Production-perception relationship of Mandarin tones as revealed by critical perceptual cues

Keith K. W. Leung and Yue Wang

Language and Brain Lab, Department of Linguistics, Simon Fraser University, 8888 University Drive, Burnaby BC V5A 1S6, Canada
kwll23@sfu.ca, yuew@sfu.ca

Abstract: The relationship of lexical tone production and perception has not been well studied. Using Mandarin tone, this research tests the hypothesis that a production-perception link is revealed by critical perceptual cues. The critical status of perceptual tone cues was determined by perceptual cue weights, showing fundamental frequency (F0) contour as being more critical than height. Then, tone production features were examined for critical F0 contour (slope, curvature, turning-point location) and non-critical F0 height (mean, onset) cues. A production-perception correlation was found for F0 contour but not height cues, suggesting that critical perceptual cues dictate the relationship between production and perception.

1. Introduction

Speech perception theories predict a link between speech production and perception (e.g., Diehl et al., 2004; Galantucci et al., 2006), and have been tested by segmentally-based studies which have revealed a production-perception correlation (e.g., Beddor, 2015; Fox, 1982; Newman, 2003). However, it is not yet clear whether certain cues are more pertinent than other cues in establishing production-perception links. The current study examines the extent to which a production-perception correlation can be established based on the critical status of perceptual cues using Mandarin tones.

One proposal states that production-perception relationships exist in critical perceptual cues based on the finding that, among the three fricative acoustic cues (peak frequency, frication centroid, and skewness), only peak frequency can reveal a production-perception correlation (Newman, 2003). The relative importance of peak frequency was inferred from the goodness rating results, supported by acoustic modeling (Jongman et al., 2000). However, the critical status of these cues was not determined independently. In another study, Shultz et al. (2012) correlated the cue weights given to the production and perception of voice onset time (VOT), the primary cue of English stop perception, and onset F0, the secondary cue. Although the two cues did not reveal a statistically significant production-perception correlation, VOT displayed a trend of positive correlation, whereas onset F0 did not. The difference was attributed to the non-critical status of onset F0. These findings are informative for additional research on the influence of perceptual cues’ critical status on production-perception relationship.

Based on the aforementioned findings, the current study examines whether the Mandarin tone production-perception relationship is driven by critical perceptual cues. As reviewed earlier, the issue of a production-perception relationship as a function of the critical status of perceptual cues has not been thoroughly studied or consistently concluded (Newman, 2003; Shultz et al., 2012). Studying the production-perception relationship of lexical tones has an advantage, because the critical status of tone perception cues has been well established by the fact that native Mandarin listeners consistently weight F0 contour cues more strongly (i.e., reflecting a more critical status) than F0 height cues (Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007; Massaro et al., 1985). Moreover, systematic variations in F0 contour (relative to height) cues can better modulate Mandarin tone perception (Chang et al., 2016; Moore and Jongman, 1997; Shen et al., 1993), indicating the critical nature of the F0 contour cues. However, the critical nature of the F0 contour and height cues may change in different Mandarin tone contexts. For instance, Mandarin listeners weight F0 contour more strongly than F0 height only in resynthesized tones with low-to-mid F0 height levels (Massaro et al., 1985). Taken together, following Newman...
(2003), the $F_0$ contour should show a stronger production-perception relationship than the $F_0$ height. Little previous research has attempted to relate lexical tone production and perception in terms of the relative critical status of individual cues. Thus, the uniqueness of this study is to test this proposal using critical and non-critical cues.

As a lexical tone language, Mandarin has four tone categories differing in $F_0$ height and contour. The four tones are high-level (T1), high-rising (T2), low-dipping (T3), and high-falling (T4). The perceptually critical $F_0$ contour cue is usually represented by the overall $F_0$ slope of the tone contour (Jongman et al., 2017). The $F_0$ curvature, as modeled by a quadratic function, can also capture Mandarin tone contour shapes acoustically (Shih and Lu, 2015; Tupper et al., 2018). Moreover, the temporal location of the $F_0$ turning point (TP) and the $F_0$ decrease from onset to TP ($\Delta F_0$), which influence tone contour shape, have been shown to be critical in T2 and T3 perception (Moore and Jongman, 1997). In addition, varying TP alone has been shown to influence T2 and T3 perception (Shen et al., 1993). A tone with a TP also has a greater $F_0$ curvature value than a tone with a linear contour (Shih and Lu, 2015). These two cues are thus considered critical $F_0$ contour cues. On the other hand, cues characterizing $F_0$ height, the secondary (non-critical) cues, usually refer to $F_0$ mean (Jongman et al., 2017), and $F_0$ onset (Massaro et al., 1985).

