Kathryn H. Franich: JASA Express Letters [http://dx.doi.org/10.1121/1.4955003] Published Online 22 July 2016

Internal and contextual cues to tone perception in Medʉmba

Kathryn H. Franich
Department of Linguistics, University of Chicago, 1115 East 58th Street, Room 224, Chicago, Illinois 60637, USA
kfranich@uchicago.edu

Abstract: This study presents results of an identification experiment with speakers of Medʉmba, a Grassfields Bantu language, aimed at evaluating the relative effects of f_0 and duration in cuing tonal contrasts, as well as the role of lexical vs non-speech pitch contexts in biasing tone perception. Results show that duration is a cue for tone perception, with the influence of duration strongest where target f_0 values were lower. Lexical tone perception is also sensitive to the identity of a preceding lexical tone, but not to the presence of a preceding non-speech pure tone.

Date Received: October 30, 2015 **Date Accepted:** May 18, 2016

1. Introduction

This work investigates the role of target-internal and contextual factors in tone perception in Medʉmba, a Grassfields Bantu language spoken in Cameroon. Despite the large body of work examining perceptual properties of tone languages (Abramson, 1979; Gandour, 1981; Xu, 1994; Brunelle, 2009; Kirby, 2010), little work has examined such properties in African languages. Medʉmba, like many African tone languages, contrasts only H and L "level" tones, where f_0 remains relatively consistent across the tone-bearing unit; this is in contrast to better-studied East Asian languages, in which contour tones—tones with a dynamic falling or rising f_0 profile—make up at least part (if not most) of a language's tonal inventory. Among the few studies on tone perception in African languages, Hombert (1976) and Omozuwa (1991) both found f_0 level and direction of f_0 change to be primary cues for tone perception of disyllabic nouns in Yoruba and Edo, respectively, two Volta-Niger languages. Connell (2000) found that speakers of Mambila, a Benue-Congo language with four lexical tones, showed much clearer categorical boundaries for high (H) and low (L) tones than for the two mid (M) tones in the language. Mixdorff *et al.* (2011) examined perception of tones in Sesotho, a Southern Bantu language, finding that f_0 can outweigh vowel quality as a cue in lexical identification.

Besides f_0 and vowel quality, duration has been shown to be an important cue associated with both the perception and production of lexical tone cross-linguistically. Specifically, while L tones have been found to be produced with longer duration (Gandour, 1977), listeners tend to hear L tones as shorter than H tones when duration is held constant (Yu, 2010; Gussenhoven and Zhou, 2013). This effect is thought to arise from a compensatory mechanism on the part of speakers to normalize for f_0-related perceptual or articulatory biases. While the phenomenon has been documented for English and Chinese languages, it has yet to be explored in an African tone language.

Tone perception has also been found to be influenced by the pitch of words in the surrounding context. Surrounding pitch context has been found to have a strong influence in studies on talker normalization in tone perception (Leather, 1983; Wong and Diehl, 2003; Francis *et al.*, 2003; Francis *et al.*, 2006). While contextual effects clearly played a role in these studies, it is unclear whether such effects were specific to linguistic stimuli, or whether they represent a more general auditory effect of f_0. Huang and Holt (2009, 2011) found that contextual non-speech pure tones which mimic the f_0 of a real talker can elicit perceptual biases in tone categorization, suggesting tone perception may rely on domain-general auditory processing mechanisms.

Manipulations for talker normalization studies either involved different talkers saying a carrier sentence, or synthetically altered stimuli meant to create the impression of different talkers. An additional question thus concerns how much intra-speaker effects of pitch context can influence tone perception. Two early studies aimed to examine such effects by looking at Mandarin tone identification in paired syllables.

Lin and Wang (1985) found that raising the onset f_0 of the second syllable (which bore a falling tone) made subjects more likely to hear a rising tone on the first syllable, even though the f_0 of the first syllable was unchanged throughout the experiment. However, a later study by Fox and Qi (1990) using a similar methodology found little effect of context. A potential issue with both of these studies was that they directly compared perception of level and contour tones in context, though more recent work has found contextual effects to be stronger for level tones than for contours (Francis *et al.,* 2003).

