly in the literature (Van Santen and Hirschberg, 1994; Prieto et al., 1995; Arvaniti et al., 1998). But the current study has provided further support for the somewhat more controversial claim that the timing of the f_0 peak also depends on the duration of the prevocalic and postvocalic consonants (Van Santen and Hirschberg, 1994; Prieto et al., 1995; Arvaniti et al., 1998; Xu, 1998).

It was expected that the difference between moraic and non-moraic segments would be important in f_0 timing, due to the demonstrated importance of the distinction in Luganda phonology (Tucker, 1962; Hyman, 1992; Hyman and Katamba 2010), as well as the findings of previous studies of f_0 timing in other tone languages (Myers, 2003; Morén and Zsiga, 2006; Zsiga and Nitisaroj, 2007; DiCanio et al., 2014). However, contrary to these expectations, it was found in this study that the moraic status of a segment played no role in f_0 timing in Luganda. For example, the duration of the onset consonant C_1 was a significant factor in predicting peak delay in Luganda, although it never counts as moraic. The duration of V_1 was a significant factor, but there was no effect of the length category (long/short) of the vowel, over and above the vowel duration. Finally, the duration of C_2 was a significant predictor of the rise offset interval whether it was moraic (as in CVC or CVVN), or not (as in CV).

Likewise, we have found that syllable affiliation does not determine which segments influence the timing of the f_0 peak. The duration of C_2 has proven to be a significant factor in the timing of the f_0 peak, whether or not C_2 belongs to the same syllable as C_1 and V_1. In CVVN syllables, for example, C_2 belongs to the same syllable that contains C_1 and V_1, while in CV syllables, it belongs to the following syllable, and in CVC syllables, to both syllables. Yet in all 3 syllable types, the duration of C_2 contributes significantly to the model of f_0 timing.

In our model, the durations of C_1, V_1, and C_2 intervals are distinct independent variables, each with a different weight. It was found that such a model provided a better fit than simpler alternative models in which the whole C_1-V_1, V_1-C_2, or C_1-V_1-C_2 interval was the independent variable. These alternative models are not exactly ones in which syllable or rime duration was the predictor, since the CVC condition has a geminate consonant which is split between 2 syllables, with no clear acoustic boundary where one syllable ends and the next begins. However, the inferior fit of these models indicates that the segmental boundaries within the syllable and rime provide temporal information that is useful in predicting f_0 timing.

A reviewer notes that the rise offset comes later in CVC syllables (mean relative peak delay = 1.14) than in CV syllables (mean relative peak delay = 0.97), and suggests that this could reflect a sensitivity of f_0 timing to syllable structure, as in Dutch (Schepman et al., 2006) or Mandarin (Xu, 1998). However, this difference need not necessarily be attributed to the difference between CVC and CV in syllable structure. Within our model, it follows from the greater duration of C_2 in CVC syllables (mean = 128 ms) compared to CV syllables (mean = 60 ms). The predictor variable C_2 has a positive coefficient in the model in Table 5, so a longer C_2 is expected to be associated with a later f_0 rise offset.

The evidence provided here suggests that in Luganda the relation between moras and syllables on the one hand and the timing of f_0 events on the other is indirect. The association of a high tone with a mora or syllable identifies the C_1-V_1-C_2 sequence that is relevant to determining the timing of the f_0 contour that realizes that high tone. This is reminiscent of the analysis of Greek prenuclear peaks in Arvaniti et al. (1998), where the f_0 peak occurs just after the onset of the post-accentual syllable. Ladd (1983, 2000) argued accordingly that the phonological association of a tone to a syllable just defines the frame that phonetic alignment takes place in.

Such an approach would also be consistent with the view of the syllable being pursued within articulatory phonology, according to which the syllable is defined in terms of a network of binary coordination relations among articulatory gestures (Browman and Goldstein, 1988, 1995; Goldstein et al., 2006; Nam et al., 2009). Niemann et al. (2011) and Mücke et al. (2012) suggest that a high tone be represented as a laryngeal gesture phased with respect to the vowel constriction gesture (in-phase), which in turn is phased with other consonants in the syllable (in-phase or anti-phase). Such a coordination of articulatory gestures would be expected to yield acoustic correlates that are interdependent in duration, depending on the strength and directness of the phasing relation between them. This is what we observe in the relation of peak and fall delay to C_1, V_1, and C_2 in Luganda.

