

F0 and voice quality of coarticulated Mandarin tones

Yaqian Huang and Yuan Chai

Tonal coarticulation often induces changes in F0 (Xu, 1997; Brunelle, 2009) as well as voice quality (DiCano, 2012). In Mandarin, a tone language with four distinct pitch contours (Tone 1: high-level, 55; Tone 2: rising, 35; Tone 3: dipping, 21[4]; Tone 4: falling, 51), coarticulation has both carryover and anticipatory effects on F0. For example, in two-tone sequences, an assimilatory effect is found in the second tone, such that that has a lowered F0 when the preceding first tone has a low offset. But a dissimilatory effect is found for the first tone, which can have a raised F0 if the following second tone has a low onset (Xu, 1997; Sun & Shi, 2019). Further, Mandarin has allophonic creaky voice that covaries with F0 (Kuang, 2017). Specifically, creak in Mandarin is associated with low F0 (Kuang, 2017; Chai, 2019) and the low F0 cue is used by listeners for citation tone identification (Huang, 2020). However, few studies have investigated the connection between the changes of F0 and voice quality in tonal coarticulation. Thus, this study asks how F0 raising and lowering due to tonal coarticulation affect voice quality in Mandarin. We address this question using acoustic and electroglottographic (EGG) analyses.

Hypotheses. Given the distinct carryover and anticipatory effects on F0 in various tonal environments and that voice quality depends on F0, we have two predictions of F0 and voice quality changes in the target tone, see Table 1. First, lowering of F0 in the target tone is expected after tones with low offset (T3, T4) and/or before tones with high onset (T1, T4). F0 lowering should be accompanied by an increase in creakiness. Second, raising of F0 in the target tone is expected after tones with high offset (T1, T2) and/or before tones with low onset (T2, T3). Though, here we don't analyze two-T3 sequences due to T3 sandhi (Zhang & Lai, 2010). Accordingly, F0 raising should reduce the amount of creak.

Table 1. Tonal combinations and predictions (X represents the target tone).

Effect	F0 lowering → more creaky	F0 raising → less creaky
Environment	Low offset (T3,4)-X-High onset (T1,4)	High offset (T1,2)-X-Low onset (T2,3)
Sequences	T3-X-T1; T3-X-T4; T4-X-T1; T4-X-T4	T1-X-T2; T1-X-T3; T2-X-T2; T2-X-T3

Methods. The stimuli consisted of trisyllabic Mandarin compounds, where each of the four Mandarin tones was flanked by varying Tones 1–4, for a full range of contextual variation of 4*4*4=64 combinations. The stimuli were embedded in a carrier sentence: *wǒ jiāo nǐ STIMULI zěn me shuō* ‘I teach you STIMULI how to say.’ We recruited 24 native Mandarin speakers (12F). Each participant uttered two repetitions of the 64 sentences in a random order. We obtained F0 and voice quality measures of the middle target syllable of the trisyllabic stimuli using VoiceSauce and EggWorks, and z-scored and time-normalized each measure's values over nine equal intervals. Below, voice quality is discussed only in terms of Contact Quotient (CQ), defined as the fraction of time the glottis is closed. CQ was measured from the EGG signal using a hybrid method (Howard, 1995). Higher values of CQ are expected in creak.

Results. The F0 and voice quality changes in the four tones partly confirm the predictions; representative consistent cases are illustrated in Figure 1. For Tone 1, relative to the grand mean, its F0 was higher in all environments with expected raising, 1-1-2, 1-1-3, 2-1-2, and 2-1-3 ($p < .001$) and lower in environments with expected lowering: 3-1-1, 3-1-4, and 4-1-1 ($p < .001$). But for 4-1-4 sequences, the F0 of T1 was raised ($p < .001$), contrary to our prediction. The CQ of T1 was lower in F0-raised 1-1-2 and 2-1-2 sequences ($p < .001$) and higher in F0-lowered 3-1-1 ($p < .001$) sequences, as expected. For other sequences, the CQ did not differ.

For Tone 2, relative to the grand mean, its F0 was higher in all environments with expected raising: 1-2-2, 1-2-3, 2-2-2, and 2-2-3 ($p < .001$) and lower in all the environments with expected lowering: 3-2-1, 3-2-4, 4-2-4 ($p < .001$), and 4-2-1 ($p < .01$). The CQ of T2 was lower in the F0-raised sequences of 1-2-2 ($p < .025$) and 2-2-2 ($p < .001$) and higher in the F0-lowered sequences of 3-2-4, 4-2-1, and 4-2-4 ($p < .001$), as expected. For other sequences, the CQ did not differ.

For Tone 3, relative to the grand mean after excluding two-T3 combinations, the F0 was lower in the environments with expected lowering: 4-3-1 ($p < .01$) and 4-3-4 ($p < .001$) and higher in one environment with expected raising: 2-3-2 ($p < .001$). But its F0 was lowered in the other environment with expected raising: 1-3-2 ($p < .001$), contrary to our prediction. The CQ of T3 was higher in F0-lowered 4-3-1 and 4-3-4 sequences ($p < .001$) and lower in the F0-raised 2-3-2 sequences ($p < .001$), as expected. The CQ did not differ in 1-3-2 sequences.

For Tone 4, relative to the grand mean, its F0 was higher in the environments with expected raising: 1-4-2, 1-4-3, 2-4-2, and 2-4-3 ($p < .001$) and lower in the environments with expected lowering: 3-4-1, 3-4-4, and 4-4-1 ($p < .001$). The CQ of T4 was lower in the F0-raised sequences of 1-4-2 and 1-4-3 ($p < .001$) and higher in the F0-lowered 4-4-1 sequences ($p < .001$), as expected. However, for the F0-raised 2-4-2 sequences, the CQ was higher ($p < .001$). The F0 and CQ did not differ in other sequences.

Discussion. F0 and voice quality of Mandarin tones can vary according to tonal context. First, we review the F0 results. Most of the target tones had F0 raising or lowering in the expected raising or lowering sequences. Out of the 28 combinations, two exceptions were observed in 4-1-4 and 1-3-2 sequences, for which the F0 changed in an opposite direction contrary to our expectations, and the F0 in 4-4-4 sequences did not change significantly.

Next, we review the voice quality results. Our hypothesis that the voice quality of the target tone will covary with F0 has been partly confirmed. The covarying relation was found in 14 out of the 28 combinations such that when F0 was lowered, the voice quality was more constricted, indicating a creakier quality; when F0 was raised, the tone production became less constricted, indicating a more modal voice quality. There are 13 other sequences in which F0 was raised or lowered but CQ did not change. In addition, one exception was found for the direction of the change in CQ in 2-4-2 sequences. The nonsignificant cases and the exception indicate that voice quality is not completely dependent on F0, but has some independence.

Fig. 1. Representative consistent cases of changes in F0 and CQ. All reached $p < .001$ significance.

