Prosodic organization and microprosodic effects in Shanghai Chinese

Jiayin Gao (jiayin.gao@univ-paris3.fr, Université Paris 3), Pierre Hallé (CNRS-Paris 3/5, Haskins Labs)

Background

In many languages, tonal or non-tonal, F0 is lower after voiced than voiceless consonants. This effect is referred to here as the “F0 perturbation effect” (Hombert et al., 1979). There exist, however, context-specific (Kohler, 1982) and language-specific (Chen, 2011) variations of this effect.

Shanghai Chinese (SHC) has 3 high tones, T1 (52), T2 (34), and T4 (55), and 2 low tones, T3 (23) and T5 (12). T4 and T5 are “checked tones”: short in duration and carried by syllables ending with a glottal stop. The three high tones are associated with voiceless obstruent onsets and the two low tones voiced obstruent onsets. The realization of the voiced vs. voiceless distinction in SHC obstruents is context-conditioned. Word-initially, both voiced and voiceless obstruents are phonetically voiceless in general and the voicing distinction is signaled by a low vs. high tone distinction. In non-initial position, the tonal distinction is neutralized due to a tone sandhi rule, which, to simplify, spreads the tone contour of the word-initial syllable to the next syllables. More exactly, the tone contour of an entire prosodic word is determined by the tone of its first syllable. The underlying voicing distinction of non-initial obstruents is nevertheless maintained: it is not a low vs. high tone distinction, but a phonetic voicing distinction. This context-conditioned realization of voicing is illustrated by the minimal pairs in (1) and (2).

(1) word-initial /t/-/d/: /t/ in [te33 tsz44] 胆子 ‘courage’ vs. /d/ in [te22 tsz33] 台子 ‘table’
(2) non-initial /t/-/d/: /t/ in [pt22 te33] 被单 ‘sheet’ vs. /d/ in [pt22 de33] 皮蛋 ‘hundred-year egg’

The question thus arises of whether F0 perturbations occur for [d] in (2). Previous studies have reported such F0 perturbations (Ren, 1992: 138ff; Chen, 2011). Ren (1992) averaged his data across all prosodic contexts; Chen (2011) found some variation according to prosodic context. Our study further investigates the influence of prosodic context at the lexical level.

Figure 1. Normalized F0 contours (first 100 ms of the rime of S2): (a) post-T1 and (b) post-T2 (averaged across all onsets and speakers for /e/ rime S2s).

Figure 2. Normalized F0 contours for post-T2 S2s (stop onset and /e/ rime): (a) elderly and (b) young speakers.
Experiment

Twenty-two native SHC speakers were recorded: 12 young speakers (6 male, aged 20-30) and 10 elderly speakers (4 male, aged 60-80). They read 40 disyllabic words (S1-2 for syllable 1-2) in a carrier sentence, S2 being the target syllable. Two prosodic contexts, determined by the tone of S1, were used: S1 in tone T1 (post-T1 S2) versus S1 in tone T2 (post-T2 S2). The onset of S2 was a labial or dental stop or fricative; its rime was /e/ or /a/. The averaged S2 duration was ~180 ms for the /e/ rime and ~110 ms for the /a/ rime. F0 was measured with a time step of 5 ms and converted into z-score.

Figure 1 shows S2 F0 contour in the first 100 ms of the rime, averaged across all onsets and speakers for the /e/ rime, according to the original tone of S2, which corresponds to its onset voicing, phonological and phonetic. A similar pattern obtained for /a/. The F0 perturbation effect can be observed in the post-T2 context (after voiceless vs. voiced: 233 vs. 216 Hz, Figure 1b), but not in the post-T1 context (213 vs. 216 Hz, Figure 1a), contrary to previous reports. Figure 2 shows that, in the post-T2 context for the /e/ rime and for stops, the F0 perturbation effect lasts longer for young than elderly speakers (about 100 ms vs. 70 ms). Chen (2011) reported that the F0 perturbation effect only affects the first 50 ms of the syllable. Her speakers were aged 60-75, an age range comparable to that of our elderly speakers. It thus seems that, in Chen’s (2011) data and in our elderly speakers’ data, the F0 perturbation effect is more locally restricted than in our young speakers’ data, whereby the time domain of the effect is wider.

Discussion

Our study mainly shows that the F0 perturbation effect is not observed in the post-T1 context. This does not imply the absence of this effect. We speculate that the prosodic organization at the word level, which determines the ballistic F0 movement on the entire word, contribute to the F0 onset height in S2. We propose that the timing of the F0 contour on prosodic words is planned independently from the timing of the segments. As shown in Figure 3, the F0 contour falls monotonically at a constant velocity over the prosodic word in the post-T1 context, regardless of segment durations. When the vowel onset of S2 occurs earlier, the F0 onset in S2 is higher on the whole prosodic word ballistic F0 trajectory. This is what happens when the word-medial consonant is voiced because it is shorter in duration than its voiceless counterpart (Shen et al., 1987): Gao (2015: 169) reported an average duration of 65 ms for intervocalic voiced obstruents, compared to 124 ms for intervocalic voiceless obstruents. This “ballistic” effect is thus opposite to the presumed F0 lowering effect (“depressor” effect) due to a voiced obstruent onset. The F0 perturbation effect in the post-T1 context is weak and runs (significantly) in the raising rather than lowering direction. In the post-T2 context, the overall F0 contour is rising instead of falling: here the ballistic effect does not cancel out the depressor effect and the overall F0 perturbation effect at rime onset clearly is an F0 lowering effect.

References