Further results on the perception of place coarticulation in Taiwanese stops

Shu-hui Peng and Terrance M. Nearey

Department of Linguistics, University of Alberta, Edmonton, Alberta T6G 2E7, Canada

Abstract: Previous study of place coarticulation using logistic discrimination showed that 87% of Taiwanese word-final dental or velar stops in stop consonant clusters (CVstop#stop2V(C)) spoken in various speech rates were correctly classified based on formant transitions of the preceding vowel. This automatic classification was significantly (though only moderately) correlated with categorization by Taiwanese speakers in a concept formation test. The present study examined the perception of the stops by a phonetically-trained listener in a goodness rating paradigm. The phonetic judgments of the listener was compared with the categorization by Taiwanese speakers, acoustic analysis, and EPG measurements.

BACKGROUND

Previous EPG (electropalatography) study (1) investigating place coarticulation in Taiwanese dental-velar or velar-dental stop sequences separated by word boundaries (/pat#kag/, /pa#khag/, /lak#yang/, /lak#hag/) showed that the timing of the two place gestures was affected by speech rate. The latency of the second gesture relative to the first gesture decreased as speech rate increased. A perception study using the concept formation paradigm also showed that as speech rate increased, the first stop of the sequence was often perceived with the characteristics of the second stop. In fast speech rate, the stop in CVt syllables was more frequently perceived as /k/ than as /l/. However, regression analysis showed that this categorization pattern often could not be consistently predicted from EPG measurements.

Peng and Nearey (2) examined the extent to which perceptual categorization of place coarticulation correlates with acoustic properties of the stop sequences. The categorization of word-final stops by Taiwanese listeners was compared with discriminant function analysis based on spectral measurements in the stops and the preceding vowel. However, the correlations between the predicted categorization and the observed categorization were modest ($R^2$ less than 0.167).

The present study examines the perception of the place coarticulation by a phonetically-trained listener with a category goodness rating paradigm. A phonetician was asked to rate the category goodness of the first stop of the stop sequences. The results of the rating were compared with the categorization by Taiwanese listeners in the previous concept formation test. Both the rating by the phonetician and the categorization by Taiwanese listeners were then compared directly with regression to acoustic measurements and gestural (EPG) measurements.

METHOD

The test stimuli used in the rating test were the excised first syllables of six two-syllable compounds: /pat#ka~/, /pat#ka~/, /la#kyang/, /la#khag/, /pa#tikag/, /pa#khag/. Each compound was produced 11 times ranging from slow rate to fast rate. Each of the repetition of the syllable was repeated ten times in the rating test. The phonetician was asked to rate the final stop of the syllable in terms of goodness of /d/ or /l/ on a 0 to 1 scale in two separate task sessions.

The acoustic analyses were based on the formant patterns of the vowel in the first syllable. Formant frequencies were estimated using a procedure similar to that described by Talkin (3). Formant candidates were analyzed from short-term analysis of 100 ms frames multiplied by cosine$^4$ windows with a frame advance of 3.0 ms by using a12-coefficient selective LPC procedure with a high frequency cutoff of 10 kHz. Candidates were aligned to formant slots using a dynamic programming algorithm followed by 3-point median smoothing of each formant frequency track. The signals so tracked were displayed with spectrograms. Temporal regions corresponding to the beginning and end of the vocoid of the first syllable were selected and measurements were taken at the midpoint and end of this track. (Some additional acoustic measurements were also taken but cannot be reported in this summary).

The EPG measurements were taken from a previous study by Peng (4). The measurements selected were the duration of the dental or velar gestures, degree of lingual contact, the latency of the second gesture relative to the first gesture at the beginning, maximum lingual contact, and end of the the gestures, and the degree of overlapping between the two gestures.
RESULTS & DISCUSSION

The perception of place coarticulation by the phonetically-trained listener showed a similar categorization pattern as Taiwanese listeners' perception. The correlation between the categorization by Taiwanese listeners and the rating by the phonetically-trained listener was significant for both the /d/ task \((r = .789, p < .005)\) and the /k/ task \((r = .545, p < .005)\). The results of the comparison between the perception of the syllable-final stops by listeners and the acoustic and articulatory measurements are summarized in Table 1. (The cross-validation forward stepwise model selection procedure described by Hjorth (5) was used.) The categorization by Taiwanese listeners was correlated to formant measurements better than the EPG measurements. Significant correlations with EPG measurements were only found in the /t/ categorization for /tt/ (degree of lingual contact) and the /k/ categorization for /tk/ (offset latency of the velar gesture relative to the dental gesture). These results seem to suggest that some articulatory information related to the perception of the syllable-final stops was not included in the EPG measurements used in this study.

The rating by the phonetically-trained listener, in general, was also better predicted by formant measurements for /kk/ and /tt/ sequences. For /kt/ and /tk/ sequences, the syllable-final stops were better predicted by EPG measurements (duration of the velar gesture and the latency of the second gesture relative to the first gesture). Furthermore, the prediction by formant measurements for /k/ was not as good as that for /t/. This suggests that the phonetically-trained listener might have detected some other acoustic information than formant frequencies which contributed to the perception of /k/. For example, some weak frication at the transition between the vowel and the velar stop closure was often observed in tokens spoken at faster speech rate. However, because the small number of data points and the large number of correlation examined, it is not possible to draw any firm conclusion from this data. In the future study, we plan to directly correlate acoustic parameters with parameters of articulatory gestures using the productions of five other subjects.

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>Taiwanese listeners</th>
<th>phonomatic listener</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/t/ judgment</td>
<td>/k/ judgment</td>
</tr>
<tr>
<td>CC  EPG formants</td>
<td>EPG formants</td>
<td>EPG formants</td>
</tr>
<tr>
<td>kk</td>
<td>0.110</td>
<td>0.078</td>
</tr>
<tr>
<td>tt</td>
<td>0.832</td>
<td>0.368</td>
</tr>
<tr>
<td>kt</td>
<td>0.085</td>
<td>0.325</td>
</tr>
<tr>
<td>tk</td>
<td>0.343</td>
<td>0.591</td>
</tr>
</tbody>
</table>

REFERENCES