The current study examines tone perception in Medʉmba, focusing on two questions: (1) what is the role of duration in tone perception in Medʉmba and (2) how does f_0 of a preceding sound influence tone perception of a target word? If speakers are indeed sensitive to pitch cues from the surrounding context, a related question concerns the types of stimuli which can exert such an influence. Specifically, we investigate whether speakers are sensitive only to the pitch of surrounding speech sounds, or whether pitch of contextual non-speech sounds can also bias tone perception.

2. Methods

2.1 Participants

Nineteen Medʉmba speakers (9 female) aged 18–47 participated in a word identification task. Speakers were paid the equivalent of $10 U.S. and none reported any speech or hearing problems.

2.2 Stimuli

Stimuli consisted of 6 syllable types and two contrastive tones (Table 1). A female native speaker (who did not participate in the experiment) produced the syllables, which were then resynthesized in PRAAT using PSOLA. A seven-step f_0 continuum was created for each syllable which ranged from 185 to 275 Hz, increasing at each step by 15 Hz; f_0 remained constant throughout the syllable. The highest and lowest values on the continuum were based on the average f_0 values of the model speaker's H and L tones, as assessed from production data collected in a separate experiment. A three-step duration continuum was then created from the resynthesized syllables with values of 100, 175, and 250 ms. The original syllables from the model talker were produced with durations of around 175 ms, on average. To achieve the three target durations, stimuli were either trimmed or lengthened by splicing in copies of vowel wavelets taken from within the target syllable. In addition, pure tones consisting of a single sine wave with a frequency corresponding to either the highest or lowest points on the f_0 continuum were generated as "context pure tones" to be played 40 ms before the target syllables. Manipulations resulted in the creation of 42 total stimuli per syllable (7 f_0 values $\times$ 3 durations $\times$ 2 context pure tones), for a total of 252 trials (42 stimuli $\times$ 6 blocks) per subject [Franich (2016)]. All stimuli were normalized for intensity.

2.3 Procedure

All data were 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 10 in. Macbook Pro. 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 word they heard; they could not listen to a stimulus more than once. Stimuli were presented in blocks according to syllable, with stimulus order randomized within each block and block order randomized by subject. Subjects were given a brief tutorial on how to use the trackpad (as many had not had previous experience with a computer) and were given one full block of training (responses were not recorded) before the actual experiment began. The study was self-paced, and took around 25 min to complete on average, including short breaks between blocks.

Table 1. Translations of Medʉmba words.

	Syllable	H Tone	L Tone
1)	fə	"sheet of paper"	"cadaver"
2)	tsə	"season of famine"	"in-laws"
3)	la	"pomade"	"pineapple"
4)	sa	"dance"/"game"	"star"
5)	to	"hole"	"belly button"
6)	ko	"summit"	"lance"

3. Results

H tone responses were modeled using mixed effects logistic regression with the LME4 package for R. The model included five predictors, Target_F_0 (f_0 of the target syllable), Duration (duration of target syllable), Prec_F_0 (f_0 of target syllable in the preceding trial), PrecResponse (response to target syllable in the preceding trial, H or L), and PureTone (level of context pure tone, H or L). The first three continuous variables were mean-centered to limit collinearity effects and the last two categorical variables were sum-coded. Random intercepts for Subject and Syllable (segmental content of the syllable, ignoring tone) were included, and random slopes were included for all aforementioned variables. Model selection proceeded with likelihood ratio tests.

As expected, there was a main effect of Target_F_0 ($\beta = 1.3801$, $p < 0.0001$) with H responses correlated positively with f_0 of the target syllable. There was also an effect of Duration ($\beta = -0.2936$, $p < 0.001$), with fewer H responses at longer durations. Furthermore, there was an interaction between Target_F_0 and Duration ($\beta = 0.1569$, $p < 0.0001$). Figure 1 shows that target syllable duration had a stronger effect on perception where target f_0 was lower, and less of an effect where target f_0 was higher. This is perhaps attributable to a perceptual compensation effect whereby subjects "add" additional duration to syllables with higher pitch to normalize for f_0-related articulatory effects on duration in production (Gussenhoven and Zhou, 2013).