4.2 Early and Late High-Tone Categories

The authors of the transcription-based literature on Luganda tone (Tucker, 1962; Stevick, 1969; Hyman and Katamba, 1993, 2010) have agreed in their description of the tone contrast under investigation here, transcribing it as a contrast between falling and high-level tone, attested only in long syllables (CVC, CVVN, CVV). This description led us to expect that the key acoustic difference between the classes would be a greater fall delay in the high-level class compared to the falling class. Instead, we have found that, where there is a difference between the 2 classes, both peak delay and fall delay are greater in one class than in the other, and the fall delay depends on peak delay. The contrast is, therefore, more accurately described as one between an earlier rise-fall pattern and a later one, which we have referred to as early high tone and late high tone.

One way to represent such a contrast would be in terms of an association of H to the first mora of a syllable for early high tones and to the second mora for late high tones, as in Figure 1a and b. This would amount to representing the contrast as one between a falling tone and a rising tone. However, as can be seen in Figures 6 and 7, both tone types in fact involve a rise in f_0 over most of the vowel, with a peak attained in the second half of the vowel or after the end of the vowel. Moreover, the difference in peak delay between early and late high tone is 13 ms in CVV syllables, where the difference is greatest. This difference is significant and consistent across speakers, but it is unexpectedly small if we are thinking of it in terms of association to the first or the second half of a syllable. For reference, the mean duration of the C_1-V_1 interval in a long syllable in this sample is 170 ms.

An alternative to representing the contrast in terms of tone association would be to represent it in terms of a distinctive feature. In the feature system of Ladd (1983), a tone that is (+delayed peak) has a pitch contour aligned later in the syllable than a comparable tone that is (-delayed peak). Remijsen (2013) and Remijsen and Ayoker (2014) proposed a feature (late-aligned) with a similar interpretation, because they found that in Shilluk and Dinka the contrast between an early fall and a late fall was not limited to long syllables. Such a featural representation of the timing contrast in Luganda would be consistent with the fact that both classes are characterized by a rise in f_0 over the course of the vowel.

Another way in which the results in this study differed from the expected findings involved the types of syllables in which the tone contrast occurred. In the previous, transcription-based literature (Tucker, 1962; Stevick, 1969; Hyman and Katamba, 1993; 2010), authors agreed that the early-late contrast was limited to long syllables, but also that it was attested in every kind of long syllable: CVC, CVVN, and CVV. In this study, on the contrary, it was found that there was a significant phonetic difference between early and late high tones only in the CVV and CVVN syllable types. In CVC syllables, peak delay and fall delay were greater in the late high-tone class than in the early high-tone class; once segmental durations are taken into account (Table 6), but the difference was not significant. The contrast is thus limited to syllables with long vowels. This fits in with the proposal of Gordon (2001) and Zhang (2004) that contour tones are often limited in languages to syllables with longer sonorant intervals because only in such syllables there is a long enough interval to provide reliable perceptual evidence of a timing contrast.

Even in the CVV syllable type, in which the timing difference between early and late high tones is the greatest, the difference is surprisingly small, though consistent and significant. In early high tones, the rise offset occurred on average at 66% of the C_1-V_1 interval (mean relative peak delay), and in late high tones at 80% of that interval – a difference of 14%. The corresponding difference in the mean peak delay between the 2 tone categories was 13 ms.

For comparison, the difference in the mean relative peak delay between falling and rising tones in the related language Kinyarwanda (Myers, 2003) ranged for different speakers from 40 to 50%, and the difference in the mean peak delay between the 2 tones ranged from 106 to 135 ms. Pisoni (1977) showed that the onset of 2 pure tones of different frequencies had to be at least 20 ms apart for listeners to be able to distinguish which one began first (cf. Hirsh, 1959; Stevens, 1997). House (2004), generalizing over previous studies of tone perception in speech, estimated that the difference had to be at least 50 ms to be perceptible.

