Tonal Aspects of Languages 2016
24-27 May 2016, Buffalo, New York

Perception of Tonal Contours in Medʉmba

Kathryn Franich

Department of Linguistics, University of Chicago, USA
kfranich@uchicago.edu

Abstract

This paper explores cues to contour tone perception in Medʉmba, a Grassfields Bantu language of Cameroon. Like many Bantu languages, Medʉmba contrasts only level high (H) and low (L) tones, but contour tones can result where H and L tones associate to a single syllable. Here, we examine the relative influence of vowel duration and tonal slope in Medʉmba speakers' perception of contour tones, as well as the influence of a following downstep on contour tone perception. Results of a tone identification task indicate that duration, rather than tonal slope, was the primary cue to contour perception. At longer vowel durations, tonal slope also played a role in contour perception, but only when the the context following the contour tone was not downstepped. We explain this finding in terms of an f_0-related auditory masking effect and suggest that such effects may be partially responsible for the relatively more common occurrence of downstep in languages with level tone inventories over those with mixed contour and level tone inventories.
Index Terms: Tone, acoustics, perception, Bantu

1. Introduction

Though there has been considerable growth in the study of tone perception over the last two decades, African languages have still received very little attention in this domain. Given how different the tonal inventories of these languages look compared with better studied languages such as Mandarin and Thai, adequate generalizations about tone perception in human language cannot be made without better coverage of African languages. The present study aims to fill this gap by examining the relative influence of tonal slope, duration, and downstep on the perception of contour tones in Medʉmba, a Grassfields Bantu language.

One of the fundamental differences observed between the tone systems of African languages and those of many East Asian and indigenous Mexican languages is that, while the latter language groups tend to contrast both level and contour tones, African languages overwhelmingly contrast only level tones. As a result, existing studies on African tone have focused exclusively on the perception of tones which are level. Unsurprisingly, much of this work has focused on the role of f_0 in cuing tonal contrasts. Findings from a study of Mambila, a Benue-Congo language with four lexical tones, showed much clearer categorical boundaries for H and L tones than for the two M(id) tones in the language [1]. For Edo, a Volta-Niger language, f_0 level and direction of f_0 change were both found to be important cues for tone perception of disyllabic nouns [2]. Tonal slope has received less attention as a cue in level tone perception, though work on Yoruba, another Volta-Niger language, has shown that a falling pitch contour is the most salient cue for native listeners in perceiving L versus M tones in word-final position [3]. Though M tones were found in this study to have both longer duration and higher amplitude than L tones in production, increasing either one of these properties alone did not lead listeners to perceive a higher number of M tones in the experiment. A shortcoming of the experiment was that two-way interactions between slope, duration, and amplitude were not examined, so the precise relationship between these cues in tone perception was left somewhat unclear. Recent work examining the relative weighting of pitch height, duration and tonal context on the perception of H and L level tones in Medʉmba found that duration, in addition to f_0 height, had a strong effect on tone perception, with more L tones perceived at longer durations [4]. Subjects in that study were also more likely to hear a H tone if the tone of the preceding trial had had a relatively lower f_0 value, and a L tone if it had had a relatively higher f_0, demonstrating a contrastive effect of pitch context.

1.1. Contour Tones in Medʉmba: Slope and Duration

Despite the prevalence of phonemic level tones, many Bantu languages make frequent use of contour tones, which result when a H and a L tone associate to a single syllable (Fig. 1).

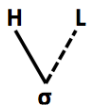

Figure 1: Autosegmental diagram of African-type contour tones

Medʉmba, a Grassfields Bantu language spoken in southwestern Cameroon, contrasts only H and L tones, but allows contour tones in a variety of contexts. Some of these contours occur as the result of docking of segmentless floating tone morphemes to neighboring syllables. For example, in the so-called 'associative construction', two nouns are joined together by a floating associative morpheme which docks to the head noun to form a contour (1). The resulting form in (1) has a HL falling tone on the first of the two nouns.

(1) mɛ́n ` mvə̀n → mɛ̂n mvə̀n
child AM chief child.of chief
'The chief's child'

HL falling contour tones are also common in Medʉmba loanwords borrowed from English. For example, the loanword *tîblì* 'table' carries a HL contour on the first syllable, and a L tone on the second syllable.