For context effects, there was no effect of PureTone ($p > 0.8$), indicating that perception was not influenced by the frequency of non-speech sounds. However, there was a strong effect of both Prec_F_0 ($\beta = -0.4115$, $p < 0.0001$) and PrecResponse ($\beta = 0.3204$ p < 0.0001), indicating that both actual f_0 and perceived tone of the syllable on the previous trial predicted perception of the target. A positive correlation between H tone responses and preceding response indicates that participants heard more H tones when they had judged the previous tone as H, and more L tones when they had judged it as L. As can be seen in Fig. 2, there was an inverse relationship between target H responses and f_0 of the preceding trial, showing that lower f_0 values of the preceding syllable encouraged more H responses, and vice versa.

There was also a significant three-way interaction between Target_F_0, Prec_F_0, and PrecResponse ($\beta = 0.5765$, $p < 0.0001$). Figure 3 shows this interaction, with the x axis representing f_0 of the target syllable, solid and dotted lines representing subjects' responses from the preceding trial (H vs L, respectively), and each pane representing the f_0 of the syllable on the previous trial. Proportion of H responses was predictable based on f_0 of the target, but only in those cases where participant judgments on the preceding trial syllable were the most *congruous*, or consistent with the actual f_0 of that syllable. In other words, if the syllable from the previous trial was on the higher end of the f_0 spectrum and the subject indicated hearing a H tone on that trial, this would be considered a "congruous" outcome; if they indicated a L tone on the same trial, this would be an "incongruous" outcome. Where outcomes of the preceding trial were incongruous, target f_0 was not a good predictor of target response, and subjects instead showed a tendency to match the target response with the preceding trial response. Examination of individual subjects' data revealed this pattern was consistent across speakers. This suggests that subjects use both acoustic information from the previous syllable and their own categorization patterns to generate expectations about

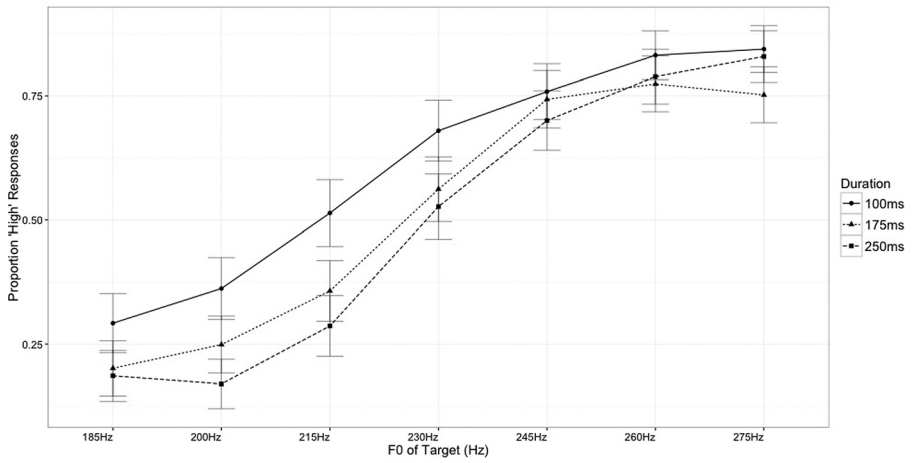

Fig. 1. Proportion of H tone responses on target by F0 and duration.

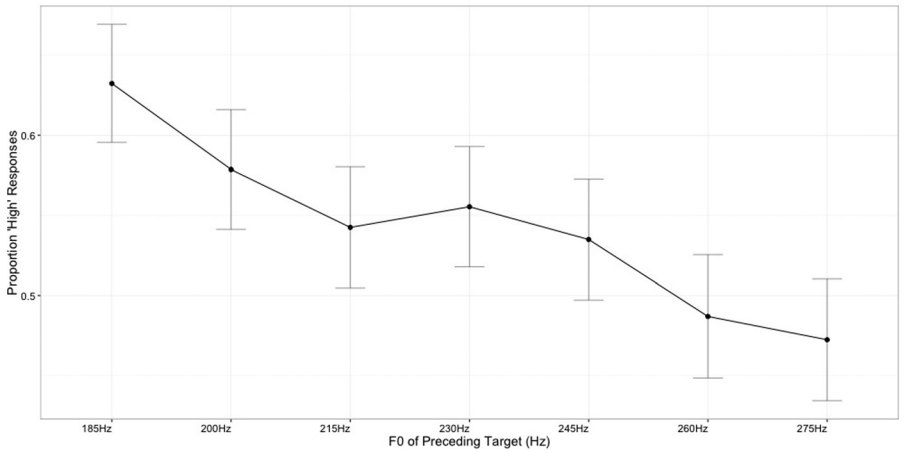

Fig. 2. Proportion of H tone responses on target by F0 of preceding trial.

upcoming tones; when expectations are not consistent with incoming acoustic information, subjects are able to rely less on the acoustic signal and resort to other strategies to identify target tones.