The current study uses T2 as the target tone, since it has a low-to-mid $F_0$ height level and $F_0$ contour should have a more critical status than $F_0$ height in this context (Massaro et al., 1985). Moreover, its contrast with T1 involves a change of $F_0$ contour in terms of direction (i.e., rising versus level), and a change of $F_0$ height in terms of the overall $F_0$ mean and $F_0$ onset (Chang et al., 2016). In addition, the T2-T3 contrast involves a change in critical $F_0$ contour cues ($F_0$ curvature and TP), as well as non-critical $F_0$ height cues ($F_0$ onset) (Moore and Jongman, 1997). Therefore, this study can examine the production-perception correlation using critical (contour) and non-critical (height) cues separately.

In this study, $F_0$ slope, $F_0$ curvature, TP, $F_0$ mean, and $F_0$ onset were obtained from production and perception data. To examine a tone production-perception relationship as a function of the critical status of perceptual cues, this study performs a production-perception correlation for each cue, and expects to find a strong, positive correlation for critical $F_0$ contour cues ($F_0$ slope and $F_0$ curvature, and TP as a related temporal cue). In contrast, the non-critical $F_0$ height cues ($F_0$ mean and $F_0$ onset) should reveal a weaker correlation.

2. Method

2.1 Participants

Twenty-five native Mandarin female speakers who were undergraduate students at Simon Fraser University served as participants in this study (mean age: 22.2).

2.2 Production task

Participants produced four commonly used Mandarin monosyllabic words containing the syllable zhu with four tones, meaning “pig” in T1, “bamboo” in T2, “lord” in T3, and “pillar” in T4. The task was self-paced and conducted in a sound-attenuated booth. Each tone word was repeated 6 times. A total of 600 productions were recorded (4 tone words × 6 repetitions × 25 participants).

All T2 productions were analyzed acoustically in Praat (Boersma and Weenink, 2018). Two phonetically-trained native Mandarin listeners evaluated the accuracy of the stimulus tones. Five out of 150 T2 productions were error productions and were removed. $F_0$ values were then obtained at the 101 equidistant time points along a tone contour to provide data for the subsequent polynomial fits following Shih and Lu (2015) and Tupper et al. (2018). Equation (1) was used for $F_0$ normalization (Wang et al., 2003),

$$T = \frac{\log x - \log L}{\log H - \log L} \times 5,$$

(1)

where $x$ was $F_0$ value in Hz at any given point, $L$ and $H$ were the minimum and maximum $F_0$, respectively, of all four tones produced by the speaker.

Equations (2) and (3) were used to estimate $F_0$ slope and $F_0$ curvature, the critical cues (cf. Tupper et al., 2018),

$$F(t) = mt + k,$$

(2)

$$F(t) = at^2 + bt + c,$$

(3)

where $t$ represented the time elapsed from the tone onset. The linear coefficient of Eq. (2) ($m$) and the quadratic coefficient of Eq. (3) ($a$) represented $F_0$ slope and $F_0$ curvature, respectively.
Another critical cue, TP, was obtained at the temporal location relative to the total duration.

For non-critical cues, $F_0$ mean was obtained by averaging the $F_0$ values obtained from all measurement points in Praat. $F_0$ onset in normalized frequency $T$ was obtained at the starting time point of the tone contour.

2.3 Perception task

One female native Mandarin talker who did not participate in the other tasks of this study produced $zhu$ with four tones which were judged as correct productions of the intended tones by two phonetically-trained native Mandarin-speaking research assistants. Based on these natural productions, the tone contours of the perception stimuli were resynthesized by separately manipulating TP and $F_0$ onset, creating a perceptual space that situated the T2-like stimuli between two end points that simulate a T1 and T3.

All the resynthesized tone contours were set at duration of 410 ms, the mean duration of the speaker’s T1, T2, and T3 productions. The $F_0$ onset series (high to low bound) was created based on the talker’s T1 (295 Hz) and T3 (159 Hz) (Left panel of Fig. 1). Intermediate steps were created at a step size of 8 Hz (Jongman et al., 2017). As a result, the $F_0$ onset range encompassed the $F_0$ onset from T1 to T2, and to T3.

