SOMETIMES LOW REALLY IS JUST THE OPPOSITE OF HIGH: PERCEPTION OF LOW F0 TARGETS BY TONE AND NON-TONE LANGUAGE SPEAKERS

Jonathan Barnes[1], Chris K. C. Lee[1], Alejna Brugos[2], Stefanie Shattuck-Hufnagel[3], & Nanette Veilleux[2]

[1] Boston University, [2] Simmons University, [3] Massachusetts Institute of Technology
jabarnes@bu.edu, chriskcl@bu.edu, alejna.brugos@simmons.edu, sshuf@mit.edu, veilleux@simmons.edu

ABSTRACT

Previous perceptual studies show that High pitch accents with maximum F0 extended in time sound higher than sharp-peaked counterparts with identical maximum F0 to listeners of various language backgrounds. It is not clear, however, whether this ‘plateau’ effect extends to the perception of Low pitch accents. We report results from two perception tasks involving American English (NE) and Cantonese-English bilingual (BiL) speakers, aimed at uncovering influences of contour shape on perception of Low and High pitch accents in English. We found that while the ‘plateau’ effect is demonstrated by both groups in the case of high pitch accents, only BiL participants judged a low pitch accent with a temporally extended minimum F0 to be lower than a sharp-turn contour with identical minimum F0. We discuss our findings from the perspective of the language-specific functional load placed on discrimination of Low pitch targets, and an averaging-based model of F0 scaling perception.

Keywords: Intonation, Pitch perception, tone scaling, tonal timing, Tonal Center of Gravity

1 INTRODUCTION

Intonation researchers have learned a great deal over the past several decades about the production and perception of High intonational pitch accents, both within and across languages. Considerably less is known about corresponding aspects of their Low counterparts. There is, in fact, a broad sense in the literature that Low and High tones may behave differently in implementation (e.g., [1], “High is not just the opposite of Low”). The nature of those differences, however, awaits further investigation.

Much attention has been paid recently, for example, to the effect of global F0 contour shape on tone perception. One of the best-known contour shape effects concerns differences between sharp-peaked and “plateau”-shaped High accents ([2]–[7], inter alia). Specifically, it has been demonstrated in a variety of languages that all else equal, a plateau-shaped pitch accent will sound higher in pitch to listeners than an analogous sharp-peaked accent with identical maximum F0 [4], [8], [9] (illustrated schematically in Fig. 1a). One possible explanation for this ([10]–[12]) rests on the assumption that perception of tone scaling involves not precise identification of a single target F0 level (maximum or minimum), but rather, a process of F0 averaging over some domain of interest à la [13], [14]. If this is correct, then we also might expect to find the same pattern for Low accents as for High, i.e. that an L* should sound lower to listeners when the minimum is extended to create a low flat region (a “basin”, if you will), than when it is reached only as a sharp fall-rise pattern (a “gully”, depicted schematically in Fig. 1b).

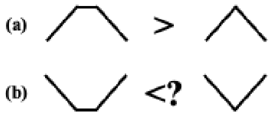

Figure 1: Schematics showing the effect of contour shape on F0 scaling of high (a) and low (b) pitch accents. “>” = *sounding higher than*; “<” = *sounding lower than*.

We know of only one study investigating potential perceptual differences between low “basins” and “gullies”: [15] report on results showing that while Low basin-shaped pitch accents did in fact sound lower than narrow gully analogues, (British English-native) listeners were significantly less adept at discriminating pitches in the lower F0 range, reducing the salience of the plateau effect for Low accents as compared with Highs. Similar results involving reiterant speech and non-speech stimuli, furthermore, lead them to speculate that Low perception may be worse than High for domain-general psychoacoustic reasons. At the same time, they wonder whether speakers of languages with more Low F0 targets might not prove better at its perception, perhaps yielding a different pattern of experimental results.