Though Medʉmba has no vowel length contrast, contour tones are produced with longer duration than level tones, in line with findings from many other languages [5, 6]. One question that arises, then, is whether Medʉmba listeners are sensitive to durational differences in contour tone perception. Interestingly, though durational differences between contour and level tones

http://dx.doi.org/10.21437/TAL.2016-17

have been noted for many languages, previous work examining the effect of duration on contour tone perception in languages such as Mandarin and Taiwanese has found only a small effect of duration on tone perception [7, 8]. In fact, [8] actually found that non-native listeners utilized duration more than native listeners in differentiating between Mandarin L and rising tones. Some work examining perception of whispered speech in Mandarin has shown that speakers may exploit secondary cues such as duration more when f_0 information is absent, however, suggesting that the relative informativity of duration in a given context (or a given language) may contribute to its use as a cue in perception [9]. One of the central questions we seek to explore in this study is how f_0 slope and duration may interact in cuing tonal contrasts in Medʉmba.

1.2. Downstep

Another intriguing property of contour tones in Medʉmba concerns their behavior with respect to downstep[1]. As in many other Bantu languages, HLH sequences (or HLLH sequences) in Medʉmba typically result in lowering of the second H (HL$^{\downarrow}$H), and lowering of the register of all subsequent H and L tones within a clause. However, this process is more restricted when the H tone undergoing downstep is part of a HL falling contour; in this case, only another falling contour can trigger the process, and even then, the process is blocked for some words. All loanwords bearing falling contours (2a) as well as many native words comprised of associative NPs and bearing falling contours (2b) can trigger downstep. There is a class of native words/NPs for which the process is blocked, however (3a). [2].

(2) a. All loanwords and many native words bearing HL contours trigger downstep
tî̱bli ˬnâʔ ʒúmɔ́
table DPAST dry
'The table was dry.'

b. mô̱nzə̱ ˬnâʔ ʒúmɔ́
large.AM.path DPAST dry
'The road was dry.'

(3) a. Downstep blocked after certain native words bearing HL contour
mbâ̱kù (*ˬ)nâʔ ʒúmɔ́
pot.AM.leg DPAST dry
'The ankle was dry.'

Thus, application of downstep in Medʉmba can be said to rely on a combination of phonological and lexical factors. An interesting question which has yet to be investigated is the extent to which processes such as downstep may influence tone perception. Using noncewords embedded as subjects in sentential contexts such as those in (2-3), we examine how the presence or absence of downstep on the following context influences perception of a syllable's tone. Depending on subjects' listening strategies, there are a number of ways in which downstep may influence perception. Using a'top-down' approach and treating nonce forms like loanwords (since novel words also tend to be incorporated according to loanword phonology), subjects might apply their lexical knowledge which dictates that downstep can only occur as a result of a tonal contour on the subject, thus perceiving a tonal contour anytime downstep is present. If, however, subjects incorporate a more 'bottom-up' strategy, downstep might be found to have a contrastive effect, such that listeners hear more high tones preceding a downstepped syllable than a non-downstepped syllable, as has been found in studies of speaker normalization for lexical tone [11, 12, 13]. Another possibility is that downstep will have a more complex interaction with slope and duration cues.

[1] *Downstep* here will refer primarily to 'automatic downstep' (lowering of the second H in a HLH sequence where the L is associated to a segment), following [10].

[2] We note that a change is in progress whereby fewer native words trigger downstep in this environment, possibly leading to a system in which downstep after HL contours is simply conditioned along loan/native word lines.

2. Method

Seventeen native Medʉmba speakers (8 female) were paid the equivalent of $10 US to participate in a tone identification task. Stimuli consisted of six disyllabic nonce forms embedded in the context X *nâʔ ʒúmɔ́* 'X was dry' (Table 1). The six forms were selected as the most natural-sounding by three native speakers who chose from a list of phonotactically licit nonce words.

Table 1: Nonce forms used in contour perception experiment

kuzim	məʉ̱	tsito
ləkam	bô̱sam	batʃim

Sentences were originally recorded by a female native speaker who was asked to produce the nonce forms with a falling tone on the first syllable and a low tone on the second syllable. For the first vowel, based on the range and average duration produced naturally by the speaker for level and contour tones in a separate production study, a three-point duration continuum (100 ms, 200 ms, 300 ms) was synthesized by trimming or lengthening the original syllable using PSOLA in Praat [14]. Resulting vowels were then subjected to f_0 resynthesis for the creation of a five-point slope continuum; the endpoints of the continuum were again based on the range produced naturally by the speaker for level and contour tones. The f_0 of the first half of each vowel was kept at 300 Hz, while the f_0 target at vowel offset ranged from 300 Hz (resulting in a completely flat f_0) to 220 Hz; f_0 was interpolated linearly between the midpoint and offset of the vowel (Table 2).