TP endpoints were based on the earliest and latest TP location of the talker’s T2 and T3 productions (i.e., 0% and 60% of the total tone duration). Each $F_0$ onset had a TP series, except for the high bound $F_0$ onset endpoint (a level tone), with the interval of 10% of the total duration (Right panel of Fig. 1). The change in TP consequently alters tone contour shape and is thus expected to critically influence perception (Moore and Jongman, 1997; Shen et al., 1993). $\Delta F_0$, another cue that can modulate T2-T3 perception (Moore and Jongman, 1997), was not varied since it would lead to a change in the relative position of $F_0$ onset—the other cue varied in this study.

As a result, an 18-step ($F_0$ onset) × 7-step (TP) grid of stimuli was formed, so that each tone item had a distinctive set of critical $F_0$ slope, $F_0$ curvature, and TP values, as well as non-critical $F_0$ mean and $F_0$ onset values.

The procedures of the perception task followed the Method of Adjustment task (Johnson et al., 1993), which required participants to select their preferred exemplars of T2. When the task began, a stimulus grid consisting of 126 boxes (18 $F_0$ onsets × 7 TPs) and a rating scale was displayed on a computer screen. When the participant clicked a box, one of the 126 $zhu$ stimuli was played over the headphones. Participants were instructed to first listen to each of the corner stimuli (i.e., the boundary tones) and rate each stimulus on a scale of 1 (poor exemplar of T2) to 5 (good exemplar of T2). Then, they had to find the box in the grid that best represented a T2 for them and rate their choice on the same rating scale. The rating data were used to examine whether a participant’s preferred exemplar of T2 fell inside the stimulus grid (i.e., the preferred exemplar should be rated higher than the boundary tones). To ensure participants determined the location of the preferred T2 exemplar auditorily (not visually), the orientations of the two axes or position of the axes were switched for each repetition, forming 8 orientation combinations (2 $F_0$ onset orientations × 2 TP orientations × 2 axis positions). This task was repeated 16 times (8 orientation combinations × 2 repetitions) for each participant.

![Fig. 1. Schematic representations of perceptual stimulus series of $F_0$ onset (left panel) and TP (right panel).](image-url)
3. Results

Participants’ T2 productions and perceived preferred T2 exemplars had a mean F0 slope of 4.82 T/s and 10.52 T/s, representing a rising contour; a mean F0 curvature of 31.88 T/s² and 18.92 T/s², indicating an upward opening parabolic shape; a mean TP of 25% and 30%; a mean F0 mean of 2.32 T and 2.56 T; and a mean F0 onset of 2.14 T and 1.02 T, respectively. All production and perception data are presented in Fig. 2, which show that the participants produced and perceived T2 with comparable ranges of acoustic data for all cues. In addition, the polynomial fits showed that F0 curvature yielded a higher R² value than F0 slope in both production (F0 curvature: 96% vs F0 slope: 67%) and perception (F0 curvature: 97% vs F0 slope: 90%).

To examine the correlations among the five cues, Spearman’s rank-order correlations were carried out for production and perception data separately. For production, TP was significantly correlated with F0 slope [\( \rho(23) = -0.529; \ p = 0.007 \)], F0 curvature [\( \rho(23) = 0.702; \ p < 0.001 \)], and F0 mean [\( \rho(23) = -0.398; \ p = 0.049 \)]. For perception, F0 slope and F0 curvature were significantly correlated [\( \rho(23) = 0.436; \ p = 0.030 \)]. TP was significantly correlated with F0 curvature [\( \rho(23) = 0.935; \ p < 0.001 \)] and F0 mean [\( \rho(23) = -0.561; \ p = 0.004 \)].

In addition, paired sample t tests showed that the preferred T2 exemplars [Mean = 4.8, standard deviation (SD) = 0.52] were rated significantly higher than the four corner stimuli (Mean < 3.05, SD < 0.98) [t(24) > 8.79, ps < 0.001], indicating that the participants’ preferred T2 stimuli fell inside the stimulus grid.

The above data show that the participants produced and perceived T2 with comparable ranges of acoustic data for all cues, indicating that a parity between their production and perception was maintained, which paves the way for correlation analysis below. In addition, the correlation analysis among the five acoustic cues showed that only certain cues were highly correlated with each other (e.g., TP and F0 curvature), suggesting that it was possible to compare the coefficients of the production-perception correlations for the cues that were not highly correlated (e.g., F0 slope and F0 curvature).

To further quantify the relationship between production and perception, a Spearman’s rank-order correlation was conducted to relate production data to perception data for each acoustic cue. Based on the assumption that the production-perception correlation would be positive, one-tailed correlation analysis was conducted.

