THE ROLE OF REDUNDANT TEMPORAL CUE ON PERCEIVED VOWEL DURATION: EVIDENCE FROM TONE LANGUAGE SPEAKERS

National Chiao Tung University

Yu-An Lu & Sang-Im Lee-Kim
yuanlu@nctu.edu.tw; sangimleekim@nctu.edu.tw

ABSTRACT

This study explores whether temporal information associated with different lexical tones influences the perception of vowel duration. In a perception experiment, Mandarin and Korean listeners rated the duration of duration-controlled CV syllables carrying one of the four lexical tones in Mandarin and a reduced Tone 3 (T3half, X^{21}). The results showed that perceived vowel duration by Korean listeners reflected general perceptual biases: contour tones were rated as longer than level tones, and high-*f0* tones (e.g., X^{55}) were rated as longer than low-*f0* tones (e.g., X^{21}). In direct contrast, although contour tones were generally rated as longer than level tones, Mandarin listeners overestimated the duration of vowels carrying T3, the longest tone in production: X^{214} (T3full) was rated as the longest, and X^{21} (T3half) was rated as longer than X^{55}. The findings suggest that perceived vowel duration is guided by redundant temporal cues available in one's native language as well as general auditory biases.

Keywords: perceived vowel duration, temporal cue, tone, pitch contour, Mandarin

1. INTRODUCTION

Perceived vowel durations have been shown to correlate with pitch height, pitch movement and ones' native prosodic system (e.g., [4, 7, 8, 10, 13]). For example, Gussenhoven & Zhou [4] tested tone (Mandarin) as well as non-tone (Dutch) language speakers in their perceived duration of vowels of different pitch heights and contours with duration manipulated on a 4-step continuum. The results showed that, independent of language background, the listeners perceived vowels of equal durations with high-*f0* as longer than those with low-*f0*. The results were interpreted as evidence of perceptual compensation: due to physiological restrictions, vowels with high-*f0* tend to be short in production, but are corrected in perception and are thus perceived as longer.

Pitch movement has also been shown to have an effect on perceived vowel duration [3, 6, 7, 8, 11]. For example, Cumming [3] tested French and German speakers' relative perceived duration of falling/rising/complex vs. level tones. The results showed that independent of their native language, listeners consistently gave longer duration judgments to dynamic *f0* contours than static *f0* contours. The results were taken to reflect a general perceptual bias wherein listeners simulate the articulatory gestures involved in the production of dynamic tones. Along the same lines, other studies have shown that vowels with rising contours are perceived as longer than those with falling contours (e.g., [8]), reflecting physiological difficulties in the production of a rising contour produced against natural airflow dynamics.

In addition, the prosodic system of one's native language has also been shown to have an effect on perceived vowel duration. For example, Šimko et al. [10] showed that languages with a quantity system (Estonian, Swedish and Finnish) made more precise duration judgments; speakers of Mandarin, a language lacking vowel length contrasts, on the other hand, performed poorly. Among the languages with vowel length contrasts, Estonian and Finnish use pitch movement to co-signal quantity contrast. These speakers showed even higher duration/pitch sensitivity.

Building upon the literature demonstrating the impact of pitch height, pitch movement, and native prosodic systems on the perception of vowel duration, the present study reports another factor — the phonetic knowledge of one's native language — that contributes to the perception of vowel duration. Specifically, it was tested whether redundant temporal cues associated with lexical tones systematically influence how vowel durations are perceived.

2. TARGET LANGAUGE AND PREDICTION

The current study tests the perception of Mandarin vowels carried by different lexical tones. Mandarin is a quantity insensitive language, but the phonetic durations of lexical tones vary. In a corpus study, Wu & Kenstowicz [12] established the following durational hierarchy: T3 (X^{214}, *M*= 407 ms) > T2 (X^{35}, *M*= 364 ms), T1 (X^{55}, *M*= 333 ms) > T4 (X^{51}) *M*= 286 ms). In addition, the canonical T3 (X^{214}), the longest tone in production, has a reduced variant T3half (X^{21}) when followed by a tone other than itself [14]. T3half is short in duration and low in pitch, and its *f0*-

trajectory no longer forms a contour shape (see Figure 1).

