The Impact of Accidental Gaps on Tonal Categorization

Tzu-Hsuan Yang, Shao-Jie Jin & Yu-An Lu

National Chiao Tung University (Taiwan)
thy2105@tc.columbia.edu, courtney0419003.fl04@nctu.edu.tw, yuanlu@nctu.edu.tw

Previous studies have demonstrated that speech perception may be shaped by a speaker's linguistic knowledge. For example, Ganong [1] employed a continuum of stops varying in VOT, where one endpoint was a word and the other a non-word (e.g., *task*-**dask*, **tash*-*dash*), and asked English native speakers to identify the sound they heard as a [t] or [d]. The results showed that the boundary of categorical perception was biased towards the endpoint which was a word, reflecting an effect of lexical knowledge. Using a similar method, Massaro and Cohen [2] showed that English speakers gave more [r] responses on a [tr]-*[tl] continuum but more [l] responses on a *[sr]-[sl] continuum, reflecting a bias from phonotactic restriction. Other kinds of knowledge that are relevant to this phenomenon include phonological relationship (e.g., [3]), syllable structure (e.g., [4]) and morphological alternation (e.g., [5]). While previous studies have demonstrated a sensitivity to segmental features, there is little evidence of speech sound categorization being shaped by suprasegmental information. The present study aims to investigate the role of suprasegmental information in speech perception by examining the effect of accidental gaps on Mandarin speakers' tonal categorization.

Mandarin is a tone language with four phonemic tones: high-level Tone 1 [X^{55}], rising Tone 2 [X^{35}], falling-rising Tone 3 [X^{214}], and falling Tone 4 [X^{51}]. However, not every allowable syllable carries all four tones. For instance, the syllable [ts^{h}u] can be combined with T1 ([ts^{h}u]^{55} "coarse"), T2 ([ts^{h}u]^{35} "die" in Old Chinese), and T4 ([ts^{h}u]^{51} "vinegar"), but not with T3 (*[ts^{h}u]^{214}). These syllable-tone combinations that could but do not exist are termed "accidental gaps" (e.g., [6]). The present study investigates whether tonal categorization may be biased by the knowledge of accidental gaps.

To examine Mandarin speakers' tonal categorization, we conducted a forced-choice identification experiment in which 20 Taiwanese Mandarin speakers (16F, 4M; ages 20-37, M = 22) participated. Mandarin CV syllables carrying a tone along one of two continua were selected: T1-T2 (level to rising) or T1-T4 (level to falling), among which, tone pairs were further selected so that each pair contained a word on one end and an accidental gap on the other (*T1-T2, T1-*T2, *T1-T4, T1-*T4). The eight syllables were naturally produced and recorded by a phonetically trained female Taiwan Mandarin speaker. 10-step continua were resynthesized from these endpoints using TANDEM-STRAIGHT [7]. This method allows for the resynthesis to be done from the entire CV endpoints so that *f0*, as well as other secondary cues to tone perception (e.g., duration and creakiness) were manipulated proportionally [8, 9]. Figure 1 shows the *f0* trajectories and duration of the resynthesized stimuli on the continua (left: T1-T2; right: T1-T4). Participants were instructed to listen to each of the stimuli randomly presented to them in E-Prime [10] and judge whether they heard T1 or T2/T4 by pressing the corresponding key on the keyboard as soon as they were sure. Participants' responses and response times (RTs) were recorded.

Mixed-effects logistic regression models were fitted using the lme4 package in R [11] to interpret the results. The dependent variable was the participants' tonal categorization, with T1 responses coded as 0 and T2/T4 responses coded as 1. The models included Endpoint (e.g., T1-*T4, *T1-T4) and Step (1 to 10, normalized), and an interaction term for Endpoint and Step. The models also included random intercepts and slopes for Participant and Token. We fitted separate models for each tone continuum (T1-T2 and T1-T4). The results showed an effect of Endpoint along the T1-T4 continua (p=.002), indicated by the earlier boundary shift on the *T1-T4 continuum (solid line) than on the T1-*T4 continuum (dash line). The results along the T1-T2 continua also showed a similar trend (p=.06). The RT results provide further support to this perceptual bias, in that, first, the RTs peaked at different steps along the continua, with earlier peaks for the *T-T continua and later for the T-*T continua; second, although listeners successfully

identified the gap syllables, they generally spent more time in making the decisions. To summarize, the findings in the present study showed that Mandarin listeners' responses were biased by accidental gaps. The results suggest that listeners' perceptual categorization is sensitive to segmental as well as suprasegmental (tonal) information.

