CUE SHIFTING IN THE PERCEPTION OF SHANGHAINESE TONE SANDHI PATTERNS

Zihao Wei

Chinese University of Hong Kong
zihaowei@cuhk.edu.hk

ABSTRACT

The role of cue reweighting in reshaping tone sandhi patterns is rarely discussed. The consensus in Shanghainese disyllabic prosodic words is that F0 lexical tonal differences are neutralized in the second syllable and that VOT and closure duration are the primary cues in maintaining the voicing/register contrast. To investigate perception strategies in different age groups, an AXB forced choice experiment is used. The findings of the current study are generally consistent with the hypothesis that change occurs when listeners reanalyze coarticulatory effects used in speech production (e.g., Ohala, 1993). The F0 split in sandhi patterns is most likely due to younger speakers no longer associating stable VOT differences with the onset stop voicing contrast in the second syllable and instead begin to associate the difference between high- and low-register words with the voicing-condition F0 difference. This cue shift exemplifies how perception leads to the process of reanalysis.

Keywords: Tonogenesis, Cue weighting, Sound change, Sandhi pattern, Shanghainese

1. INTRODUCTION

It is well-known that obstruents affect the F0 of the following vowel, and the vowels following a voiceless onset tend to have relatively higher F0 in many languages (e.g., Kirby & Ladd, 2015). Furthermore, in tonal languages where F0 differences indicate lexical contrasts, the relation between voicing and F0 is often manifested as co-occurrence restrictions between tone and the onset consonant of tone-bearing syllables.

In Shanghainese, a disyllabic prosodic word gets its surface sandhi tone by delinking the tone of the second syllable and spreading the tonemes of the first syllable over the two syllables. For example, the prosodic word /tsɔ^{3a} vɛ^{3b}/ 'fried rice' that could be autosegmentally represented as /ʦ^{h}ɔ^{MH} vɛ^{LH}/ has a surface form [ʦ^{h}ɔ^{M} vɛ^{H}] (e.g., Yip, 1995). In Shanghainese, as traditionally described, the functional cost of neutralizing tone in second syllables is mitigated by the preservation of vocal fold vibration. Some previous studies suggested that vocal fold vibration is no longer the primary cue (e.g., Chen & Wang, 2012), while others show that the two categories have multidimensional acoustic and articulatory correlates, including a significant difference in Closure of Duration of onset stop (CD) (Shen, Wooter & Wang, 1987) and in breathiness concentrated mainly at the onset of the following vowel (e.g., Ren, 1992).

The lexical contrasts can still be maintained when the primary cue changes. Such a change is reminiscent of tonogenesis, a type of cue shift that has been studied extensively (e.g., Hyman, 1973). The result of our previous production study (Wei, 2022) shows that the tone sandhi pattern in Shanghainese is changing: in the second syllable of a sandhi domain, F0 is overtaking CD and VOT as the primary cue in younger generations. In this work, we will examine the role of these cues in perception and whether there are generational differences in the relative importance of these different cues. Research on the mapping between production and perception cues is very inspiring when looking at an ongoing diachronic change. Only when we know how cues shift in both production and perception can we further draw a picture of the first trigger in a particular case of sound change where the cue shifting in both production and perception on the timeline is well established.

Cue weighting is a process of conceptualizing, quantifying, and ranking the cues associated with a contrast, which encompasses both perception and production (Schertz, 2019). The weighting of a cue for a phoneme may show variations in both production and perception depending on age, and the weights of cues in production and perception do not always coincide. Research on tonogenesis in Korean finds that speakers of different age groups prefer different strategies in speech perception (e.g., Kang & Guion, 2008; Kang, 2009, 2014; Silva, 2006; Wright, 2007). Meanwhile, studies on the production and perception of /u/- and /ʊ/-fronting of Standard Southern British English (e.g., Harrington, 2012; Harrington, Kleber, & Reubold, 2008; Kleber, Harrington, & Reubold, 2012) revealed that the relationship between production and perception is unstable during diachronic change. As a result, the comparison of cue weighting in perception and production bears on the theoretical issue of the processes underlying sound change.

2. EXPERIMENT

To test the hypothesis that both F0 and VOT are prominent cues in the perception of domain-medial onset stops, two separate sets of AXB forced choice experiments are used to investigate the participants' perception strategies in different age groups of Shanghainese speakers.