If perceived vowel duration is language-independent, listeners' perceived duration should follow the general perceptual biases outlined below. Each bias makes a specific prediction about vowels carried by different Mandarin tones.

Bias 1. Vowels with a dynamic *f0* should be perceived longer than those with a static *f0*:
T2 (X^{35}), T4 (X^{51}), T3 (X^{214}) > T1 (X^{55}), T3half (X^{21})

Bias 2. Vowels with a rising contour should be perceived longer than those with a falling contour: **T2 (X^{35}) > T4 (X^{51})**

Bias 3. Vowels with a high *f0* should be perceived longer than those with a low *f0*:
T1 (X^{55}) > T3half (X^{21})

However, if the perceived duration is, to some extent, informed by the phonetic knowledge associated with different lexical tones in the native language, T3 is likely to be overestimated due to its long duration in production. The lexical link to the T3 category may further lead to the overestimation of T3half as well. To test these predictions, the present study examined perceptual patterns of native speakers of Mandarin in comparison with those from Korean listeners without any experience in tone languages.

3. EXPERIMENT

3.1. Methodology

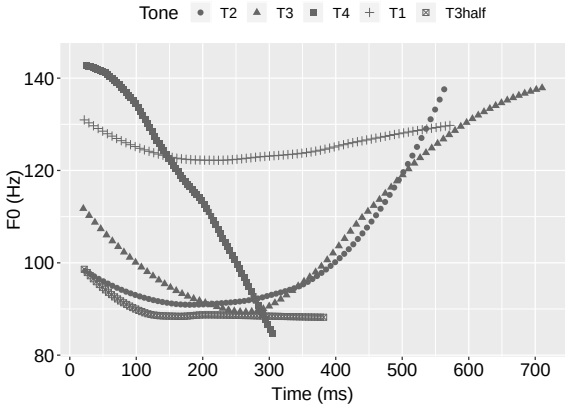
3.1.1. Participants

20 Taiwan Mandarin speakers (10F, 10M; ages 19-36; *M*= 22.89) were recruited to serve as the target group, and 20 Korean speakers (12F, 8M; ages 19-34; *M*= 24.7) served as the non-tone language baseline group. Although Korean is traditionally known to be quantity-sensitive, vowel length contrasts have been shown to have levelled for young Seoul Korean speakers [5]. None of the Korean participants had learned Mandarin or spoke Korean dialects with pitch accents. None of the participants reported hearing or speaking deficiencies. All participants were compensated monetarily for their time.

3.1.2. Stimuli

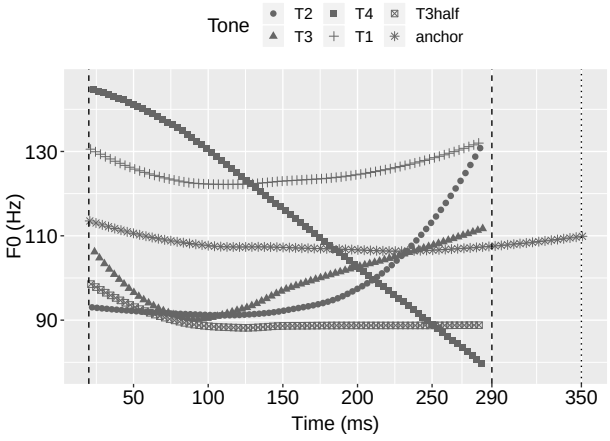
Four CV syllables, [pa], [pi], [ta], and [ti], were selected as the target syllables. Unaspirated consonants were chosen because Gussenhoven & Zhou [4] showed that speakers might perceive aspiration as part of the vowel duration [4]. Two vowels [i] and [a] were used to increase the variability of the stimuli. All syllable-tone combinations are well-formed and possible words in Mandarin to avoid lexical biases for or against certain tones. These syllables were produced by a phonetically trained male native Taiwan Mandarin speaker. Figure 1 shows the *f0* trajectories and durations of the recorded tokens.

Figure 1: Mean *f0* trajectories and durations of the naturally produced tokens



These tokens were then resynthesized into a five-step duration continuum, 290–320–350–380–410 ms, falling within the duration range of Mandarin syllables, using the Pitch Synchronous Overlap and Add (PSOLA) algorithm in Praat [2]. Figure 2 shows the example stimuli resynthesized to 290 ms. A fixed stimulus [pa] set to be 350 ms in mid-level tone was also resynthesized to serve as an anchor.