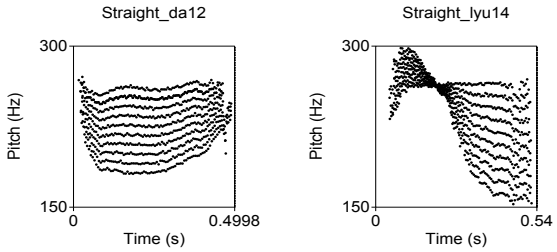

Fig.1 f0 trajectories of T1T2 continua for the token [ta] (left) and T1T4 continua for the token [ly] (right).

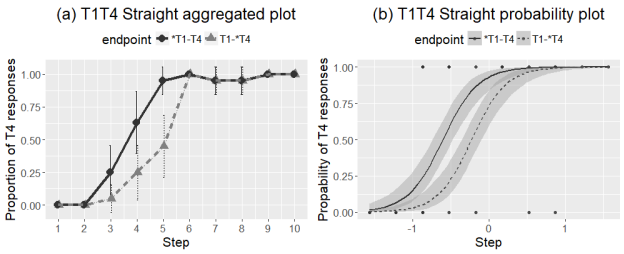

Fig.2 Aggregated and probability plot on the T1-T4 continua showing an earlier boundary on the *T1-T4 continuum than on the T1-*T4 continuum.

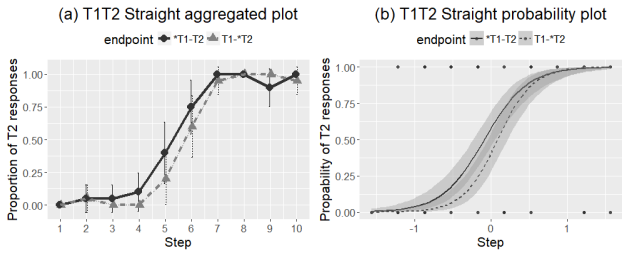

Fig.3 Aggregated and probability plot on the T1-T2 continua showing an earlier boundary on the *T1-T2 continuum than on the T1-*T2 continuum.

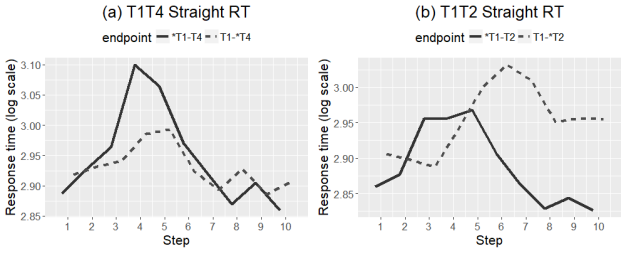

Fig.4 RTs of the identification showing earlier peaks for the *T-T continua and later for the T-*T continua.

References

[1] Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance 6*: 110-125.

[2] Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception and Psychophysics 34*(4), 338–348.

[3] Boomershine, A., Hall, K. C., Hume, E., et al. (2008). The impact of allophony versus contrast on speech perception. In: Avery P, Dresher E and Rice K (eds) *Contrast in Phonology*. Berlin: Mouton de Gruyter, 143-172.

[4] Dupoux, E., Kakehi, K., Hirose, Y., et al. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance 25*: 1568-1578.

[5] Ahn, M. (2008). Morphologically conditioned perceptual bias. *Proc. Chicago Linguistic Society, 44*:1-15.

[6] Duanmu, S. (2000). *The phonology of Standard Chinese*. Oxford: Oxford University Press.

[7] Kawahara, H., Morise, M., Takahashi, T., et al. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. Acoustics, Speech and Signal Processing Las Vegas, NV, 3933-3936.

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

[9] Yu, K. M. (2010). Laryngealization and features for Chinese tonal recognition. Proc. INTERSPEECH-2010 Makuhari, 1529-1532.

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

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