Where the preceding syllable was in the middle of the f_0 continuum (230 Hz), f_0 of the target syllable was a good predictor of participant response regardless of whether subjects rated the preceding syllable as H or L. This is likely the result of perceptual assimilation, such that subjects could interpret these preceding 230 Hz syllables as either H or L reference points in perception of the target tone. On trials where the preceding syllable was 230 Hz, judgments of target tones occurring in the middle of the f_0 continuum were biased by judgments on the preceding syllable. Specifically, more H tone responses were recorded where the preceding response was H, and vice versa. However, responses converged at either end of the target continuum (185 and 275 Hz), presumably because target f_0 values at these points were extreme enough that subjects relied less on the preceding context.

4. Discussion

This study examined two types of cues to tone perception in Medumba, including the relative weighting of target-internal phonetic properties of duration and f_0, and the contextual influence of preceding speech and non-speech tones. Findings indicate that f_0 and duration are both cues to tone perception in the language, with duration being an especially salient cue for tonal differences where the target syllable carries a lower f_0 value. In particular, lower f_0 syllables with short durations were perceived as H tones more often than their longer counterparts. This result could be due to a compensatory effect as proposed by Yu (2010) or Gussenhoven and Zhou (2013), though the nature of such an effect remains unclear in the absence of corroborating production data.

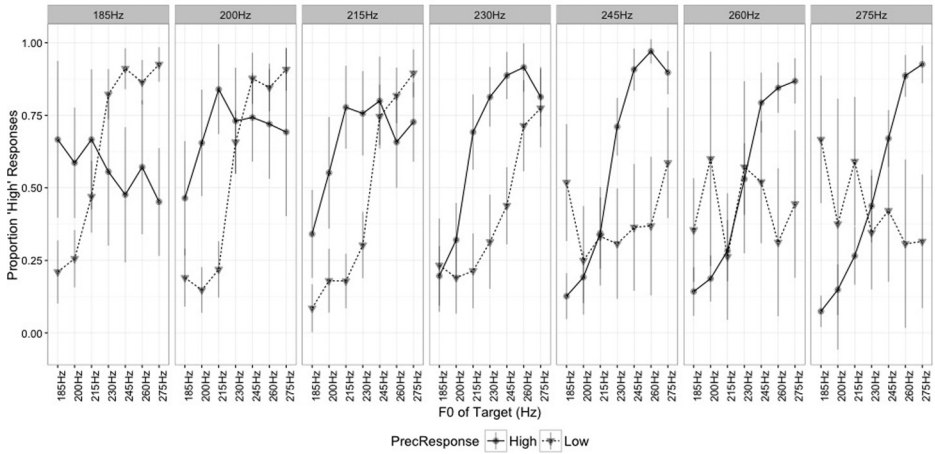

Fig. 3. Proportion of H tone responses on target by F0, preceding response, and F0 of preceding trial (individual windows represent F0 of preceding target syllable within block).

The f_0 of the target syllable within a preceding experimental block had a strong effect on perception of the target tone, suggesting that contextual pitch information does have an influence on tone perception. In general, preceding f_0 had a "contrastive" effect, with lower f_0 values leading to more H tone responses on the target syllable, and vice versa. Subjects' tone identification on the target was also positively correlated with their preceding response. The importance of contextual information was made especially clear from the three-way interaction between target syllable f_0, preceding syllable f_0, and perceived tone of the preceding syllable. Specifically, conflicts between category assessment and actual f_0 value of the preceding syllable influenced listeners' tone identification on the target, making it less predictable from target f_0 values.

Interestingly, H and L non-speech tones had no effect on subjects' perception of tone. This is somewhat unexpected given findings by Huang and Holt (2009, 2011) which indicate that contextual non-speech stimuli can elicit similar perceptual effects as speech stimuli. Non-speech stimuli used in the present study differed from those used in previous studies in that ours consisted of a single tone, whereas theirs consisted of several tones of different frequencies whose overall frequency/f_0 was raised or lowered to create the surrounding context. It is possible that the complexity of processing both speech and non-speech stimuli in rapid succession effectively ruled out any potential effects of the non-speech stimuli, as subjects were likely attending much more directly to the speech stimuli. Future work should investigate whether lengthening the duration of a single non speech context tone might lead to larger effects on target tone perception.