2.1. Participants

All the experiments in this study were conducted online using the Gorilla online application (https://app.gorilla.sc/admin/home) with 76 native Shanghainese speakers. There were 36 young participants (age range 20-40, mean 27.83), 24 middle-aged participants (age range 40-60, mean 49.63), and 16 elder participants (age > 60, mean 64.23). All participants spoke Standard Mandarin with different degrees of Shanghainese interference.

2.2. Stimuli

Three acoustic properties of the second syllable were manipulated along acoustic continua: the F0 slope, the VOT and the closure duration of the onset. The synthesized stimuli were produced based on the natural speech of a female Shanghainese speaker in her 20s. We chose two minimal pairs: 小巴 /ɕiɔ^{hL} pa^{H}/ "minibus" vs. 小排 /ɕiɔ^{hL} ba ^{H}/ "rib", 大巴 /da^{lL} pa^{H}/ "bus" vs. 大排 /da^{lL} ba^{H}/ "chop". The duration of the target vowel was normalized to 160 ms, and the onset duration of the domain-medial /p/ (i.e., Closure duration + VOT) was set to 125 ms, while that of the domain-medial /b/ was 35 ms. Next, the F0 at the offset of the vowel of /pa/ was set to 260 Hz, and that of the offset of the vowel of /ba/ was set to 220 Hz as in a mid-high flat tone. The normalization is based on the natural range of the speaker. Finally, five fundamental frequency curves with different slopes were generated on each of the second syllables by decreasing the F0 of their initial points in 10Hz steps. The specific stimuli information of F0 after the manipulation is presented in Figure 1.

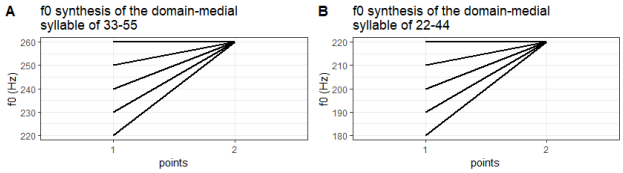

Figure 1: Steps of F0 onset of synthesized stimuli

Table 1 illustrates the VOT, closure duration, and F0 manipulations for one token as a cube in a 3-dimensional acoustic space (2 sandhi patterns × 4 F0 onset × 3 closure duration × 4 VOT = 60 stimuli). The 3a-1a/3b sandhi tone is shown on the top left side, and the 3b-1a/3b sandhi tone is shown on the bottom left side. Altogether, the stimuli varied in VOT, closure duration, and F0 slopes.

	ɕiɔ^{33} Pa^{55}	da^{22} Pa^{44}
	2nd syllable	2nd syllable
F0 onset (Hz)	260, 250, 240, 230, 220	220, 210, 200, 190, 180
Closure Duration (ms)	115, 70, 25	115, 70, 25
VOT (ms)	Voiceless: 10 Voiced: -25, -70, -115	Voiceless: 10 Voiced: -25, -70, -115

Table 1: Demonstration of the manipulation of VOT (four levels), Closure Duration of the onset (three levels), and F0 onset (five levels) of the second syllable.

2.3. Procedures

Each participant was then asked to listen to each stimulus five times in the experiment. As a result, each speaker had to identify 300 tokens (60*5). Participants were asked to complete a training session consisting of 20 disyllabic prosodic words to understand the requirements of the experiment. The training instructions asked participants to listen to a stimulus and to select one of the two corresponding disyllabic prosodic words that appeared on the screen by pressing A or D within ten seconds after hearing the recording. The formal experiment consisted of four blocks containing 75 tokens, with a mandatory one-minute break after each section. All responses given in less than 160 ms were removed as we considered them too fast to be valid. For the remaining data points, we remove data points smaller than Q1 - 1.5 * (Q3 - Q1) (Q1: lower tertile, Q3: higher tertile), and data points larger than Q1 + 1.5 * (Q3 - Q1). The final response times obtained for all data points are between 160 ms and 3000 ms. In the perception study, we use the LDA to weight different cues in perception (Schertz, 2019).

