Perception of Synthesised Hindi Geminate and Cluster Sounds

Nisheeth Shrotriya and S.S. Agrawal

Speech Technology Group
Central Electronics and Engineering Research Institute Centre,
CSIR Complex, NPL Campus, Hillside Road, New Delhi-110 012, INDIA
E-mail: sagrawal@ceerid.ernet.in

Abstract: This paper describes the differences in some acoustic parameters of geminate, non-geminate and cluster sounds of Hindi, which can be used for their perceptual distinction. For distinction between geminate and non-geminate sounds, although the parameters such as duration of the plosive gap and preceding vowel, and the formant frequencies and intensity of the burst are important but the closure duration plays the most significant role. This difference is of the order of 1:1.5. In case of the cluster (-C1C2-) words containing stop consonants, when C1 and C2 belong to different places of articulation, a burst is observed in the middle of the closure. When the closure on either side of this burst is reduced by different amounts the sounds are perceived differently.

INTRODUCTION

Speech perception is one of the most important methods for evaluating the role acoustic parameters for synthesizing good quality speech. A number of perception experiments have been performed to find their importance, including closure duration, for the identification of consonants [1, 2, 3, 4]. The present study is based on the perception experiments of the synthetic stimuli, after manipulating the closure durations of the singleton, geminate and cluster consonants.

PROCEDURE

In the present experiment all the non-geminate and geminate words were synthesized using the PC based KLATT synthesizer [5]. The synthetic non-geminate words were segmented in to two parts and the segmented files were patched by adding, silence (in case of the unvoiced stops) and voice bar (in case of the voiced stops) in their closure. However, in case of the geminate words instead of adding, the desired amount of silence and or voice bar was removed from the closure, in steps of 10 msec. The limit of removing the silence was set to 120 msec (12 steps). For perception tests ten listeners, having Hindi as their native language, were chosen. They heard the stimuli through a TEAC cassette deck player and were asked to write down the word they heard. However, in case of the cluster words only those words were synthesized for which the two stop consonants (-C1C2-) differed in their place of articulation. In such cases an additional burst/spike of duration less than 5 ms, occurs in the middle region of the closure. To find out the effect of this burst on perception the closure duration on either side of the burst was reduced in steps and perception experiments were conducted.

RESULTS

1. The subjects' responses were scored according to whether they heard a non-geminate or a geminate consonant. The mean percentage of geminate responses at different closure durations to the synthetic stimuli is shown in Fig 1. The (+) sign indicates geminate responses to stimuli created from original non-geminates by incrementing their closure; whereas the (.) sign indicates geminate responses to stimuli created from original geminates by decrementing their closure duration. The average displacement (as could be seen through fig. 1) between the curves is about 9 msec. However, for individual cases the displacement varies from a minimum of 9 msec to a maximum of 30 msec. The average displacement values (in msec) for different places of articulations are given below:

Labials > Dentals > Palatals > Velars > Affricates
18.625 > 18.125 > 17.625 > 14.125 > 13.250

2931
Though the amount of the silence required to be added to convert a non-geminate sound into a geminate sound depends upon, whether the stop is voiced-unvoiced or aspirated-unaspirated, but on an average if an additional silence/voice bar of duration 135 msec is added to the closure of the stop it is converted into a geminate. Similarly, if the same amount is removed from the closure of a geminate it becomes a non-geminate. This is an interesting observation.

**TABLE 1.** Effect of removal of various amount of closure durations on the perception of the cluster word.

<table>
<thead>
<tr>
<th>With Burst</th>
<th>With out Burst</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closure reduced by (msec)</td>
<td>Word Perceived</td>
</tr>
<tr>
<td>upto 60</td>
<td>Cluster</td>
</tr>
<tr>
<td>&gt; 60 Non cluster</td>
<td>upto 30 Geminate</td>
</tr>
</tbody>
</table>

**FIGURE 1.** The mean percentage of geminate responses at different closure durations.

2. In case of the cluster words for which the two stop consonants differ in their place of articulation, perception experiments indicate that when the closure duration on either side of the burst is reduced up to a particular value the sound remains same, however beyond that value the sound changes. For example, in case of the cluster word /s/-/kbar/, the total closure duration is about 180 msec and the burst occurs in the middle. When the closure duration on either side of this burst is reduced up to 60 msec the sound remains same but beyond this value the cluster word is perceived as a non-cluster word i.e. /s/-/kbar/. The same experiment was conducted with the natural speech which gave similar results. Further, in case of the natural speech, even if the burst is removed from the closure, the perception experiments indicate that the sound remains same. But when the closure duration (in addition to the removal of burst) is reduced by about 20-30 msec the sound changes to a geminate word (/s/-/kbar/); further reduction in the closure duration changes the sound to a non-geminate word i.e. (/s/-/kar/). However, in case of the synthetic speech, the cluster word (/s/-/kbar/) after removing the burst from the closure was perceived as the geminate word (/s/-/kbar/). These values are shown in Table1. Though the present results are based on the study of a small number of data, however, studies with a large data base are continued.

**ACKNOWLEDGMENTS**

The authors are thankful to the Director CEERI Pilani for permitting us to publish this work. They are also thankful to Shri A S S Sarma and S. Nigam for their discussions and useful comments during the preparation of the paper. The award of Senior Research Fellowship by Council of Scientific and Industrial Research of India to one of the authors (Nisheeth Shrotriya) is gratefully acknowledged.

**REFERENCES**