5. Conclusions

This study has shown that Medumba speakers are sensitive to both internal and contextual tonal cues in perceiving tonal categories. An important finding was that duration, in addition to f_0, influences tone perception; such effects have not been widely described for African tone languages. Also of interest was that the pitch of preceding speech sounds had a strong effect on tone perception of the target syllable, but pitch of non-speech sounds did not.

Acknowledgments

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 Helena Aparicio, Christian DiCanio, Julian Grove, Ming Xiang, Alan Yu, and audiences at ACAL 47 for valuable discussion.

References and links

Abramson, A. (**1979**). "Noncategorical perception of tone categories in Thai," J. Acoust. Soc. Am. **61**, S66.

Brunelle, M. (**2009**). "Northern and Southern Vietnamese tone coarticulation: A comparative case study," J. Southeast Asian Ling. Soc. **1**, 49–62.

Connell, B. (**2000**). "The perception of lexical tone in Mambila," Lang. Speech **43**, 163–182.

Fox, R., and Qi, Y. (**1990**). "Contextual effects in the perception of lexical tone," J. Chin. Ling. **18**, 261–283.

Francis, A., Ciocca, V., Wong, N., Leung, W., and Chu, P. (**2006**). "Extrinsic context affects perceptual normalization of lexical tone," J. Acoust. Soc. Am. **119**, 1712–1726.

Francis, A. L., Ciocca, V., and Ng, B. K. (**2003**). "On the (non)categorical perception of lexical tones," Percept. Psychophys. **65**, 1029–1044.

Franich, K. H. (**2016**). All stimuli can be accessed at http://home.uchicago.edu/kfranich/Perception_Study_Sound_Files.html (Last viewed July 12, 2016).

Gandour, J. (**1977**). "On the interaction between tone and vowel length: Evidence from Thai dialects," Phonetica **34**, 54–65.

Gandour, J. T. (**1981**). "Perceptual dimensions of tone: Evidence from Cantonese," J. Chin. Ling. **9**(1), 1–33.

Gussenhoven, C., and Zhou, W. (**2013**). "Revisiting pitch slope and height effects on perceived duration," in *Proceedings of Interspeech*, Lyon, France, pp. 1365–1369.

Hombert, J.-M. (**1976**). "Perception of tones in bisyllabic nouns in Yoruba," Stud. African Ling. S **6**, 109–122.

Huang, J., and Holt, L. L. (**2009**). "General perceptual contributions to lexical tone normalization," J. Acoust. Soc. Am. **125**, 3983–3994.

Huang, J., and Holt, L. L. (**2011**). "Evidence for the central origin of lexical tone normalization," J. Acoust. Soc. Am. **129**, 1145–1148.

Kirby, J. (**2010**). "Dialect experience in Vietnamese tone perception," J. Acoust. Soc. Am. **127**(6), 3749–3757.

Leather, J. (**1983**). "Speaker normalization in perception of lexical tone," J. Phon. **11**, 373–382.

Lin, T., and Wang, W. (**1985**). "Tone perception," J. Chin. Ling. **2**, 59–69.
Mixdorff, H., Mohasi, L., Machobane, M., and Niesler, T. (**2011**). "A study on tone and intonation perception in Sesotho," in *Proceedings of Interspeech*, Florence, Italy, pp. 3181–3184.
Omozuwa, V. E. (**1991**). "Acoustic cues for the perception of tones of disyllabic nouns in Edo," Stud. African Ling. **22**(2), 135–156.
Wong, P. C. M., and Diehl, R. L. (**2003**). "Perceptual normalization for inter- and intra-talker variation in Cantonese," J. Speech Lang. Hear. Res. **46**(2), 413–421.
Xu, Y. (**1994**). "Production and perception of coarticulated tones," J. Acoust. Soc. Am. **95**, 2240–2253.
Yu, A. C. L. (**2010**). "Tonal effects on perceived vowel duration," in *Laboratory Phonology 10* (Mouton de Gruyter, Berlin), pp. 51–168.