Figure 2: Example stimuli [pa] resynthesized in 290 ms and the anchor stimuli (*) set to be 350 ms in mid-level tone



3.1.3. Procedure

The 100 resynthesized stimuli (4 syllables [pa, pi, ta, ti] x 5 tones [T1, T2, T3, T3half, T4] x 5 duration steps [290–320–350–380–410 ms]) were presented twice in two blocks using E-Prime [9]. The 200 trials

were randomized for each participant and presented in an AX task in which "A" was the anchor stimulus ([pa] fixed at 350 ms in mid-level tone) and "X" was the target stimulus. The ISI was set as 800 ms. Participants were instructed verbally in their respective languages as well as with written instructions on the computer screen. They were asked to listen to each pair of sounds and judge the relative duration of the target stimulus compared to the anchor stimulus. The judgments were made on a 7-point scale, with 1 being the shortest, 4 the same, and 7 the longest.

Six practice trials were presented before the experiment to familiarize participants with the task. These trials contained the experimental stimuli in either the longest duration step (410 ms) or the shortest (290 ms) randomly chosen from the four syllables and five tone combinations. The experiment was conducted in sound-attenuated booths using high-quality headphones in two separate locations, one at National Chiao Tung University for Taiwan Mandarin listeners and the other at National Seoul University for Korean listeners. The total duration of the experiment was around 20 minutes.

3.2. Results

To interpret the results, a linear mixed-effects regression model was fitted to the data in R using the *lme4* package [1]. The dependent variable was the participants' judgments of vowel durations converted to *z*-scores for each speaker. The model included fixed effects for Group (Mandarin, Korean), Tone (T1, T2, T3, T3half, T4), Duration (5 steps, converted to *z*-scores), and Vowel ([a], [i]). The model included random intercepts for Participant as well as by-participant random slopes for the fixed factors. In addition, the model included an interaction term for Group, Tone, and Duration.

The model using T4 as a baseline showed T3 and T2 were perceived as longer than T4 (p=.015 and p=.003, respectively) which were in turn perceived as longer than T3half and T1 (p=.007 and p=.002, respectively). The results reflect Perceptual Bias 1: all the dynamic tones (T2, T3, T4) were judged to be longer than the level tones (high-level T1 and low-level T3half) independent of language background. In addition, the Duration–Group interaction was significant (p=.002), reflecting the steeper slopes for the Korean group (Figure 3) than the Mandarin group (Figure 4). This result indicates Korean listeners have a higher sensitivity to acoustic vowel durations.

The three-way interactions between Tone–Duration–Group were significant for some tones, and as such, separate models were fitted for each language group to interpret the results. Except for the Group factor, the rest of the structure of the statistical model remained identical to the larger model. The results of the Korean group showed T2, a rising tone, was perceived as longer than T4, a falling tone (p=.002), consistent with Perceptual Bias 2. The perceived duration of T3 was not significantly different from that of T4 (p=.208), but T4 was judged to be longer than T1 (p=.046) and T3half (p<.0001). In a separate model with T1 as a baseline, it was established that T1 was perceived as longer than T3half at a marginal level (p=.053). This result confirms Perceptual Bias 3 whereby high-level tones are perceived longer than low-level tones, all else being equal.

Figure 3: Mean of estimated perceived vowel durations by tone by Korean listeners