How can we account for the differences between our results, and the expectations founded on the descriptions of the tone contrast in the literature (Tucker, 1962; Cole, 1967; Stevick, 1969; Hyman, 1982; Hyman and Katamba, 1993, 2010)? We could just point out that our study was the first controlled experimental study based on objective acoustic measurements, while previous work on the tone contrast was based on subjective auditory impressions. We do think that experimental data are more reliable than transcription data, but if generations of observers of the language were all mistaken, what accounts for the inter-transcriber consistency in their descriptions of the contrasting tone categories, and in their assignment of items to tone categories?

It is unlikely that the participants in the current study simply spoke a different regional dialect of Luganda than those consulted in previous studies. The participants in the current study were from all over the Luganda-speaking area of Uganda, and showed similar results as discussed earlier.

However, it could be that there is an ongoing diachronic change affecting the 2 tone classes that distinguishes the speakers who were the consultants for earlier studies, and the contemporary Kampala residents who were the consultants for this study. That is, it might be that there used to be a more robust contrast between the 2 classes, reflected in those earlier studies, but that it has been reduced over the recent generations, particularly in the CVC syllable types.

If there is an ongoing merger of the 2 tone classes, the small but significant difference between them could be due to the contrast being in a state of near merger or incomplete neutralization (Port and O'Dell, 1985; Di Paolo and Faber, 1990; Labov, 1994; Warner et al., 2004; Yu, 2007). Contrasting sounds in near merger cannot be reliably distinguished by native speakers, but are still associated with small but significant acoustic differences, and discriminability at greater than chance level. Ernestus and Baayen (2006) proposed a plausible account of near merger in terms of co-activation of related items or exemplars in the lexicon.

If there is an ongoing merger in progress, it would be expected that older speakers of the language would tend to have a greater acoustic difference between early and late tone than younger speakers. We checked this in our sample of 10 speakers, and found no such effect of age. However, the variance in age in the sample is small, and could well be too little to make such an effect detectable.

In the case of an ongoing merger, it would also be expected that older recordings would show more difference than contemporary ones. One piece of suggestive evidence that is compatible with this view is found in the recorded materials for the Luganda language course of Kamoga and Stevick (1968), available online thanks to the Live Lingua Project (https://www.livelingua.com/fsi-luganda-course.php). In this small older sample of Luganda, the distinction between early and late tone is considerably easier to hear than in our contemporary sample. In a sample of the CVV syllables in Kamoga and Stevick (1968), the mean relative peak delay was 0.46 ms for early tones and 0.72 ms for late tones, a difference of 26%. The mean peak delay in CVV syllables was 140 ms for early tones and 251 ms for late tones, a difference of 111 ms. In this sample, the contrast was robust in both closed CVVN syllables and open CVV syllables.

This is a very small sample, with just one speaker, and the materials and the conditions of the recording are quite different from those in our study. As a reviewer quite rightly points out, the greater distinctiveness of the tone categories in this sample could be at least partially due to a greater care of pronunciation for these pedagogical materials. A fair test would require an acoustic production study involving a sample of Luganda speakers with a broader range of ages than in the present study, and a perceptual listening test to determine whether speakers currently can distinguish these 2 tone classes.

Acknowledgements

The work reported here was partly supported by a grant from the Department of Linguistics at the University of Texas at Austin. There was no conflict of interest in this work. We would like to thank the following for their help with this project: Larry Hyman, Francis Katamba, Richard Meier, Judith

Nakayiza, Kizza Mukasa Jackson, Patrice Sanyondo, Medadi Ssentanda, Nkonge Douglas Kiyinikibi, Mariam Kyabangi, Racheal Nakirayi, Christopher Buyondo, Lukwago Moses, and the associate editor and 2 reviewers for *Phonetica*.