In this study, we have two goals. The first is to determine whether low basin pitch accents in fact sound lower than sharp gully analogues, as the averaging-based model of F0 scaling perception put forward in [9] (called TCoG-F) predicts. The second is to determine whether the presence or magnitude of the effect could be influenced at all by the language experience of the listeners in question. To do this, we conducted a perception experiment involving Low and High pitch accents of English, judged by two sets of participants: native speakers of American English, and Cantonese-English bilinguals from Hong Kong. Cantonese is a lexical tone language contrasting several tone categories with a Low F0 component: low level (T6) vs. low fall (T4) vs. low rise (T5). The tone inventory of Cantonese is illustrated in Fig. 2. It should be noted not only that Cantonese contrasts Low tones with different dynamic properties (fall, rise, level), but also that F0 level is one of the most salient acoustic cues, if not the primary cue [16] to the contrast between T6 and T4. (T4 reaches a lower minimum typically than T6.) The English spoken by this population is typically described as influenced in its phonology by its Cantonese substrate ([17], [18]).

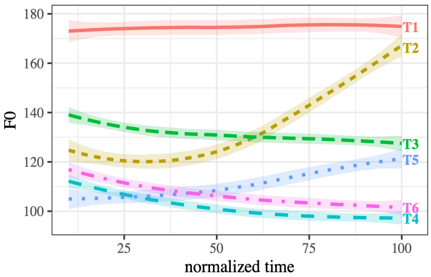

Figure 2: The six lexical tones in Cantonese produced by a male native speaker in citation form.

2 METHOD

Our research questions are thus (1) whether a basin-shaped L*-accented syllable would yield lower F0 scaling judgments relative to an analogous gully-shaped accent (just as high plateaux yield higher judgments relative to sharp peaks), and (2) whether listeners of different linguistic backgrounds might differ in these judgments, depending on the functional load their native languages place on the discrimination of low F0 levels. The study consists of two parts, one focused on replicating prior results involving high F0 plateaux, the other focused on low basins and their narrow gully analogues.

2.1 Stimulus creation

The base stimulus for both parts of the experiment was the sentence "*DAY* might fit", with nuclear accent on *DAY,* produced by a male native speaker of American English, then resynthesized using [19]. In both the **High** and the **Low** judgment tasks, each trial consists of a pair of contours, one we will call the "target", and another serving as the "reference". In the **High** judgment task, both contours take the form H* L-H% (the "uncertainty" contour of [20]). Target contours then bore one of four different shapes realizing the High nuclear pitch accent. In all four, the accented syllable began with a 250 ms F0 rise, from 141 Hz to 200 Hz. This was followed by either (1) an immediate F0 fall, creating a sharp peak coinciding with the end of the accented vowel, (2) a 25-ms plateau at the maximum F0, followed by a fall, (3) a 50-ms plateau, or (4) a 75-ms plateau. To keep the initial rise constant across all targets, rime duration for *DAY* increased in proportion with the duration of the plateaux.

In the **Low** judgment task, all stimuli bore the intonation contour L*+H L-H%, again with *DAY* bearing the nuclear accent. As with the High task, target stimuli took four different shapes. All of them began with a 250 ms fall from 141 Hz to 100 Hz, followed by (1) an immediate rise, creating a narrow gully, (2) a 25 ms low basin, followed by a rise, (3) a 50 ms low basin, or (4) a 75 ms low basin. Low accent contours were thus effectively the mirror image of High accent contours, in terms of shape and alignment of key turning points.

In both tasks, the reference contours that targets were paired with in each trial were identical to targets, except that F0 was flat during the accented syllable *DAY* at one of seven reference F0 levels. In the High task, the highest of these levels was at 200 Hz, identical to the F0 maximum in the High target contour. The remaining six levels formed a continuum declining in half semitone increments. (Reference contours remained perceptually plausible instantiations of H* L-H%.) In the Low task, the lowest reference level was flat throughout the accented syllable at 100 Hz, identical to the minimum in the Low target contour, with remaining reference levels forming a continuum increasing in half semitone increments. In the High task, for each target-reference pair, participants were asked to judge which of the two contours reached a higher pitch during the accented syllable. In the Low task, by contrast, they were asked to judge which of the two accented syllables sounded lower. Fig 3. illustrates these reference levels and target contours.