The second syllable was given a 220 Hz target starting from the vowel onset, which corresponded to the average f_0 for this syllable in the model speaker's productions. Pitch was then interpolated linearly between the offset of the initial vowel and the onset of the second vowel. In addition to manipulations on the nonce forms, tone of the sentential context was altered to simulate downstepped and nondownstepped contexts. Without being asked, the model speaker intuitively downstepped the high of the contour tone on the second word *nâʔ*, hence lowering the tonal register of the remainder of the sentence. This condi-

Table 2: Onset and offset f0 values for 5 tonal slopes

Slope Level	Onset f_0	Offset f_0
Slope 0	300 Hz	300 Hz
Slope 1	300 Hz	280 Hz
Slope 2	300 Hz	260 Hz
Slope 3	300 Hz	240 Hz
Slope 4	300 Hz	220 Hz

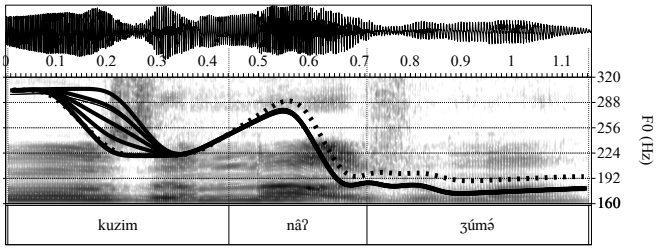

Figure 2: Pitch tracks of 5 f0 contours, with and without downstep on subsequent sentential context (non-downstepped version = dotted line)

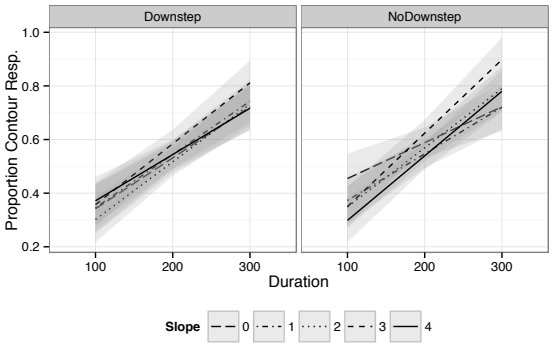

Figure 3: 3-way interaction: Duration × Slope × Downstep

tion was referred to as the 'downstep' condition. To simulate the 'non-downstep' condition (again, based on results from the speaker's natural productions in similar tonal contexts), tonal targets starting with the midpoint of the second word through the end of the sentence were raised by 15 Hz from the original production. (Figure 4). All sentences were normalized for intensity. In total, there were 180 trials (6 forms × 3 durations × 5 slopes × 2 downstep conditions).

All data was collected by the author in Bangangté, Cameroon in September of 2015. The experiment was run in a quiet room using Praat version 5.3.84 on a 13" Macbook Pro. Prior to beginning the experiment, participants were told they would be hearing sentences in which the first word sounded like Medumba, but had no meaning. They were asked to listen to the first syllable of the word and identify whether they heard a high tone or a falling tone. For those participants who were not familiar with the names of Medumba tones, several example sets of real Medumba words were given which contained high and falling tones. After this, participants were briefly quizzed to see if they could identify the tones of some additional Medumba words which had not been previously discussed. One subject had very low accuracy on the quiz; her data was thus excluded from the study. All participants were given several practice trials to get used to the experiment before beginning. Participants listened to stimuli through headphones and used the computer's trackpad to click on one of two buttons on the screen, corresponding to the tone they heard; they could not listen to the stimulus more than once. Stimuli were presented in random order by subject. The experiment was self-paced and took around 20 minutes to complete. A short break occurred every 30 trials.

3. Results

Data were analyzed using mixed-effects logistic regression (lme4 package [15] in *R* [16]). Fixed effects of DURATION (duration of the target syllable), SLOPE (f_0 slope of the target syllable), and DOWNSTEP (two levels, Downstep or No Downstep) were included. DURATION and SLOPE were treated as continuous variables and were centered to avoid collinearity effects. Random intercepts for SUBJECT and WORD (segmental content of nonce forms, ignoring tone) were included, each paired with random slopes for DURATION, SLOPE, and DOWNSTEP. Model selection proceeded with likelihood ratio tests, starting with the maximal model, and using stepwise backward elimination of each term until the optimal model was reached.