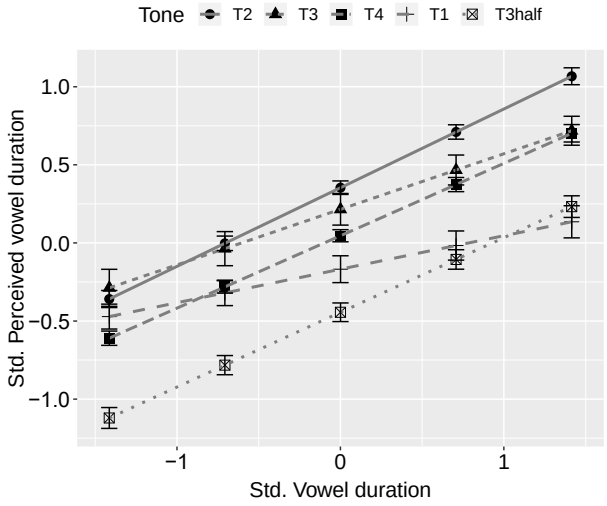
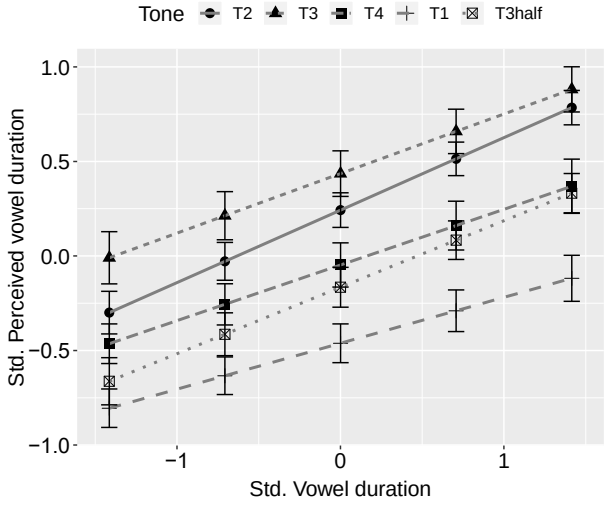


Figure 4: Mean of estimated perceived vowel durations by tone by Mandarin listeners



In contrast, the results of the Mandarin group (Figure 4) showed some patterns that could not entirely be accounted for based on the general perceptual biases. In particular, T3 was perceived as significantly longer than T4 (p=.041), which was not different from T2 (p=.098). A clear cross-linguistic

difference arises such that Korean listeners judged T2 the longest, while Mandarin listeners judged T3 the longest. In addition, T3half was not significantly different from T4 (p=.546) which was perceived as longer than T1 (p=.016). Again, T3half, the shortest tone for the Korean listeners, was equivalent to other contour tones for the Mandarin listeners.

The aggregated data of the perceived vowel duration carried by different tones are ordered below for each language group. Contour tones are indicated in italics and T3 categories in bold.

(longer) (shorter)
Kor: *T2(35)* > ***T3 (214)***, *T4 (51)* > T1(55) > **T3half(21)**
Mand: ***T3 (214)*** > *T2(35), T4(51),* **T3half(21)** > T1(55)

4. GENERAL DISCUSSION

The results of the perception study showed that, overall, dynamic tones were perceived as longer than static tones. This pattern was observed for both groups, indicating that Perceptual Bias 1 is stronger than linguistic experience. However, differences between language groups were evident, suggesting the importance of linguistic experience. In the following, we consider the source of the observed group differences.

First, between the level tones, T1 was perceived as longer than T3half for the Korean group while a reversed pattern was found for the Mandarin group. Consistent with the common pattern (Perceptual Bias 3) in which syllables with a high-*f0* elicit a longer perceptual duration than those with a low-*f0*, the results of the Korean group can be taken to support perceptual compensation: due to articulatory constraints, high-*f0* tones are often shorter in production than low-*f0* tones, and this asymmetry is corrected in perception such that high-*f0* tones are overestimated compared with low-*f0* ones, all else being equal.

The reversed pattern in the Mandarin group is thus intriguing. We attribute this to the lexical association between the members of the T3 category. Although T3half itself is low in pitch, the phonological knowledge of tonal allophones seems to have linked T3half to its unreduced counterpart, T3full, the longest tone in production. This seems to have driven the reversal of low-*f0* T3half above high-*f0* T1 in perception.

Second, among the contour tones, Korean listeners perceived T2 as longer than T3 while Mandarin listeners conversely perceived T3 as longer than T2. The Korean group's performance can, again, be taken as a reflection of perceptual compensation; syllables with a high-*f0* (T2, 3$\underline{5}$) are perceived longer than those with a low-*f0* (T3, 214). On the other hand, the overestimation of T3 by the Mandarin listeners reflects their phonetic knowledge of lexical tones, namely T3 has the longest duration in production.