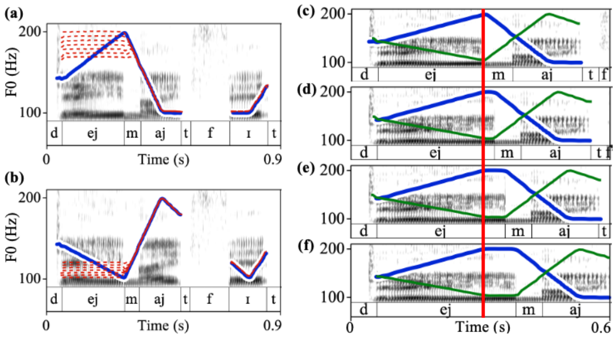

Figure 3: (left) F0 contours superimposed on spectrograms for reference (red) and sharp-turn (blue) target contours, for (a) high and (b) low judgment tasks. (right) The first 600 ms of target contours showing varying plateaux (blue) and basin (green) durations.

2.2 Procedure

The two tasks were presented to the participants in separate blocks (with order counterbalanced across participants). Participants listened to pairs of target contours and corresponding reference levels (order of presentation randomized) and decided which

DAY sounded higher (or lower, depending on the task). 18 additional **control trials** comparing reference levels separated by 2 or 3 continuum steps were included as a measure of participants' accuracy in discriminating pitch levels. There were 130 trials (4 shapes × 7 reference levels × 2 orders × 2 repetitions + 18 controls) in each task, presented in random order.

19 native American English (NE) speakers (13 female, 5 male, 1 non-binary, aged 18-21) and 22 Cantonese-English bilingual (BiL) speakers (13F, 9M, aged 22-33) reporting no speech or hearing deficits participated in the study. Fig. 4 shows accuracy rates in control trials for the two groups of participants. It should be clear that while BiL participants performed well in both tasks, the Low judgment task was challenging for NE participants. Since both groups generally had little difficulty discriminating control trials for the High task, we used participants' accuracy rates in those trials, with a threshold of 70%, as a criterion for inclusion in further analysis. Data from one speaker of each group were discarded.

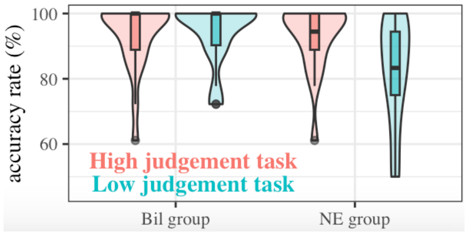

Figure 4: Participants' accuracy rates in control trials.

3 RESULTS

3.1 High judgment task

Fig. 5 displays results of the High judgment task, pooled across subjects separated by group. Lines represent the percentage of trials in which a given shape (*sharp turn*, *25-ms plateau*, *50-ms plateau*, *75-ms plateau*) was judged **higher** than each of its seven reference levels. Two trends are immediately obvious: (1) Across contour shapes, participants tend to judge target *DAY* as higher than the lowest reference levels, but lower than the highest levels. (2) As plateau length increases, participants' 'higher target' judgments decline later, suggesting that longer plateaux result in higher perceived pitch accent scaling (thus replicating previous findings). More interestingly, despite minor differences in the percentage of 'higher target' responses at each step, the two groups of participants display essentially the same response patterns. These observations are supported by mixed-effects logistic regression analysis, using reference level (continuous variable from 0 to 6, with 0 as the lowest level), target contour shape (continuous variable from 0 to 3, with 0 representing sharp turn), and participant group (factor variable with BiL as reference) as fixed effects, and participant and trial included as random intercepts. The best model, based on all-subset selection, contains a main effect of target contour shape (β=.433, z=10.72, p<.001), a main effect of reference level (β=-1.17, z=-12.42, p<.001) and an interaction between reference level and group (β=.193, z=-3.21, p=.001). No main effect of participant group is selected in the best model. This suggests that at least for high F0 target perception, listeners of different linguistic backgrounds are affected by contour shape in qualitatively the same way: high pitch accents with an extended plateaux sound higher than analogous sharp-peak accents with identical maximum F0.