A significant positive correlation was found for curvature [\( \rho(23) = 0.402; \ p = 0.024 \)], slope [\( \rho(23) = 0.378; \ p = 0.032 \)] and TP [\( \rho(23) = 0.391; \ p = 0.027 \)] (Fig. 3). No significant result was found for F0 onset [\( \rho(23) = 0.080; \ p = 0.352 \)] and F0 mean [\( \rho(23) = -0.022; \ p = 0.543 \)]. These results suggest that even though all cues displayed comparable ranges of acoustic values from both production and perception results, only critical F0 contour cues (i.e., curvature, slope, and TP) displayed a significant production-perception correlation.

4. Discussion and conclusion

This study examines whether a production-perception relationship is revealed through perceptually critical acoustic features (Newman, 2003), by comparing the Mandarin tone production-perception correlations established by F0 contour (critical) and F0 height (non-critical) cues (Gandour, 1983). The results of the current study revealed a significant positive production-perception correlation for F0 curvature, F0 slope, and TP, but not for F0 mean or F0 onset. This study defines the more strongly weighted F0 contour cues as perceptually more critical than the other acoustic cues.
less strongly weighted F0 height cues (Francis et al., 2008; Gandour, 1983; Massaro et al., 1985). Therefore, more critical perceptual cues contribute to a stronger a production-perception link compared to less critical perceptual cues, supporting the hypothesis of this study (Newman, 2003).

The current results further show that the strength of the production-perception relationship may differ among the critical cues. Among the two possible acoustic correlates of the F0 contour, the polynomial fits yielded a higher R² value for F0 curvature than for F0 slope, and it was the case for both participants’ T2 productions and their preferred T2 exemplars in perception. Therefore, F0 curvature explains the variance of T2 contours better than F0 slope, presumably because F0 curvature captures greater details of the T2 contour shape than F0 slope (Shih and Lu, 2015; Tupper et al., 2018). Future perception studies should also consider F0 curvature to be a representative cue for F0 contour of Mandarin tones. More importantly, the F0 curvature yielded a stronger production-perception correlation than the F0 slope (ρ = 0.402 versus 0.378). In addition, our production data did not show a significant correlation between F0 curvature and F0 slope. Therefore, this finding further extends the previous hypothesis, in that the strength of production-perception relationship may depend on the level of critical status of cues.

Likewise, the temporal feature TP was another critical cue that showed significant correlation strength. The critical status of TP can be attributed to its close relationship with other F0 contour cues. Note that TP showed significant correlation with F0 contour cues in both production and perception, and is also linked to the magnitude of F0 curvature in tone modeling (Shih and Lu, 2015). As a result, the current study shows that critical perceptual cues are adopted from both temporal and spectral domains in establishing production-perception links.

This study also supports our hypothesis that perceptually relevant but non-critical cues exhibit a weak production-perception relationship, as demonstrated by the results for non-critical F0 height cues: F0 mean and F0 onset. Their non-critical status is supported by perceptual weighting studies, but F0 mean has been shown to be the second perceptual dimension of a Mandarin tone perceptual space next to the F0 contour (Gandour, 1983). Additionally, the F0 onset also contributes to the perception of T1 and T2 in addition to the primary F0 contour cues (Chang et al., 2016; Massaro et al., 1985). These findings thus suggest that production-perception links are not likely established through non-primary, non-critical cues, even though these cues may contribute to perception to some degree (Jongman et al., 2000; Newman, 2003; Shultz et al., 2012).

Finally, this study focused on the production-perception relationship of one target tone, T2. However, the critical status of F0 contour cues versus height cues may change in different Mandarin tone contexts (Massaro et al., 1985), which may then lead to different production-perception relationship patterns for different Mandarin tones. Such patterns may further demonstrate how the effects of the critical status of perceptual cues on the production-perception relationship can be aligned with the intrinsic characteristics of individual speech sounds.

Taken together, this study showed that critical tone perceptual F0 contour cues yielded a positive production-perception correlation, whereas non-critical F0 height cues did not, supporting our hypothesis and the previous segmentally-based predictions (Newman, 2003). In addition, the critical status of tonal cues applies in both spectral and temporal domains. These findings extend our understanding of the cues that are pertinent to production-perception links, thus informing the relationship of speech production and perception in general.

Acknowledgments
We thank the attendees of the 176th ASA Meeting and 2018 Acoustics Week of CAA in Victoria, Canada for their valuable feedback, SFU Language and Brain Lab members Anisa Dhanji, Joanna...
Xie, and Dahai Zhang for their assistance. This project was supported by a Discovery Grant from NSERC of Canada (Grant No. 2017-05978).

References and links