The results, however, appear to contradict the findings in Gussenhoven & Zhou [4]. In their study, the performance of the Mandarin listeners did not follow the trend that was found in the current study. In line with the predictions based on perceptual compensation, Mandarin listeners perceived vowels as *longer* when their acoustic durations are *shorter*.

However, there is a crucial difference between the current study and the previous one. In Gussenhoven & Zhou [4], the stimuli were produced by a Russian speaker and were manipulated into different *f0* heights/contours which were not familiar to Mandarin listeners, while the stimuli in the current study maintained the pitch heights/contours of naturally produced tokens by a native Mandarin speaker. The only acoustic property being manipulated was the duration. The stimuli in the current study were manipulated from Mandarin speech and are therefore likely to have enabled a processing at the phonological level for the Mandarin listeners, which, in turn, drove a strong linguistic experience effect. Gussenhoven and Zhou's study, on the other hand, was designed to tap into pure phonetic processing and thus induced a pattern consistent with the language-independent general perceptual biases.

The higher sensitivity to vowel durations by the Korean group may reflect the contribution of the prosodic system of the native language. Although the vowel length contrasts have reportedly been levelled in the production of young Korean speakers, they might still retain the contrasts at the phonological level. Mandarin listeners' lower sensitivity to vowel duration, on the other hand, may be a reflection of the lack of length contrasts in their native sound system [10].

To summarize, the findings in the present study suggest the perception of vowel duration cannot be solely explained by universal mechanisms. Rather, perceived vowel duration is guided by both redundant temporal cues associated with different lexical tones drawn from one's linguistic experience as well as general perceptual biases.

5. ACKNOWLEDGEMENTS

We would like to thank the attendees at NCTU Sound Workshop for their comments and the RAs for running experiments, Yu-Ming Chang, Pei-Chun Chen, Yangyu Chen, Shao-Jie Jin, Cheng-Huan Lee, Yu Nan, Waan-Rur Lu, Yen-Ju Lu, and Sarang Jeong. The project was supported by MOST106-2410-H-009-031 to Yu-An Lu and MOST107-2410-H-009-016-MY3 to Sang-Im Lee-Kim.

6. REFERENCES

[1] Bates, D., Maechler, M., Bolker, B., & Walker, S. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Soft.* 67(1), 1–48.

[2] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot Int.* 5:9/10, 341–345.

[3] Cumming, R. 2011. The effect of dynamic fundamental frequency on the perception of duration. *J. Phonet.* 39(3), 375–387.

[4] Gussenhoven, C., & Zhou, W. 2013. *Revisiting pitch slope and height effects on perceived duration*. INTERSPEECH 2013, Lyon, France.

[5] Kwon, K.-K. 2003. Prosodic change from tone to vowel length in Korean. *Development in Prosodic Systems*. 67–89.

[6] Lehiste, I. 1976. Influence of fundamental frequency pattern on the perception of duration. *J. Phonet.* 4(2), 113–117.

[7] Pisoni, D. B. 1976. Fundamental frequency and perceived vowel duration. *J. Acoust. Soc. Am.* 59(S1), S39–S39.

[8] Rosen, S. 1977. The effect of fundamental frequency patterns on perceived duration. *Speech Transmission Laboratory—Quarterly Progress and Status Report*. 18, 17–30.

[9] Schneider, W., Eschman, A., & Zuccolotto, A. 2002. *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.

[10] Šimko, J., Aalto, D., Lippus, P., Włodarczak, M., & Vainio, M. 2015. *Pith, perceived duration and auditory biases: Comparison among languages.* 18th ICPhS, Glasgow, Scotland, UK, August.

[11] Wang, W. S. Y., Lehiste, I., Chuang, C. K., & Darnovsky, N. 1976. Perception of vowel duration. *J. Acoust. Soc. Am. 60*(S1), S92–S92.

[12] Wu, F., & Kenstowicz, M. 2015. Duration reflexes of syllable structure in Mandarin. *Lingua,* 164, 87–99.

[13] Yu, A. C. 2010. Tonal effects on perceived vowel duration. *Lab. Phonol*. 10, 4(4), 151–168.

[14] Zhang, J., & Lai, Y. 2010. Testing the rol of phonetic knowledge in Mandarin tone sandhi. *Phonology*. 27, 153–201.