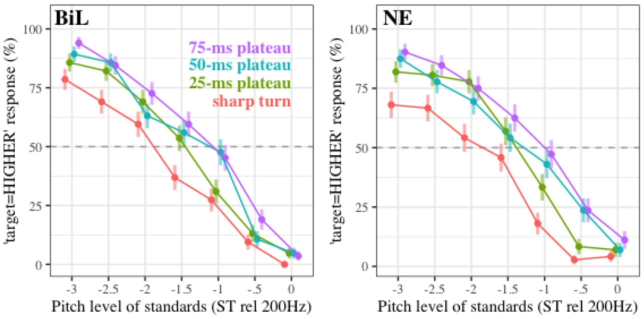

Figure 5: Percentage 'target is HIGHER' judgments for the four contour shapes, as a function of the reference level against which they were compared.

3.2 Low judgment task

Fig. 6 displays results from the Low judgment task, pooled across subjects separated by group. Lines represent the percentage of trials in which a given shape was judged **lower** than each reference level. Unlike in the high task, the two groups differ markedly in terms of how they perceived the scaling of target pitch accents: NE participants' 'lower target' response rate is almost always below 50%, meaning that they tend to perceive target *DAY* as higher in pitch than all but the highest reference levels. Additionally, NE participants show no clear pattern involving the shape or temporal extent of the Low pitch accents. In other words, low-basin contours sounded no lower to NE listeners than sharp-turn or gully-shaped low accents.

The response pattern of the BiL participants, by contrast, is essentially the mirror image of that seen in the High judgment task: (1) Across contour shapes, target *DAY* contours were judged **lower** in pitch than the highest reference contours, but higher than the lowest references. (2) As plateau length increases, participants' 'lower target' judgments increase earlier. Adopting the same statistical analysis reported above, the best model, based on all-subset selection, contains interaction effects between both reference level and group (β=-.498, z=-9.54, p<.001), and contour shape and group (β=-.343, z=-4.57, p<.001). Pairwise comparisons via model re-leveling show that contour shape has a significant effect on BiL participants' responses (β=.411, z=7.56, p<.001), but not NE participants' responses (β=.068, z=1.31, p=.190).

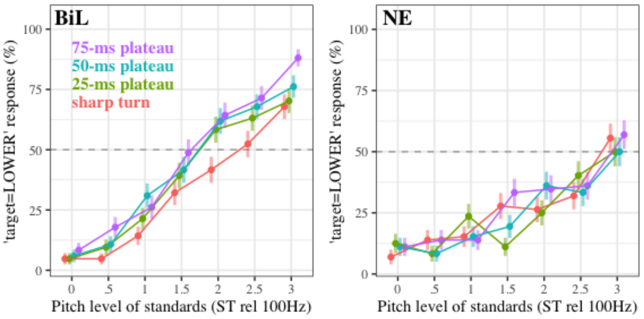

Figure 6: Percentage 'target is LOWER' judgments for the four contour shapes, as a function of the reference level against which they were compared.

4 DISCUSSION AND CONCLUSION

The lack of a "plateau effect" for NE participants in the Low judgment task could be related to their overall poor performance in judging the relative scaling of Low accents. ([15] implicate this same lack of acuity in the attenuation of the plateau effect that they report for Lows.) The fact that BiL listeners show both a plateau effect for Lows, and equivalent performance in discriminating Low and High F0 levels, is consistent with such an account.

It is not immediately obvious why NE listeners consistently rated Low target contours relatively higher in pitch than did their BiL counterparts. Note, however, that if listeners indeed evaluate pitch accent scaling through some form of F0-averaging over accented syllables, as suggested by [10]–[12] and others, then their responses are in fact in the expected direction: Mean F0 levels during accented *DAY* syllables in the low judgment task were as follows: 122.4 Hz (sharp turn), 120.3 Hz (25-ms plateau), 118.2 Hz (50-ms plateau), and 117 Hz (75-ms plateau), while the highest reference level was 119 Hz. NE listener responses effectively mirror this pattern. In this case, however, we must ask both why the BiL listeners failed to show this pattern, and why neither group showed its analogue (with peaks sounding lower than nearly all reference levels) in the High judgment task. See [3] for an explanation of a similar pattern in terms of weighting F0 samples by "recency".

