PERCEPTION OF CUES TO A STOP VOicing CONTRAST BY
NORMAL-HEARING CHILDREN AND ADULTS

VALERIE HAZAN  LISA HOLDEN-PITT  SALLY REVOLLE  DONNA EDWARD
Dept. of Phonetics and Linguistics
Univ. College London
London NW1 2HE, UK

Sensory Communication Research Lab.
Galldet University
Washington, DC 20002, USA

ABSTRACT

The contribution of two acoustic cues, Voice Onset Time (VOT) and vowel onset transitions, to the perception of a /taed/-/daad/ contrast was examined for normal-hearing children and adult controls.

INTRODUCTION

The important speech acoustic cues affecting voicing identification of initial stops are voice onset time (VOT) and the vowel onset transitions. Differences in the stimuli used among these studies may also be a factor in the variations found among these results.

This paper describes an experiment that examined further the use of VOT versus vowel onset transitions for onset stop voicing distinction by normal-hearing children and adults. Both synthetic and spoken stimuli were tested.

METHOD

Stimuli

Three continua of spoken /dadj/-/dad/ stimuli and two of synthetic /dadj/-/dad/ were used as the test syllables. Each continuum comprised eight stimuli among which VOT varied nominally from 18 to 60 ms in 6 ms steps.

Vowel onset transitions were also present in three of the five continua's stimuli. The three continua of natural stimuli were derived from two spoken utterances - a /taed/ and a /daad/ - that had been selected for their average acoustic characteristics from a larger pool of syllables [6].

In one continuum, TAD/VOT, a /taed/ utterance served as a base stimulus; the /taed/ burst was appropriately shortened to yield the desired VOT durations for the constituent stimuli of the continuum. The resultant /t/ burst were copied for use in another condition, DAD/VOT. The base stimulus of this continuum was a /daad/ stem from which the /d/ burst had been removed and replaced by the /t/ bursts of different durations. These same stimuli were used in a third continuum of synthetic syllables, DAD/VOT/vowel cutback, but here, the /t/ was progressively cut back within most of the continuum to approximate the VOT/vowel transition relationship found in natural speech.

The two remaining continua contained stimuli developed via software synthesizer [7]. These were used in a fourth continuum, the synthetic version of the natural continuum contained an initial burst, with major energy peaks at 1620 Hz, 2600 Hz and 4000 Hz. Vowel formant values were selected for a steady state state but constituted a best fit to the natural vowel. F1 varied from 500 Hz to 288 Hz, F2 from 1650 Hz to 1355 Hz, F3 from 2650 Hz to 2433 Hz and F4 from 3700 Hz to 3500 Hz. In both continua, the respective VOTs approximated those used on the natural continuum. In the "Synthetic TAD/VOT" continuum, F1 was fixed at 500 Hz throughout the stimulus range and contained no initial burst, whereas the F1 on the vowel cutback was therefore neutralized. In the "Synthetic DAD/VOT/F1 cutback" continuum, the F1 transition was systematically varied in front of initial burst throughout the continuum, with a starting frequency of 460 Hz at the voiced extreme of the stimulus range. The transition duration of F1 was 36 ms.

Subjests

Ten normal-hearing children 7-9 years of age and five normal-hearing adults served as paid listeners. The younger subjects were children of employees at Gallaudet University. All children and adults had pure tone thresholds (3FA mean of +5.1 and 2 kHz) better than 15 dB HL.

RESULTS AND DISCUSSION

Put stimuli were presented in single-interval identification trials with "TAD," "DAD" response alternatives. Pictorial sketches and orthographic labels of TAD and DAD were displayed. A touch-sensitive screen was used as a response terminal.

The order of stimulus presentation followed a simple adaptive tracking procedure developed for use in perceptual experiments with young children [8]. The approximate length of the test for the children was 48 trials.

The children were tested during five 30-minute sessions that occurred throughout the three-week period. The tests were administered to the adults in two sessions of about one hour each. The listeners presented 15 db SPL. Stimulus presentation and response tallies were under computer control (DEC PDP-11/23 and 11/34).

