

Canonicity effect on perceived vowel duration: Evidence from Taiwan Mandarin speakers

Yu-An Lu & Sang-Im Lee-Kim

NATIONAL CHIAO TUNG UNIVERSITY

Perceived vowel duration is known to be influenced by f_0 characteristics of the vowel. For example, listeners tend to perceive vowels with dynamic f_0 as longer than those with static f_0 , vowels with rising contours as longer than those with falling contours, and vowels with high f_0 as longer than those with low f_0 (*General Perceptual Biases*) [1, 2, 3, 8]. However, studies have shown that ones' native phonological system may further shape perceived vowel duration [4]. Using Taiwan Mandarin (TM) as a test case, the present study examines the effect of canonical forms in speech perception [5, 6]. In particular, a reduced T3half [X²¹], low in pitch and short in duration, arising from contextual variation is compared to fullT3 [X²¹⁴]; the former occurs more frequently in TM, while the latter is the abstract canonical form of the T3 category with its full concave f_0 contour. The *Canonicity Bias* predicts that despite its low frequency and shorter phonetic durations in TM (see the corpus results below), T3 is likely to be overestimated due to its canonical complex contour and its variant T3half would likewise be overestimated based on its phonological association with the T3 category.

Using a corpus study based on a list of 112 Mandarin monosyllabic words produced by 10 young TM speakers, we first established relative durations associated with lexical tones as follows: T2 [X³⁵] ($M = 517$ ms) > T1 [X⁵⁵] (467 ms) > T3 [X²¹⁴] (404 ms) > T4 [X⁵¹] (348 ms). Unlike Mainland Mandarin where T3 is unambiguously longer than other tones [7], T3 in TM is much shorter and reduced. In a perception experiment, TM (10F, 10M) and Seoul Korean (SK) listeners (12F, 8M), a control group, were presented with duration-controlled CV syllables ([pa], [pi], [ta], [ti]) in five duration steps (290-320-350-380-410 ms) carrying one of the four lexical tones in Mandarin and the reduced T3half [X²¹]. They were asked to rate the relative duration of these stimuli on a scale of 1 (much shorter) to 7 (much longer) compared to an anchor stimulus ([pa] in 350 ms). The results showed that perceived vowel duration by SK listeners reflected general perceptual biases (Figure 1): contour tones (T2, T3, T4) were rated as longer than level tones (T1, T3half), and high- f_0 tones were rated as longer than low- f_0 tones (T1 > T3half, T2 > T3). TM listeners, on the other hand, overestimated the duration of vowels carrying T3 and T3half (T3 as the longest; T3half > T1). In a follow-up shadowing experiment, TM (15F, 5M) and SK speakers (8F, 12M) were asked to repeat the same stimuli in the perception experiment as accurately as possible. While SK speakers showed a faithful imitation of the auditory input as shown in the less variable temporal differences between different lexical tones, TM speakers' production largely mirrored the temporal manifestations of lexical tones observed in the corpus study (Figure 2). Interestingly, T3half was again hyperarticulated, produced as longer and with similar f_0 trajectories as fullT3 (Figure 3).

The finding that syllables carrying T3 were perceived as the longest by TM listeners, despite their short phonetic duration in production, indicates the significance of the canonical complex dynamicity of the T3 contour. The effect of canonicity is further evidenced by TM listeners' overestimation of T3half in production, demonstrating a clear association with the T3 category. This result provides evidence for the psychological reality of the stored knowledge of abstract canonical representations of lexical tones. These results are indicative of the pervasiveness of

canonicity in phonological processing and contribute evidence to the literature the empirical coverage of which has been limited, to date, to the segmental domain.