To the question, however, of why BiL listeners performed both better on the discrimination of Low levels, and more symmetrically across the High and Low judgment tasks, we offer an explanation based on differences in the language experience of the two groups. As mentioned above, Cantonese is a lexical tone language that contrasts multiple pitch levels. Low tones realized at different F0 levels are used to signal different lexical meanings (e.g. low level T6 vs. low falling T4). Assessing the relative pitch of F0 targets in the lower portion of the pitch range is thus a frequent and necessary task in the perception of Cantonese.

English does make use of both High and Low targets in its intonation system (H* vs. L*, H- vs. L-, H% vs. L%). Implementation of Lows, however, is such that while High targets undergo a great deal of contextual variation (signaling, e.g., differences in focus type, paralinguistic emphasis, syntactic embedding, presence or absence of downstep, etc.), Low targets are reportedly less variable, most often hovering near the bottom of the speaker's pitch range [21]. Reports of how Low tonal targets across languages behave under differences in focus or emphasis are conflicting at best. (Early reports of lowering, e.g., [22] coexist with reports of no change, or even raising, in other sources–see [23, Sec. 10.4.1] for an overview.) Likewise, in English, there is no L*- analogue to the contrastive accentual downstepping patterns we find for High pitch accents. In fact, when bitonal accents such as L*+H participate in downstepping patterns, it is the peaks of the H, more saliently than the valleys of the lows, that encode the relation [24, p. 226].

Arguably, therefore, for NE speakers, the relative scaling of Low pitch targets is not linguistically meaningful, or at least has dramatically lower functional load than the relative scaling of Highs. This may cause NE listeners to attend to it less in speech, and to perform worse when judging it in experimental contexts. L1 Cantonese speakers, by contrast, must regularly assess the relative scaling of Low pitch targets, and therefore may both attend to it more strongly in speech, and judge it more accurately in experimental contexts. We might hypothesize then that the NE listeners' pattern of results in the Low task represents a non-linguistic processing mode for pitch, while the NE pattern in the High task, and the BiL pattern in both tasks, represents a linguistic processing mode. If this is true, however, then the plateau effect must be a phenomenon specific to linguistic processing of pitch, which may be contradicted by the non-speech results from [15]. Further research will be necessary to resolve these questions.

Jeon and Heinrich [15] suggest that the lower functional load English places on the discrimination of Low F0 targets may be a result of universally weaker domain-general auditory abilities in this area. If this is true, then the performance of our BiL listeners suggests that this deficit can be overcome through experience with one's native language. Alternatively, we might wonder whether in English the causality relation is better seen as reversed: Perhaps the lack of functional load on the discrimination of relative scaling among Lows is in fact the source, rather than a consequence, of listeners' weaker performance in perception.

5 REFERENCES

[1] J. P. Evans, “High is not just the opposite of Low,” *J. Phon.*, vol. 51, pp. 1–5, 2015.

[2] J. Barnes, A. Brugos, S. Shattuck-Hufnagel, and N. Veilleux, “On the nature of perceptual differences between accentual peaks and plateaux,” in *Understanding Prosody*, O. Niebuhr, Ed. Berlin: De Gruyter, 2012, pp. 93–118.

[3] M. D’Imperio, “The role of perception in defining tonal targets and their alignment,” PhD dissertation, The Ohio State University, 2000.

[4] J. ’t Hart, “F0 stylization in speech: straight lines versus parabolas,” *J. Acoust. Soc. Am.*, vol. 90, no. 6, pp. 3368–3370, 1991.

[5] R.-A. Knight, “The shape of nuclear falls and their effect on the perception of pitch and prominence: peaks vs. plateaux,” *Lang. Speech*, vol. 51, no. Pt 3, pp. 223–244, 2008.