3. RESULTS

Looking at Figure 2 below, we can see that the slope of the F0 curve affected the perception of the voicing contrast in second syllables in all age groups, irrespective of VOT and closure duration. Broadly speaking, the lower the F0 was (i.e., the larger the slope of the F0 curve was), the more likely the stop onset was perceived as voiced, and the higher it was,

the more likely the stop onset was perceived as voiceless. In general, CD played a secondary role in the perception of second syllable, but this secondary role was much stronger in older than younger participants. VOT, on the other hand, had little effect on the perception in either of the three age groups.

In general, young speakers highly relied on the slope of F0 in their perception, and their perception was categorical: changes in CD had a relatively small effect on the perception of the voicing contrast. In the perception of elder and middle-aged participants, F0 was less categorical and CD played a greater role (middle-aged speakers fall between older and younger ones). Finally, VOT had little effect in the perception of all three age groups.

In summary, we found that the perceptual strategies of the three age groups changed over time:

- The elder group relied heavily on both CD and the slope of the F0 curve, but did not seem to rely on VOT in perception, despite using in in production.
- The middle-aged group also used F0 slope as a primary cue but relied significantly less on CD.
- The young group tended to use the slope of the F0 curve as the primary cue and used CD as a much more peripheral cue than other groups.

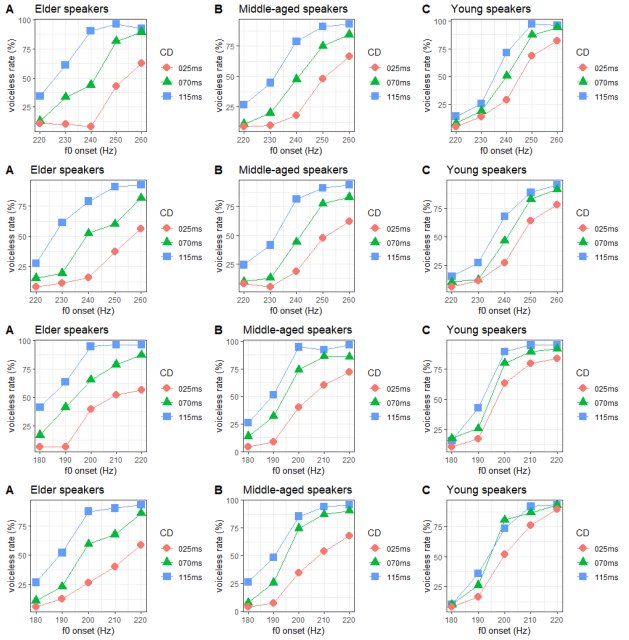

Figure 2: Rate of perceived as voiceless in 3a-1a/3b sandhi tone, VOT > 0 (1st row); in 3a-1a/3b sandhi tone, VOT < 0 (2nd row); in 3b-1a/3b sandhi tone, VOT > 0 (3rd row), and in 3b-1a/3b sandhi tone, VOT < 0 (4th row)

	Age group	X-squared	*p*-value	error rate
	Elder	666.86	$< 2.2e^{-16}$	23.37%
/pɔTɛ3a-1a/3b/	Middle	1057	$< 2.2e^{-16}$	22.32%
	Young	2016.3	$< 2.2e^{-16}$	18.41%
/bɔTɛ3b-1a/3b/	Elder	603.88	$< 2.2e^{-16}$	24.22%
	Middle	1166.1	$< 2.2e^{-16}$	20.45%
	Young	1982.1	$< 2.2e^{-16}$	17.75%

Table 2: LDA Results of the perception experiments

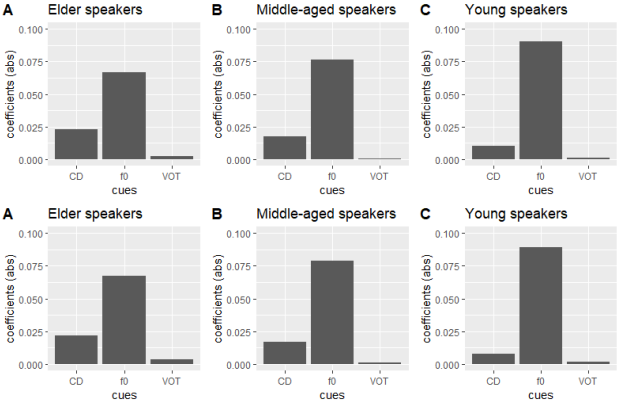

Figure 3: LDA results of perception, in sandhi tone 3a-1a/3b (top) and 3b-1a/3b (botton).