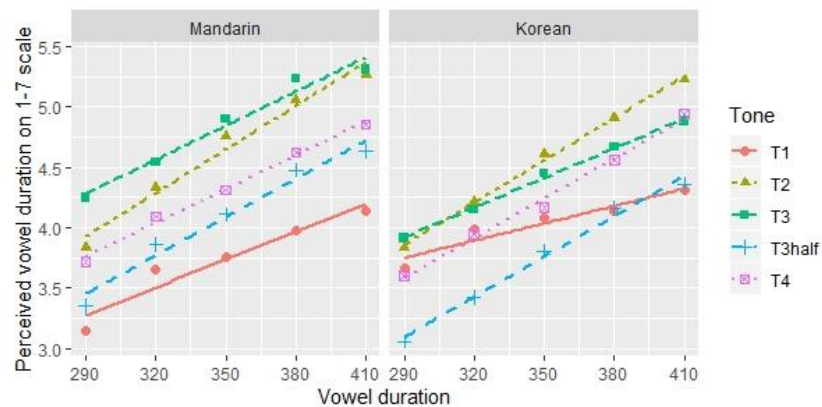


Figure 1. Predicted regression lines and average perceived vowel duration as a function of tone and duration of the stimuli

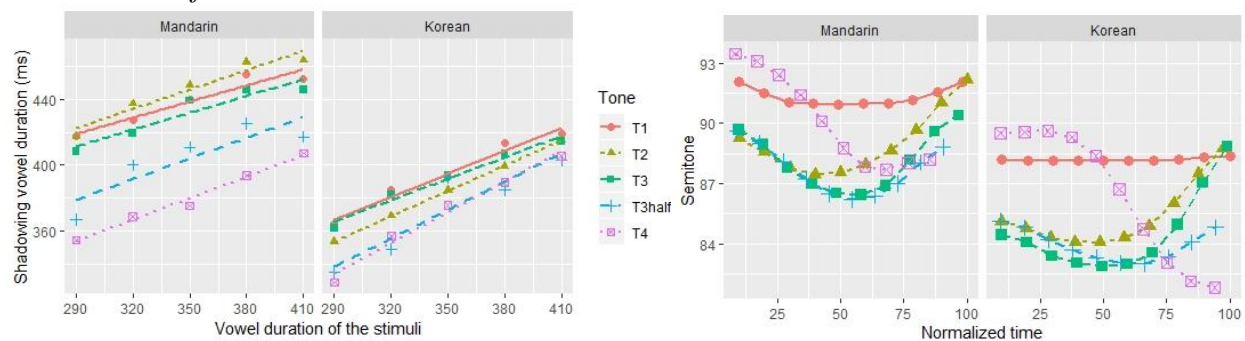


Figure 2. Predicted regression lines and average shadowing vowel duration as a function of tone and duration of the stimuli

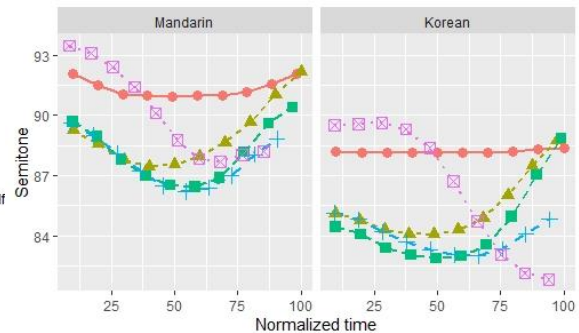


Figure 3. Time normalized shadowing contours as a function of tone

References

- [1] Cumming, R. (2011). The effect of dynamic fundamental frequency on the perception of duration. *Journal of Phonetics*, 39(3), 375-387.
- [2] Gussenhoven, C., & Zhou, W. (2013). *Revisiting pitch slope and height effects on perceived duration*. Paper presented at the INTERSPEECH 2013, Lyon, France.
- [3] Rosen, S. (1977). The effect of fundamental frequency patterns on perceived duration. *Speech Transmission Laboratory—Quarterly Progress and Status Report*, 18, 17-30.
- [4] Šimko, J., Aalto, D., Lippus, P., Włodarczak, M., & Vainio, M. (2015). *Pith, perceived duration and auditory biases: Comparison among languages*. Paper presented at the 18th ICPHS, Scotland, UK, August 10-14, 2015.
- [5] Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, 52, 322-338.
- [6] Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, 60, 487-501.
- [7] Wu, F., & Kenstowicz, M. (2015). Duration reflexes of syllable structure in Mandarin. *Lingua*, 164, 87-99.
- [8] Yu, A. (2010). Tonal effects on perceived vowel duration. *LabPhon* 10, 4(4), 151-168.