Procedure

The stimuli were presented in single-interval identification trials with TAD, DAD response alternatives. Pictorial sketches and orthographic labels of TAD and DAD were displayed. A touch-sensitive screen was used as a response terminal.

The statistical similarity between the groups for the synthetic continuum was significant. The statistical similarity between the groups for the natural continuum was significant. The statistical similarity between the groups for the synthetic continuum was significant. To determine which conditions contributed to this effect, Tyree's test of honestly significant differences (had) was applied. For the natural stimuli, the phoneme boundary of the TAD/VOT continuum was significantly shorter than that for other conditions of natural and synthetic stimuli [9]. The phoneme boundary obtained for the synthetic version of TAD/VOT was also shorter than that found for the synthetic continuum with F1 onset transitions. Overall, however, this result fell just short of statistical significance. These findings indicate that the contribution of voicing cues to the voicing contrast in the vowel onset required a shorter VOT than stimuli which contained these cues. Others have reported similar results with normal-hearing [2] and some hearing-impaired adults [2]. Overall, however, VOT appears to prevail over cues in the vowel onset for effects on initial stop voicing perception. Indeed, categorization of the stimulus was achieved despite conflicting spectral cues in the TAD/VOT and DAD/VOT continua.

A significant difference in identification function gradients was obtained between listener groups (i.e. adult versus child hearing status). The children obtained similar /d/-/t/ phoneme boundaries. This outcome would suggest that the general age difference between the two listener group was insignificant with respect to their use of the VOT and vowel onset cues for locating /d/-/t/ phoneme boundaries.

REFERENCES

effect is not found for the children's group where very little difference is observed with respect to the function gradient among the five conditions. When the two listener groups are examined separately for condition effects, neither show a significant difference in function gradient among conditions. However, for the adult group, a greater distinction in function slope is observed among conditions [(F(4,16)=2.37,p=.01)] than is seen for the children's group [F(4,36)=4.69,p=.01]. Large standard error measures were obtained for identification function gradients for the natural and synthetic TAD/VOT and natural DAD/VOT showing greater inter-individual variability in conditions with conflicting spectral cues.

CONCLUSION

Results confirm the primary importance of the temporal VOT cue over the spectral vowel onset cue to the voicing contrast in initial plosives. Vowel onset characteristics were, however, shown to have a clear secondary effect, as shown by a shift in boundary, when the cue is absent, in both children and adults. Although children gave very similar labeling to edited natural stimuli than adults, they seemed less affected by a removal of vowel onset cues.

High quality synthetic speech did provide a good match to results obtained with natural edited stimuli, for both adults and children. However, greater inter-individual variations in labeling were found both for adults and children. As a result, the shift in boundary between the TAD/VOT and DAD/VOT/cutback conditions, which had been strongly significant in the natural edited stimuli was found to be of statistical significance using synthetic stimuli. This would suggest that edited natural stimuli, by providing more homogeneous results, may be more reliable than synthetic stimuli in cue weighting experiments. However, there are limitations in the types of cues which may be altered through computer processing of natural speech, so that synthetic speech does still provide the greatest flexibility when constructing stimulus continua in which spectral rather than temporal patterns are varied.

REFERENCES


<table>
<thead>
<tr>
<th>TEST CONDITIONS</th>
<th>Natural stimuli</th>
<th>Synthetic Stimmuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAD/VOT</td>
<td>DAD/VOT</td>
<td>TAD/VOT</td>
</tr>
<tr>
<td>Vowel Cutback</td>
<td>DAD/VOT</td>
<td>Vowel Cutback</td>
</tr>
<tr>
<td>Boundary</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Children</th>
<th>Adults</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table I: Mean phoneme boundary values (in milliseconds)

| ID Function |
|            |
| Gradient   |
| (Slope)    |
| Children   |
| Adults     |
| Total      |
|            |
|            |

Table II: Mean identification function gradients