References

Arvaniti A, Ladd DR, Mennen I (1998): Stability of tonal alignment: the case of Greek prenuclear accents. J Phonet 26:3–25.
Arvaniti A, Ladd DR, Mennen I (2006): Tonal association and tonal alignment: evidence from Greek polar questions and contrastive statements. Lang Speech 49:421–450.
Baayen RH, Davidson DJ, Bates DM (2008): Mixed-effects modeling with crossed random effects for subjects and items. J Mem Lang 59:390–412.
Banti G (1988): Two Cushitic systems: Somali and Oromo nouns; in van der Hulst H, Smith N (eds): Autosegmental Studies on Pitch Accent. Dordrecht, Foris, pp 11–50.
Bates D, Maechler M, Bolker B, Walker S (2014): Lme4: Linear Mixed-Effects Models Using Eigen and S4. http://www.R-project.org.
Blevins J (1993): A tonal analysis of Lithuanian nominal accent. Language 69:237–273.
Boersma P, Weenink D (2013): Praat: doing phonetics by computer. Version 5.3.53. http://www.praat.org/.
Browman C, Goldstein L (1988): Some notes on syllable structure in Articulatory Phonology. Phonetica 45:140–155.
Browman C, Goldstein L (1995): Gestural syllable position effects in American English; in Bell-Berti F, Raphael L (eds): Producing Speech: Contemporary Issues. New York, AIP Press, pp 19–33.
Bruce G (1977): Swedish Word Accents in Sentence Perspective. Lund, CWK Gleerup.
Chebanne A, Creissels D, Nkhwa H (1997): Tonal Morphology of the Setswana Verb. München, LINCOM Europa.
Clements, GN (1986): Compensatory lengthening and consonant gemination in LuGanda; in Wetzels L, Sezer E (eds): Studies in Compensatory Lengthening. Dordrecht, Foris, pp 39–77.
Cole, D (1967): Some Features of Ganda Linguistic Structure. Johannesburg, Witwatersrand University Press.
DiCanio C, Amith JD, Castillo Garcia R (2014): The Phonetics of Mora Alignment in Yoloxóchitl Mixtec. Paper Presented at the 4th International Symposium on Tonal Aspects of Languages, Nijmegen, The Netherlands. http://www.isca-speech.org/archive/tal_2014/papers/tl14_203.pdf.
D'Imperio, M (2000): The Role of Perception in Defining Tonal Targets and their Alignment. PhD Dissertation, Ohio State University. https://etd.ohiolink.edu/!etd.send_file?accession = osu1243021045&disposition = inline.
D'Imperio M, House D (1997): Perception of questions and statements in Neapolitan Italian; in Kokkinakis G, Fatotakis N, Dermatas E (eds): Proceedings of Eurospeech. 5, Vol. 1, pp 251–254. http://www.isca-speech.org/archive/eurospeech_1997/e97_0251.pdf.
Di Paolo M, Faber A (1990): Phonation differences and the phonetic content of the tense-lax contrast in Utah English. Lang Var Change 2:155–204.
Ernestus M, Baayen RH (2006): Paradigmatic effects in auditory word recognition: the case of alternating voice in Dutch. Lang Cogn Process 22:1–24.
Frota S (2002): Tonal association and target alignment in European Portuguese nuclear falls; in Gussenhoven C, Warner N (eds): Laboratory Phonology 7. Berlin, Mouton de Gruyter, pp 387–418.
Goldsmith, J (1990): Autosegmental and Metrical Phonology. Oxford, Basil Blackwell.
Goldstein L, Byrd D, Saltzman E (2006): The role of vocal tract gestural action units in understanding the evolution of phonology; in Arbib MA (ed): Action to Language via the Mirror Neuron System. Cambridge, Cambridge University Press, pp 215–249.
Gordon, M (2001): A typology of contour tone restrictions. Stud Lang 25:405–444.
Gussenhoven C (2000): The lexical tone contrast of Roermond Dutch in Optimality Theory; in Horne M (ed) Prosody: Theory and Experiment. Amsterdam, Kluwer, pp 129–167.
Herbert, R (1975): Reanalyzing prenasalized consonants. Stud Afr Linguist 6:105–123.
Hirsh IJ (1959): Auditory perception of temporal order. J Acoust Soc Am 31:759–767.
House D (2004): Pitch and alignment in the perception of tone and intonation; in Fant G, Fujisaki H, Cao J, Xu Y (eds): From Traditional Phonology to Modern Speech Processing. Beijing, Foreign Language Teaching and Research Press, pp 189–204.
Howie JM (1974): On the domains of tone in Mandarin: some acoustical evidence. Phonetica 30:129–148.
Hubbard K (1995): "Prenasalised consonants" and syllabic timing: evidence from Runyambo and Luganda. Phonology 12:235–256.
Hyman L (1982): Globality and the accentual analysis of Luganda tone. J Linguist Res 2:1–40.
Hyman L (1992): Moraic mismatches in Bantu. Phonology 9: 255–265.
Hyman L (1993): Register tones and tonal geometry; in van der Hulst H, Snider K (eds): The Phonology of Tone: The Representation of Tonal Register. Berlin, Mouton de Gruyter, pp 75–108.
Hyman L, Katamba F (1993): A new approach to tone in Luganda. Language 69:34–67.

