

INTERACTION BETWEEN LEXICAL TONE AND INTONATION IN KINSHASA LINGALA

Yeong Woo Park, Francisco Torreira, Heather Goad

McGill University

yeongwoo.park@mail.mcgill.ca; francisco.torreira@mcgill.ca; heather.goad@mcgill.ca

ABSTRACT

We investigate the phonetic realization of lexical tones in phrase-final syllables in Kinshasa Lingala, where intonational and lexical targets coincide and potentially conflict. We analyze read speech data from nine Kinshasa Lingala speakers. By employing stringent and objective criteria to identify intonational tones, we observe the following: (1) We confirm the existence of intonational targets, with extra high and extra low f_0 values. (2) Contrasts between high and low lexical tones phrase-finally are unequivocally preserved when preceding a high intonational tone, but not when preceding a low intonational tone. (3) Different patterns of coarticulation arise in specific combinations of lexical and intonational tones. (4) Speakers often use devoicing in conjunction with phrase-final low tonal targets, as well as when phrase-final syllables contain /i/ regardless of its associated tone.

Keywords: lexical tone, intonation, prosody, phrase-final position, Kinshasa Lingala

1. INTRODUCTION

The interaction of lexical tone and intonation is an important aspect of linguistic systems, given that both implicate f_0 but serve different linguistic functions. Hyman and Monaka [6] propose that the two may interact in three different ways across languages: (1) *accommodation*, where lexical and intonational tones co-occur but do not affect each other; (2) *submission*, where intonational tones “invade and override” lexical tones; and (3) *avoidance*, where intonational tones fail to surface.

We propose that the *accommodation* pattern can be further divided into two possibilities: (1) *sequential realization*, where lexical and intonational tones are realized sequentially, possibly with substantial coarticulation; and (2) *coalescence*, where both lexical and intonational contrasts are preserved but are coalesced into a single tonal target (e.g., a mid tone), from which lexical tones can be reconstructed. The present study focuses on how lexical and intona-

tional tones interact in the Kinshasa dialect of Lingala, a Bantu language with a two-way lexical tonal contrast, and what appears to be a two-way intonational contrast in phrase-final position.

Previous descriptions of Lingala ([4], [5], [9], [11], [13]) report that the language has high and low lexical tones,¹ (H and L from now on), respectively. Guthrie [4] further claims that the realization of this lexical tonal contrast is affected by an extra-low tone that occurs phrase-finally, which we attribute to the presence of a low intonational tone. This type of coarticulatory pattern has been described as *superimposition* in other Bantu languages (e.g., Embosi [10], Limbum [3]).

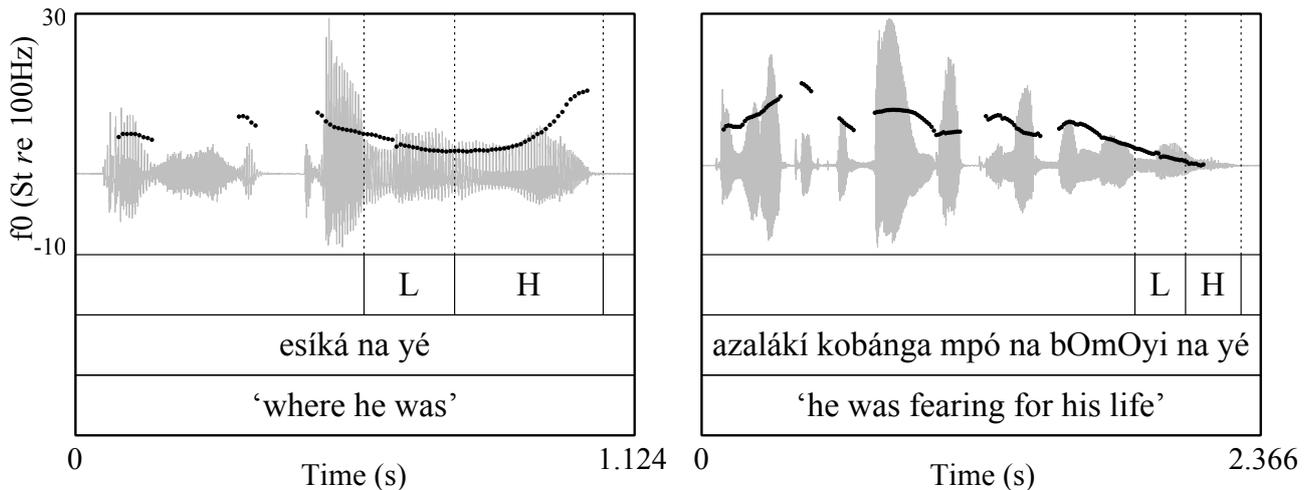
In our data, we have observed very low f_0 values at the end of many utterances where the last syllable carries an H tone, as well as unexpectedly high f_0 values in cases where the phrase-final syllable carries a L tone. Fig. 1 illustrates both of these patterns in the words *na yé* ‘his’, where the last syllable carries a high lexical tone, but opposite f_0 trajectories are observed. While this eliminates the possibility of the *avoidance* pattern for Kinshasa Lingala - at least for phrase-final H lexical tones - three options remain regarding the realization of the contrast between high and low lexical tones in phrase-final position: 1) *submission*; 2) *sequential realization*; and 3) *coalescence*. We probe these different possibilities by inspecting the phonetic realization of lexical tones in phrase-final syllables in Kinshasa Lingala.