[6] R.-A. Knight and F. Nolan, “The effect of pitch span on intonational plateaux,” *J. Int. Phon. Assoc.*, vol. 36, no. 1, pp. 21–38, 2006.

[7] O. Niebuhr, M. D’Imperio, B. G. Fivela, and F. Cangemi, “Are there ‘shapers’ and ‘aligners’? Individual differences in signalling pitch accent category,” in *Proc. ICPhS XVII*, Hong Kong, 2011, pp. 120–123.

[8] J. Barnes, A. Brugos, N. Veilleux, and S. Shattuck-Hufnagel, “Voiceless intervals and perceptual completion in F0 contours: Evidence from scaling perception in American English,” in *Proc. ICPhS XVII*, Hong Kong, 2011, pp. 108–111.

[9] J. Barnes, A. Brugos, N. Veilleux, and S. Shattuck-Hufnagel, “Segmental Influences on the Perception of Pitch Accent Scaling in English,” in *Proc. SP-2014*, Dublin, Ireland, 2014, pp. 1125–1129.

[10] J. Barnes, N. Veilleux, A. Brugos, and S. Shattuck-Hufnagel, “The effect of global F0 contour shape on the perception of tonal timing contrasts in American English intonation,” in *Proc. SP-2010*, Chicago, IL, USA, 2010, p. Paper 445.

[11] J. Barnes, N. Veilleux, A. Brugos, and S. Shattuck-Hufnagel, “Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology,” *Lab. Phonol.*, vol. 3, no. 2, pp. 337–383, 2012.

[12] J. Barnes, A. Brugos, N. Veilleux, and S. Shattuck-Hufnagel, “On (and off) ramps in intonational phonology: Rises, falls, and the Tonal Center of Gravity,” *J. Phon.*, vol. 85, 101020, 2021.

[13] C. d’Alessandro, S. Rosset, and J. P. Rossi, “The pitch of short-duration fundamental frequency glissandos,” *J. Acoust. Soc. Am.*, vol. 104, no. 4, pp. 2339–2348, Oct. 1998.

[14] P. Mertens, “The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model,” in *Proc. SP-2004*, Nara, Japan, 2004, pp. 549–552.

[15] H.-S. Jeon and A. Heinrich, “Perceptual asymmetry between pitch peaks and valleys,” *Speech Commun.*, vol. 140, pp. 109–127, 2022.

[16] J. Gandour, “Perceptual dimensions of tone: Evidence from Cantonese,” *J. Chin. Linguist.*, vol. 9, no. 1, pp. 20–36, 1981.

[17] C. Gussenhoven, “On the intonation of tonal varieties of Englishes,” in *The Oxford handbook of world Englishes*, M. Filppula, J. Klemola, and D. Sharma, Eds. New York, NY: Oxford University Press, 2017, pp. 569–598.

[18] K. K. Luke, “Phonological re-interpretation: The assignment of Cantonese tones to English words,” in *The 9th International Conference on Chinese Linguistics*, 2000.

[19] P. Boersma and D. Weenink, “Praat: doing phonetics by computer (Version 6.1.10),” 2020. [Online]. Available: http://www.praat.org/

[20] G. Ward and J. Hirschberg, “Implicating uncertainty: The pragmatics of fall-rise intonation,” *Language*, vol. 61, no. 4, pp. 747–776, 1985.

[21] M. Liberman and J. Pierrehumbert, “Intonational invariance under changes in pitch range and length,” in *Language Sound Structure*, M. Aronoff and R. Oehrle, Eds. Cambridge, MA: MIT Press, 1984, pp. 157–223.

[22] J. B. Pierrehumbert, “The phonology and phonetics of English intonation,” PhD dissertation, Massachusetts Institute of Technology, 1980.

[23] J. Barnes, H. Mixdorff, and O. Niebuhr, “Phonetic variation in tone and intonation systems,” in *The Oxford Handbook of Language Prosody*, C. Gussenhoven and A. Chen, Eds. Oxford: Oxford University Press, 2020, pp. 125–149.

[24] M. Grice, “Leading tones and downstep in English,” *Phonology*, vol. 12, no. 2, pp. 183–233, 1995.