Hyman L, Katamba F (2010): Tone, syntax and prosodic domains in Luganda; in Downing L, Rialland A, Beltzung JM, Manus S, Patin C, Riedel K (eds): Papers from the Workshop on Bantu Relative Clauses, ZAS Papers in Linguistics 53. Berlin, Zentrum für Allgemeine Sprachwissenschaft, pp 69–98.
Hyman L, Katamba F, Walusimbi L (1987): Luganda and the Strict Layer Hypothesis. Phonology Yearbook 4: 87–108.
Kamoga FK, Stevick EW (1968): Luganda Basic Course. Washington, Foreign Service Institute.
Katamba F (1974): Aspects of the Grammar of Luganda. PhD Dissertation, University of Edinburgh.
Kenstowicz M (1970): On the notation of vowel length in Lithuanian. Pap Linguist 3:73–113.
Kohler K (1987): Categorical pitch perception; in Eek A (ed): Proceedings of the 11th International Congress of the Phonetic Sciences. Tallinn, Academy of Sciences of the Estonian S.S.R., pp 331–333.
Kuznetsova A, Brockhoff PB, Bojesen RH (2014): Tests in Linear Mixed Effects Models. http://www.R-project.org.
Labov W (1994): Principles of Linguistic Change: Internal Factors. Oxford, Blackwell.
Ladd DR (1983): Phonological features of intonational peaks. Language 59:721–759.
Ladd DR (2000). Bruce, Pierrehumbert, and the elements of intonational phonology; in Horne M (ed): A Festschrift for Gösta Bruce. Dordrecht, Kluwer, pp 37–50.
Leben W (1978): The representation of tone; in Fromkin V (ed): Tone: A Linguistic Survey. New York, Academic, pp 177–219.
Maddieson I, Ladefoged P (1993): Phonetics of partially nasal consonants; in Huffman M, Krakow R (eds): Nasals, Nasalization and the Velum. Orlando, Academic Press, pp 251–301.
McCawley J (1978): What is a tone language?; in Fromkin V (ed): Tone: A Linguistic Survey. New York, Academic, pp. 113–131.
Morén B, Zsiga E (2006): The lexical and post-lexical phonology of Thai tones. Nat Lang Ling Theory 24: 113–178.
Mücke D, Nam H, Hermes A, Goldstein L (2012): Coupling of tone and constriction gestures in pitch accents; in Hoole P, Bombien L, Pouplier M, Mooshammer C, Kühnert B (eds): Consonant Clusters and Structural Complexity. Berlin, De Gruyter, pp 205–230.
Myers S (1999): Tone association and f_0 timing in Chichewa. Stud Afr Linguist 28:215–239.
Myers S (2003): F0 timing in Kinyarwanda. Phonetica 60:71–97.
Nam H, Goldstein L, Saltzman E (2009). Self-organization of syllable structure: a coupled oscillator model; in Pellegrino F, Marsico E, Chitoran I, Coupé C (eds): Approaches to Phonological Complexity. Berlin, Mouton de Gruyter, pp 299–328.
Niemann H, Mücke D, Nam H, Goldstein L, Grice M (2011): Tones as Gestures: The Case of Italian and German. Proceedings of ICPhS XVII. pp 1486–1489. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Niemann/Niemann.pdf.
Odden D (1995): Tone: African languages; in Goldsmith J (ed): The Handbook of Phonological Theory. Blackwell, Oxford, pp 444–475.
Odden D (1996): The Phonology and Morphology of Kimatuumbi. Oxford, Oxford University Press.
Pierrehumbert J, Steele SA (1989): Categories of tonal alignment in English. Phonetica 46:181–196.
Pisoni D (1977): Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. J Acoust Soc Am 61:1352–1361.