2. METHODS

2.1. Data collection

Read speech data were collected from 9 native speakers (5 males, 4 females) of Kinshasa Lingala. Each speaker read one of two fables, which were originally written in standard Lingala. They were adapted for Kinshasa Lingala, and then written in a form of writing commonly used in the Democratic Republic of Congo, which does not mark tonal contrasts. Data were collected in a sound-attenuated room at McGill University using Shure SM10A head-mounted unidirectional microphones

Figure 1: Opposite f0 patterns in semitones and waveforms for the same sequence of words *na yé* ‘his’ in phrase-final position, with syllable boundaries and lexical tones marked for the two last syllables.



connected to a Presonus AudioBox iTwo audio-interface (48 kHz, 32-bit).

2.2. Data preparation

For each recording, interpausal units (IPU) delimited by silences of at least 200ms were identified automatically in Praat [1] using the “To TextGrid (silences)” function. All lexical tones in the fables were annotated by a native speaker of Kinshasa Lingala with linguistic training. The last two syllables in each IPU were labelled with their corresponding tones. IPUs with non-canonical speech (e.g., mispronunciations, sudden stops, incomplete phrases, and clearly perceptible disfluencies) and non-speech noise were excluded. Thirty-nine IPUs where measured f0 values were 10 or more semitones higher or lower than the speaker’s mean f0 value were manually examined and excluded as f0 tracking errors.

Since the intonation of Kinshasa Lingala is still not well understood, we adopted stringent and conservative criteria in annotating intonational tones. Any f0 value in an IPU-final syllable that was higher or lower than the previously identified H or L lexical target in that IPU was labelled H% and L%, respectively, after the IPUs-final syllable. As a result, 140 cases of H% (119 H and 21 L) and 261 cases of L% (124 H and 137 L) were identified. The remaining 155 IPUs were labelled as M% and were excluded as they do not meet the criteria proposed for the present study.

The last syllable of each target IPU was divided into three quantiles of equal duration and labelled as Q1, Q2, and Q3. Mean f0 values in Q1 and Q2, and the last available f0 value in Q3, were measured

in semitones relative to 100 Hz. We normalized f0 measures relative to each speaker’s mean f0 value, calculated over the entire recording.

Given that Lingala syllables always end in a vowel, devoicing was coded when Praat did not find any f0 value in the second half of the final syllable.

2.3. Quantitative analyses

In order to probe the f0 trajectories observed throughout IPU-final syllables, we fitted a series of mixed-effects linear regression models using the lmerTest package in R [8]. We also fitted a logistic mixed-effects regression model to test the statistical significance of IPU-final lexical tones on the occurrence of final devoicing.