There was a strong main effect of DURATION (β=.175, t=4.68, p $<$0.001), indicating subjects perceived more contour tones as duration increased. In addition, there was a significant three-way interaction between DURATION, SLOPE and DOWNSTEP (β=.032 t=2.09, p$<$.05). As seen in Figure 5, in the No Downstep condition, items with different slopes patterned more differently from one another than they did in the Downstep condition. In particular, for larger slopes (Slopes 2-4) subjects perceived a greater number of contour tones at longer durations in the No Downstep condition. Unexpectedly, there was also an increase in contour tone responses at shorter durations where the slope of the target tone was the flattest, Slope 0, in the No Downstep condition compared with the Downstep condition. Subjects also reported hearing fewer contour tones for the steepest slope, Slope 4, at shorter durations in the No Downstep condition than in the Downstep condition.

4. Discussion

From the results outlined above, it appears subjects were not biased to interpret the noncewords in terms of Medumba loanword phonology, and thus did not apply a general listening strategy for all items involving downstep. Rather, it appears listeners paid close attention to the acoustic signal in making their judgments. Interestingly, our results indicate that vowel duration was a much more salient cue for Medumba listeners in identifying contour tones than was f_0 slope. This is surprising in light of previous work on East Asian and Indigenous Mexican languages which consistently indicates the primacy of f_0 slope and height over duration as cues in tone perception. This finding is in-line, however, with previous work on Medumba showing that duration also plays a role in level tone perception [4]. The emerging picture is that Medumba speakers, and perhaps speakers of other African languages, rely more on duration as a cue in tone perception than do speakers of languages with phonemic contour tones. Why should this be? Perhaps it is due to the extensive use of tone sandhi processes in Medumba such as downstep, which lead both f_0 height and slope to be highly variable according to tonal/prosodic context. In the face of such variation, duration may actually be a more reliable cue than f_0 for tone perception overall.

Indeed, the three-way interaction between f_0 slope, duration, and downstep would seem to support this notion. Our findings indicate that slope played no role in tone perception where the context following the target syllable was downstepped; however, in the No Downstep condition, subjects apparently utilized slope more in determining whether syllables of longer duration

were contoured or not. Why should speakers rely more on f_0 slope in the absence of downstep? One explanation for this finding comes from studies on frequency masking effects. Research has shown that 'backward masking' can take place such that a masking stimulus played after a test stimulus of a similar frequency can inhibit frequency discrimination of the test stimulus [17]. To apply this finding to the current results, since downstepping the environment following the target word has the effect of making the following environment more similar in overall pitch level to the the offset value of the target syllable (except when the target syllable carries a perfectly level tone), it may be that the lowered environment serves as a better perceptual mask for the pitch of the preceding contour tone.

This interplay between downstep and slope perception is especially interesting as we consider the distribution of downstep crosslinguistically. While some degree of gradual lowering within a clause or sentence is present in nearly all languages (tonal and otherwise), downstep of the kind driven by a phonological rule applying in HLH sequences appears to be far more common in African languages, where slope plays less of a role in defining tonal categories. Indeed, downstep as a tone sandhi process has been claimed to be absent from several contour tone languages, including Mandarin [18], Vietnamese [19] and Triqui [20]. One possibility is that downstep is absent from languages with more phonemic tonal contours precisely *because* the process makes slope-based tonal contrasts more difficult to perceive[3].

That contour tone responses increased at shorter durations where the slope of the target tone was the flattest, Slope 0, remains a puzzle, as both the duration and slope cues available should have encouraged more H tone responses. Also puzzling is the fact that subjects reported hearing fewer contour tones for the steepest slope, Slope 4, at shorter durations in the No Downstep condition than in the Downstep condition. One possibility is that subjects *were* accessing slope information for Slope 4 more easily in the No Downstep condition, but that the slope for these items was so extreme that subjects heard it as a sequence of H-L, rather than a HL contour preceding another L. Since shorter durations are overall more highly associated with H tones, the duration cue provided subjects with further reason to believe they had heard a H tone and not a falling tone.