Port R, O'Dell M (1985): Neutralization of syllable-final devoicing in German. J Phonet 13:455–471
Prieto P, van Santen J, Hirschberg J (1995): Tonal alignment patterns in Spanish. J Phonet 23:429–451.
Purcell E (1976): Pitch peak location and the perception of Serbo-Croatian word tone. J Phonet 4:265–270.
R Core Team (2014): R: A Language and Environment for Statistical Computing, Version 3.1.2. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
Remijsen B (2013): Tonal alignment is contrastive in falling contours in Dinka. Language 89:297–327.
Remijsen B, Gwado Ayoker O (2014): Contrastive tonal alignment in falling contours in Shilluk. Phonology 31:435–462.
Rietveld T, Gussenhoven C (1995): Aligning pitch targets in speech synthesis: effects of syllable structure. J Phonet 23:375–385.
Schepman A, Lickley R, Ladd DR (2006): Effects of vowel length and 'right context' on the alignment of Dutch nuclear accents. JPhonetic 34:1–28.
Silverman K, Pierrehumbert J (1990): The timing of prenuclear high accents in English; in Kingston J, Beckman M (eds): Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech. Cambridge, Cambridge University Press, pp 72–106.
Smiljanić R (2006): Early vs. late focus: Pitch-peak alignment in two dialects of Serbian and Croatian; in Goldstein L, Whalen D, Best C (eds): Laboratory Phonology 8. Berlin, Mouton de Gruyter, pp 495–518.
Stevens K (1997): Articulatory-acoustic-auditory relationships; in Hardcastle W, Laver J (eds): Handbook of Phonetic Sciences. Oxford, Blackwell, pp 462–506.
Stevick E (1969): Pitch and duration in Ganda. J Afr Lang 8:1–28.
Tucker AN (1962): The syllable in Luganda: a prosodic approach. J Afr Lang 1:122–166.
Tucker AN (1967): Introduction; in Snoxall RA (ed): Luganda-English Dictionary. Oxford, Oxford University Press, pp. xiii–xxxv.
Van Santen JP, Hirschberg J (1994): Segmental Effects on Timing and Height of Pitch Contours. 3rd International Conference on Spoken Language Processing (ICSLP 94), Yokohama, Japan. September 18–22, 1994. http://www.isca-speech.org/archive.

Vogel I, Spinu L (2013): Vowel length in Luganda. Proceedings of Meetings on Acoustics 12, 060005 (2013). http://dx.doi.org/10.1121/1.4812440.
Warner N, Jongman A, Sereno J, Kemps R (2004): Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. J Phonet 32:251–276.
Woo N (1972): Prosody and Phonology. Bloomington, Indiana University Linguistics Club.
Xu Y (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. Phonetica 55:179–203.
Xu Y (2001): Fundamental frequency peak delay in Mandarin. Phonetica 58:26–52.
Yip M (2002): Tone. Cambridge, Cambridge University Press.
Yu A (2007): Understanding near mergers: the case of morphological tone in Cantonese. Phonology 24:187–214.
Zhang J (2004): The role of constraint-specific and language-specific phonetics in contour tone distribution; in Hayes B, Kirchner R, Steriade D (eds): Phonetically Based Phonology. Cambridge, Cambridge University Press, pp 157–190.
Zsiga E, Nitisaroj R (2007): Tone features, tone perception, and peak alignment in Thai. Lang Speech 50:343–383.