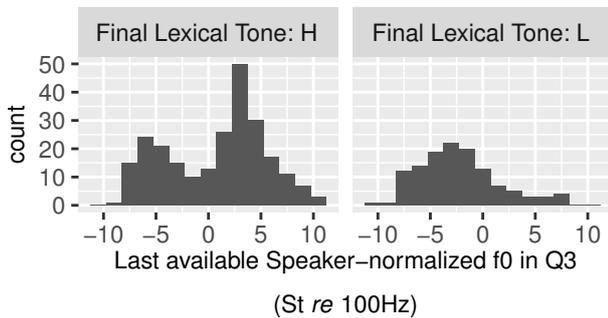
3. RESULTS

3.1. Fundamental frequency

The histograms in Fig. 2 show distributions of the last available speaker-normalized f0 values in the last quantile of all IPU-final syllables (including ones marked with M%), carrying H (left panel) and L (right panel) tones. The left panel exhibits a non-unimodal, and more of a bimodal, distribution, while the right panel is less clear. Such non-unimodal distributions suggest the influence of additional tonal targets that may be attributed to intonational tones.

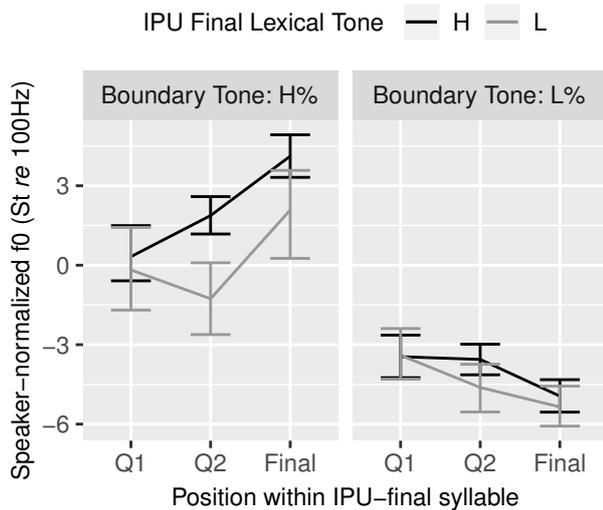
Fig. 3 shows the trajectory patterns of speaker-normalized f0 values measured in each of the three quantiles within IPU-final syllables for both lexical and intonational tones. The error bars (CI 95%) were calculated using the “mean_cl_boot” function,

Figure 2: Distribution of the last available speaker-normalized f0 in the last quintile of all IPU-final syllables (bin width = 1.5).



which does not assume normality, in the R ggplot2 [12] package.

Figure 3: Trajectory patterns of speaker-normalized average f0 values throughout target IPU-final syllables.



While the overall trends show a substantial influence of intonational tones throughout the syllable, the lexical tonal contrast appears to be preserved in certain positions. Specifically, when followed by H%, H tones have higher f0 values and L tones have lower f0 values in Q2, compared to the values in Q1. In contrast, when the intonational tone is L%, f0 is not higher for H tones in Q2 compared to the values in Q1, while L tones show a considerable drop in f0 in the same position. Also, the contrast between H and L tones in Q2 is less evident when followed by L% as compared to when followed by H%. As for Q3, the difference between H and L tones in the presence of H% is smaller when compared to the difference found in Q2, with some overlap observed;

and the difference between H and L tones seems to be neutralized in the presence of L%, with considerable overlap.

To test the statistical significance of these patterns, we fitted a series of linear mixed effects models, predicting f0 in Q1, Q2, and Q3 separately for each intonational tone. We included the lexical tones of the IPU-final syllable and the preceding syllable as fixed predictors, and speaker as a random intercept.

When the intonational tone is H%, f0 is lower in the presence of a L tone in Q2 ($\beta = -3.43$, $t = -7.308$, $p < 0.001$) and Q3 ($\beta = -2.04$, $t = -2.832$, $p < 0.01$). Additionally, the lexical tone of the preceding syllable has significant effects in Q1 ($\beta = -3.142$, $t = -7.692$, $p < 0.001$) and Q2 ($\beta = -1.47$, $t = -4.303$, $p < 0.001$). The effect sizes show what appears to be a local effect that weakens as the temporal distance from the preceding syllable increases. These results indicate that, when the intonational tone is H%, the lexical tonal contrast is preserved in Q2 and Q3, despite significant coarticulatory effects from adjacent tonal targets.