5. Conclusions

We have shown that duration is the primary cue to contour tone perception in Medumba, with duration far outweighing tonal slope as a cue. Despite this fact, we did find that downstep interacts in complex ways with duration and f_0 slope in tone perception. Specifically, where vowel duration is longer and downstep is not present, subjects appear able to rely somewhat on slope in tone identification; when downstep is present, however, this reliance goes away. We attribute these results to a frequency masking effect that downstep exerts on the preceding tonal context, and argue that this effect may explain why downstep is apparently rarer in tone languages such as Mandarin, Vietnamese, and Triqui which contrast both level and contour tones. If this explanation is on the right track, at the very least, we would expect that the degree of lowering taking place in languages with phonemic contours would be less pronounced than in languages with only level tones, such as Medumba; an important next step will be to compare more directly such effects across languages of different families and tonal inventories.

[3]Of course, in making such a claim, we must also consider the relative density of tonal inventories: since languages with phonemic contours also tend to have more tones, we must also consider the role that inventory size may play in minimizing contextual variability due to processes like downstep.

6. Acknowledgements

This work was supported by National Science Foundation Linguistics Program Grant No. BCS-1423865 (co-PIs: Kathryn Franich and Alan C.L. Yu). The National Science Foundation does not necessarily endorse the ideas and claims in this paper. Thanks to two anonymous reviewers, audiences at TAL 2016, members of the University of Chicago Phonology Reading Group, and Alan Yu for valuable discussion.

7. References

[1] Connell, B. "The perception of lexical tone in Mambila", Language and Speech, 43, 163–182, 2000.

[2] Omozuwa, V. E. "Acoustic cues for the perception of tones of disyllabic nouns in Edo", Studies in African Linguistics, 22(2), 135–156, 1991.

[3] Hombert, J-M. "Perception of tones in bisyllabic nouns in Yoruba", Studies in African Linguistics, S6, 109–122, 1976.

[4] Franich, K. "Internal and contextual cues to tone perception in Medumba", JASA Express Letters, Forthcoming.

[5] Gordon, M. "Syllable weight: Phonetics, phonology, and typology", UCLA dissertation, 1999.

[6] Zhang, J. "The effects of duration and sonority on contour tone distribution–typological survey and formal analysis", UCLA dissertation, 2001.

[7] Lin, H. B. and Repp, B. H. "Cues to the perception of Taiwanese tones", Language and Speech, 32(1), 25–44, 1989.

[8] Chang, Y-H. S. "Distinction between Mandarin tones 2 and 3 for L1 and L2 listeners", Proceedings of NACCL-23, University of Oregon, 84–96, 2001.

[9] Liu, S. and Samuel, A. G. "Perception of Mandarin lexical tone when F0 information is neutralized", Language and Speech, 47, Pt. 2, 109–138, 2004.

[10] Stewart, J. M. "The typology of the Twi tone system", Bulletin of the Institute of African Studies, Legon, 1: 1D27, 1965.

[11] Holt, L. L. "Speech categorization in context: Joint effects of nonspeech and speech precursors", JASA, 119, 4016–4026, 2006.

[12] Huang, J, and Holt, L. L. "General perceptual contributions to lexical tone normalization", JASA, 125, 3983–3994, 2009.

[13] Francis, A., Ciocca, V., Wong, N., Leung, W., and Chu, P. "Extrinsic context affects perceptual normalization of lexical tone", JASA, 119, 1712–1726, 2006.

[14] Boersma, P. and Weenink, D. Praat: doing phonetics by computer [Computer program]. Version 5.3.84, retrieved 8 November 2015 from http://www.praat.org/

[15] Bates, D., Mächler, M., Bolker, B. and Walker, S. "Fitting Linear Mixed-Effects Models Using lme4", Journal of Statistical Software, 67, 1, 1-48, 2015.

[16] R Core Team. "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/, 2013.

[17] Massaro, D. W. "Backwards recognition masking", JASA, 58, 1059-1066, 1975.

[18] Chen, M. "Tone rule typology", Proceedings of BLS 18: Special Session on the Typology of Tone Languages, 54-66, 1992.

[19] Ha, K-P. "Prosody in Vietnamese: Intonational form and function of short utterances in conversation", Asia-Pacific Linguistics 002 (SEAMLESS 001), PhD thesis, Canberra: The Australian National University, 2012.

[20] DiCanio, C. "Triqui tonal coarticulation and contrast preservation in tonal phonology", Proceedings of Sound Systems of Mexico and Central America, Yale University, 2014.