Shanghai Chinese obstruent durations vary with voicing: A phonological or phonetic effect?

Pierre Hallé (pierre.halle@univ-paris3.fr, CNRS-Paris 3/5, Haskins labs), Jiayin Gao (Paris 3 University)

Phonological voicing in Shanghai Chinese

The interdependence of tone register and phonological voicing in Shanghai Chinese (henceforth SHC) has its roots in the “tone split” diachronic change that affected Chinese dialects. In this process, the phonological voicing distinction in obstruents was transphonologized into a tone height distinction ([+voice] > low tones; [-voice] > high tones) and disappeared altogether in most dialects with the exception of the Wu dialects, to which SHC belongs. SHC underwent tone split but also maintained the voicing distinction in a particular way: In word-initial, stressed syllable onsets, the distinction is marked by tone height (voiced–voiceless ⇔ low–high); in non-initial, unstressed syllable onsets, it is marked by phonetic voicing (i.e., with vs. without glottal pulsing). In the latter case, non-initial syllable’s tone is neutralized due to a tone sandhi rule, whereby the F0 contour of a word is determined solely by the tone of its first syllable. Note that the voiceless aspirated stops of SHC are out of the scope of the present study.

The correlation between voicing and duration in obstruents

In most languages of the world, a cluster of redundant features accompanies the voicing distinction in obstruents. One of them is segmental duration: voiced obstruents are shorter than their voiceless counterparts. Yet, is the motivation for this difference phonological or phonetic in nature? This question remains largely unanswered because, in the studies relevant to that issue, phonetic and phonological voicing are usually undistinguishable. Due to its particular implementation of phonological voicing, SHC provides an opportunity to test the nature of the duration effect. Were it purely phonetic in nature, SHC word-initial obstruents should not differ in duration according to phonological voicing since they usually are phonetically voiceless in this position. On the opposite view, if duration is part of a phonological representation of voiced vs. voiceless obstruents, duration should vary according to phonological, underlying voicing. Moreover, duration should bias listeners’ categorization of word-initial obstruents into the voiced (low tone) vs. voiceless (high tone) categories.

Production study: durations of obstruent onsets according to underlying voicing

We first report the durations of SHC obstruent onsets in either word-initial or word-medial position in function of underlying voicing. In word-initial position, only fricatives were examined since phonoetically voiceless stops leave no acoustic trace of their duration. In both positions, phonologically voiced fricatives were significantly shorter than their voiceless counterparts. This also held for stops in word-medial position, where their duration could be measured regardless of phonetic voicing. The following rime exhibited an opposite duration effect, which was marginally significant: it tended to be shorter for underlyingly voiceless than voiced onsets (/e/: 208 ms vs. 235 ms after voiceless vs. voiced onsets; /a/ : 102 vs. 110 ms). The obstruent onset duration data are shown in Table 1.

A possible confound in word-initial fricatives is that phonologically voiced fricatives also are fully or partly voiced phonetically in this position, especially labial fricatives. However, when focusing on the phonetically voiceless fricatives in the data, virtually identical differences still hold (corresponding data shown within parentheses in Table 1). These findings support a phonological rather than phonetic motivation for the observed correlation between segmental voicing and duration.

Table 1. Obstruent onset duration (in ms) according to underlying tone in three production contexts: single syllable, first syllable (S1), and second syllable (S2) of a disyllabic word. The durations within parentheses are those of word-initial phonetically voiceless fricatives.

<table>
<thead>
<tr>
<th>syllable type</th>
<th>syllable’s tone</th>
<th>single syllable</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>unchecked</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(/e/ rime)</td>
<td>T3 (low, voiced)</td>
<td>130 (128)</td>
<td>111 (120)</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>T1-2 (high, voiceless)</td>
<td>196</td>
<td>165</td>
<td>119</td>
</tr>
<tr>
<td>checked</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(/a/ rime)</td>
<td>T5 (low, voiced)</td>
<td>162 (167)</td>
<td>101 (107)</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>T4 (high, voiceless)</td>
<td>208</td>
<td>163</td>
<td>133</td>
</tr>
</tbody>
</table>
Perception study: duration patterns bias listeners’ categorization of onsets as voiced vs. voiceless

We asked whether segment durations would bias listeners’ judgment of fricative voicing category in word initial position. Recall that, in this position, phonological voicing and tone register are in an equivalence relationship: voiced ⇔ low; voiceless ⇔ high. Categorizing word-initial fricative as phonologically voiceless vs. voiced thus is equivalent to categorizing syllable’s tone as high vs. low (e.g., T2 vs. T3). We recorded minimal pair syllables such as /fe/ [fɛ34] (tone T2) vs. /ve/ [vɛ33] (tone T3), with the onsets /f, v, s, z/ and the rime /ɛ/. Importantly, phonetic voicing in /ve/ or /ɛz/ items was intentionally avoided: Phonetically voiced onset stimuli—even partly voiced—were not retained. From the naturally produced syllables that were retained, we constructed syllables with time-scaled fricative onset and vowel rime so that, in one set of constructed stimuli, the C/V duration ratio was high (thus typical of a voiceless onset syllable), and, in the other set, the C/V duration ratio was low (thus typical of a voiced onset syllable). We imposed T2-T3 tone contour continua (see Figure 1) on the stimuli originally in tone T2 and, symmetrically, T3-T2 continua on the stimuli originally in tone T3. The resulting stimulus continua were presented to 23 SHC native listeners for forced-choice tone identification as T2 or T3. Our prediction was that high C/V duration ratio would favor T2 responses (equivalent to phonologically voiceless) and, conversely, low C/V duration ratio would favor T3 responses (equivalent to phonologically voiced).

This is indeed what we found. The categorization functions, pooled across subjects and onsets, are shown in Figure 2: the high, compared to the low C/V ratio, significantly shifted the responses toward T2 (by about 20% in the ambiguous region of the continuum). The response time (RT) data were in line with the categorization shift induced by the C/V ratio manipulation: the high C/V ratio yielded faster T2 and slower T3 responses, and vice versa for the low C/V ratio.

Conclusion

Our data suggest that the duration difference between word-initial voiced and voiceless obstruents is part of their phonological representations—which therefore must consist in a complex cluster of features—rather than a simple reflex of their phonetic realizations. These representations are active in both production and perception.

Selected references