When the intonational tone is L%, L tones have slightly lower f0 than H tones in Q2 ($\beta = -1.188$, $t = -3.09$, $p < 0.01$), but not in Q3. The lexical tone of the preceding syllable has a significant effect in Q1 ($\beta = -1.712$, $t = -4.056$, $p < 0.001$), but not in Q2. These results suggest that, when the intonational tone is L%, the lexical tonal contrast is preserved at least in Q2, and that it is effectively neutralized towards the end of the syllable.

3.2. Final devoicing

Recall from section 2.2 that devoicing was coded when Praat failed to find any f0 value in the second half of the final syllable. Table 1 shows the occurrence of IPU-final devoicing organized by intonational tones and lexical tones.

Table 1: Occurrence of IPU-final devoicing

Devoiced: 119 IPUs			
		Lexical Tone	
		H	L
Intonational Tone	H%	14	0
	L%	35	70

Further examination revealed that out of the 119 IPUs coded as devoiced, all 14 IPUs with H% end in the vowel /i/, which suggests that there may be an independent devoicing mechanism involving that vowel. Therefore, we excluded all 34 IPUs with final devoiced /i/ from the analysis of devoicing, in-

cluding 20 IPUs with L% (11 H and 9 L).

Eighty-five IPUs remained after removal of final devoiced /i/, 61 IPUs with L tone on the last syllable and 24 IPUs with H tone. To test for statistical significance, we fitted a mixed effects logistic regression model with the occurrence of devoicing as a response variable, IPU-final lexical tone as a fixed predictor, and speaker as a random intercept. We found that lexical tone has a significant effect on the occurrence of devoicing ($\beta = 2.112$, $z = 6.549$, $p < 0.001$). Our observations suggest that final devoicing is not driven by some phonological condition tied to tonal specification, but rather indirectly relates to low f0 targets, which are more frequent with L tones in phrase-final position.

4. DISCUSSION AND CONCLUSION

In this paper, we investigated the phonetic realization of phrase-final lexical tones in different intonational contexts in a corpus of read speech in Kinshasa Lingala. Our focus was on whether lexical tonal contrasts are preserved in this position. Our results indicate that the two-way lexical contrast is preserved in different phrase-final intonational contexts despite substantial coarticulatory pressure from adjacent tonal targets (i.e. preceding lexical tones and following intonational tones). The contrast is most prominent in Q2 of the last syllable of the target IPUs. While the tonal contrast is preserved in Q3 when followed by H%, the difference in f0 values is smaller than in Q2. Interestingly, when the intonational tone is L%, the lexical contrast appears to be preserved in Q2, though less robustly compared to H%, and neutralized in Q3 with substantial overlap of f0. That is, we observe distinct context-dependent patterns of preservation and possible neutralization of lexical tones in different location within the same phrase-final syllable, presumably driven by the same type of tonal target (i.e. intonational tone). The preservation of the lexical tonal contrast cannot be attributed to either the *coalescence* or the *submission* pattern, where we would expect a non-contrastive f0 trajectory to emerge for each intonational tone.

Our results also indicate distinctive directionalities of influence from preceding lexical tones and following intonational tones on the realization of phrase-final lexical tones. Preceding lexical tones have their strongest effect in Q1, but the effect weakens throughout the rest of the syllable, such that it is non-significant in Q3. This pattern may be attributed to a carry-over coarticulatory effect, which affects the realization of a following tone. On the other hand, intonational tones have a sub-

stantial influence in all quantiles throughout IPU-final syllables, thereby impacting the overall f0 trajectory. Such global trends may be interpreted as anticipatory coarticulation, in which speakers begin approaching f0 targets associated with L% or H% while during the production of lexical tones. Since this pattern of tonal interaction becomes stronger as the temporal distance from the intonational tones decreases, the pattern can be attributed to a local coarticulatory effect, rather than to coalescence of two tones into a single tonal target. Lastly, the smaller but nonetheless clear tonal contrasts preserved in Q3 in the presence of H% can be attributed to the carry-over effect from the IPU-final L tones. In syllables containing H%, therefore, the phonetic realizations of lexical and intonational tones appear to be substantially blended in f0.