An examination of LDA results, in Figures 6 and 7 confirmed the observations made from the raw results. VOT, an outstanding production cue among the elder speakers, was the least valued perceptual cue for all three age groups. For elder speakers, F0 was already the most important perceptual cue, followed by closure duration. This pattern was equally valid for all three age groups, the main difference between them being the relative importance of F0 slope, that gradually increased for the middle-aged and young groups. In contrast, the importance of the closure duration diminished with age. Based on the findings of the perception experiment, we can say that the perception strategies used by the three generations are similar, but vary quantitatively.

4. DISCUSSION AND CONCLUSION

In perception, F0 is always the primary cue for all three age groups, and CD is gradually losing its importance as a secondary cue. VOT, as opposed to F0 and CD, has little effect in the perception of the three age groups. However, the limited role of VOT could be due to our speaker being a 24-year-old young woman. Although in her actual pronunciation, the VOT of syllables with phonologically voiced onset stops in the second syllable is negative, we cannot exclude that Shanghainese listeners are aware that young speakers rely less on VOT and more on F0 and adjust their perception to the young voice used in the experiment. Therefore, to further investigate the changes in the tone sandhi pattern of disyllabic prosodic words in Shanghainese, we would need to test stimuli produced by middle-aged and elder speakers as well.

5. ACKNOWLEDGMENT

This work is based on a section of the author's memoir in uOttawa, supervised by Professor Marc Brunelle.

6. REFERENCES

[1] Chen, Z., & Wang, Y., Roles of F0 and closure in the voiced-voiceless distinction for initial stops in Wu dialects—take Shanghainese for example. *In Tonal Aspects of Languages-Third International Symposium*. 2012.

[2] Harrington, J., The coarticulatory basis of diachronic high back vowel fronting. In M.-. J. Solé & D. Recasens (Eds.), The initiation of sound change. Perception, production, and social factors (pp. 103–122). Amsterdam: John Benjamins Publishing Company. 2012.

[3] Harrington, J., Kleber, F., & Reubold, U., Compensation for coarticulation, /u/-fronting, and sound change in Standard Southern British: An acoustic and perceptual study. Journal of the Acoustical Society of America, 123, 2825–2835. 2008.

[4] Hyman, L., (Ed.) Consonant types and tone: Southern California occasional papers in linguistics. *Los Angeles: The Linguistic Program, University of Southern California*. 1973.

[5] Kang, K. H., & Guion, S. G., Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. The Journal of the Acoustical Society of America, 124(6), 3909–3917. 2008.

[6] Kang, K. H., Clear speech production and perception of Korean stops and the sound change in Korean stops (Doctoral dissertation). Eugene: University of Oregon. 2009.

[7] Kang, Y., Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. Journal of Phonetics, 45, 76–90. 2014

[8] Kirby, J. P., Ladd, D. R., Stop voicing and F0 perturbations: Evidence from French and Italian[C]//ICPhS. 2015.

[9] Kleber, F., Harrington, J., & Reubold, U., The relationship between the perception and production of coarticulation during a sound change in progress. Language and Speech, 55, 383–405. 2012.

[10] Ohala, J. J., Coarticulation and phonology [J]. Language and speech, 1993, 36(2-3): 155-170.

[11] Ren, N., *Phonation types and consonant distinctions: Shanghai Chinese*. Ph.D. Dissertation. The University of Connecticut. 1992.

[12] Schertz, J., Clare, E. J., Phonetic cue weighting in perception and production[J]. Wiley Interdisciplinary Reviews: Cognitive Science, 2020, 11(2): e1521.

[13] Shen, Z., Wooters, C., & Wang, W. S. Y., Closure duration in the classification of stops: A statistical analysis. *OSU Working Papers*, 35, 197–209. 1987.

[14] Silva, D. J., Variation in Voice Onset Time for Korean stops: A case for recent sound change. Korean Linguistics, 13(1), 1–16. 2006.

[15] Wei, Z. H., *Cue reweighting in Shanghainese sandhi patterns.* ExLing 2022: Proceedings 13th International Conference of Experimental Linguistics. 2022.

[16] Wright, J. D., Laryngeal contrast in Seoul Korean (Doctoral dissertation). Philadelphia: University of Pennsylvania. 2007.

[17] Yip, M., Tone in east Asian languages[J]. The handbook of phonological theory, 1995, 476: 494.