Together with the f0 trajectories observed throughout the three quantiles, the results of the mixed-effects linear regression models are indicative of *sequential realization* of lexical tones and following intonational tones in IPU-final syllables. That is, H and L tones exhibit contrastive trajectories from Q1 to Q2, regardless of the following intonational tone, and this contrast is reduced towards the end of the syllable, where intonational tones have the strongest local effect, in addition to their global effect throughout the final syllable.

Our results also indicate that the probability of IPU-final devoicing significantly increases when a L tone is present on an IPU-final syllable. Yet, together with the occurrence pattern reported in section 3.2, the results suggest that devoicing is likely not a phonological phenomenon driven by the tonal specification of phrase-final syllables, as a near categorical distribution would then be expected. In contrast, 28.24% of devoicing occurs with a H tone. Taking the global and local effects of L% in IPU-final syllables into consideration, one possible explanation is that IPU-final devoicing is closely related to low f0 values rather than to a specific tonal specification. Since speakers of other languages have been shown to use non-f0 cues to mark tonal targets (e.g., [2] & [7]), it is possible that Lingala speakers shift acoustic dimensions (i.e. from f0 to voicing) to cue lexical tonal targets as they approach the bottom limit of their f0 ranges. This and related hypotheses require further investigation.

5. REFERENCES

- [1] Boersma, P., Weenink, D. 2018. Praat: A system for doing phonetics by computer, version 6.0. 40.
- [2] Gao, J., Hallé, P. A. 2015. The role of voice quality in Shanghai tone perception. *Proceedings of the*

- 18th International Congress of Phonetic Sciences.*
- [3] Gjersøe, S., Awasom, J. N., Paschen, L. 2016. The interaction of lexical tone and phrase-level intonation in limbum. *Speech Prosody 2016* 109–113.
 - [4] Guthrie, M. 1940. Tone ranges in a two-tone language (Lingala). *Bulletin of the School of Oriental and African Studies* 10(2), 469–478.
 - [5] Guthrie, M. 1951. *Grammaire et dictionnaire de lingala : la langue actuellement la plus parlée sur les deux rives de la partie centrale du fleuve Congo; avec un manuel de conversation français-lingala*. Leopoldville : Librairie Evangelique au Congo; Longmans.
 - [6] Hyman, L. M., Monaka, K. C. 2011. Tonal and non-tonal intonation in shekgalagari. In: Frota, S., Elordieta, G., Prieto, P., (eds), *Prosodic Categories: Production, Perception and Comprehension*. Dordrecht: Springer Netherlands 267–289.
 - [7] Kuang, J., Liberman, M. 2015. The effect of spectral slope on pitch perception. *Proceedings of Sixteenth Annual Conference of the International Speech Communication Association*.
 - [8] Kuznetsova, A., Brockhoff, P., Christensen, R. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, Articles* 82(13), 1–26.
 - [9] Meeuwis, M. 2010. *A grammatical overview of Lingala*. LINCOS Studies in African Linguistics 81. Lincom.
 - [10] Rialland, A., Aborobongui, M. E. 2017. How intonations interact with tones in Embosi (Bantu c25), a two-tone language without downdrift. *Intonation in African Tone Languages* 24, 195.
 - [11] Schadeberg, T. C. 1977. Anticipation of tone: Evidence from Bantu. *Language and linguistic problems in Africa* 195–204.
 - [12] Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Use R! Springer International Publishing.
 - [13] Woods, D. R., Akouala, F. 2002. *Lingala parallel texts*. Springfield, VA: Dunwoody Press.

¹ Guthrie [4] proposes that Lingala employs High vs. Mid (“Normal”), rather than High vs. Low, as he appears to reserve the low target for the extra low f₀ value which only occurs phrase-finally.