PERCEPTION OF CUES TO A STOP VOICING CONTRAST BY

NORMAL-HEARING CHILDREN AND ADULTS

VALERIE HAZAN LISA HOLDEN-PITT SALLY REVOILE DONNA EDWARD

Dept. of Phonetics and Linguistics
Univ. College London
London NW1 2HE, UK

Sensory Communication Research Lab.
Gallaudet 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 /tæd/-/dæd/ contrast was examined for normal-hearing children and adult controls.

INTRODUCTION

The important speech acoustic cues effecting voicing identification of initial stops are voice onset time (VOT) and the vowel onset transitions subsequent to the stop burst [1]. The contribution of these cues for initial consonant voicing perception by adults and children, both normal-hearing and hearing-impaired, has been investigated in various studies [2,3,4,5,6]. However, because these cues have been co-varied in most studies, the relative importance of VOT versus vowel onset transitions for initial stop voicing distinctions remains equivocal. 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 cueing initial stop voicing distinctions by normal-hearing children and adults. Both synthetic and spoken stimuli were tested.

METHOD

Stimuli

Three continua of spoken /dæd/-/tæd/ stimuli and two of synthetic /dæd/-/tæd/ 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 /tæd/ and a /dæd/ -that had been selected for their average acoustic characteristics from a larger pool of syllables [6].

In one continuum, TAD/VOT, a /tæd/ utterance served as a base stimulus; the /t/ 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 /dæd/ 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 spoken syllables, DAD/VOT/vowel cutback, but here, the vowel was progressively cut back within most of the continuum to approximate the VOT/transition cue relationship found in natural speech.

The two remaining continua contained stimuli developed via software synthesizer [7]. These were copy syntheses generated to resemble perceptually and acoustically the utterances of /tæd/ and /dæd/ from which the natural continua were developed. The synthetic stimuli contained an initial burst, with major energy peaks at 1620 Hz, 2600 Hz and 4000 Hz. Vowel formant values were not steady state but constituted a best fit to the natural vowel. F1 varied from 500 Hz to 288 Hz, F2 from 1850 Hz to 1535 Hz, F3 from 2650 Hz to 2433 Hz and F4 from 3700 Hz to 3450 Hz. In both continua, the respective VOTs approximated those used in the natural continua. In the "Synthetic TAD/VOT" continuum, F1 started at 500 Hz throughout the stimulus range and contained no initial transition; the F1 onset cue was therefore neutralized. In the "Synthetic DAD/VOT/F1 cutback" continuum, the F1 transition was systematically varied in frequency extent throughout the continuum, with a starting frequency of 400 Hz at the voiced extreme of the stimulus range. The transition duration of F1 was 36 ms.

Subjects

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.

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 on a touch-sensitive screen 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 average length of test for the children was of 48 trials.

The children were tested during five 30-minute sessions that occurred within a three-week period. The tests were administered to the adults in two sessions of about one hour each. The listeners were tested in IAC audiometric rooms, with the stimuli presented monaurally through a TDH-39 earphone (MX 41/AR cushion) in a headset. Stimuli were presented at 75 dB SPL. Stimulus presentation and response tallies were under computer control (DEC PDP-11/23 and 11/34).

RESULTS AND DISCUSSION

For each listener, mean results were computed over the four repetitions of each continuum (three repetitions for the DAD/VOT continuum). A Maximum Likelihood technique [9] was used to fit a cumulative normal function to each set of data. Two measures of performance were derived: the phoneme boundary (50% labeling point of the fitted curve) and the gradient of the identification function (slope). The results are summarised in Table I for the two groups of listeners individually and combined. For each performance measure, analyses of variance were carried out using factors: groups (adults versus children) by conditions (the five stimulus continua) treated as repeated measures.

Phoneme Boundaries

The listener groups did not differ for the phoneme boundaries measured. The group of adults and of children obtained similar /d/-/t/ phoneme boundaries [$F(1,13)=.02$, $p=0.9$], and showed no interaction with the test conditions [$F(4,52)=.9$, $p=0.5$]. This outcome would suggest that the general age difference between the two listener groups was insignificant with respect to their use of the VOT and vowel onset cues for locating /d/-/t/ phoneme boundaries.
The statistical similarity between the groups for phoneme boundary enables their results to be combined for testing differences among conditions. An additional analysis of variance carried out for the total group of listeners revealed that the effect of test condition was significant [$F(4,56)=7.8$, $p=0.0$]. To determine which conditions contributed to this effect, Tukey's test of honestly significant differences (hsd) 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 [hsd, $p<.05$]. The phoneme boundary obtained for the synthetic version of TAD/VOT was also shorter than that found for the synthetic continuum with F1 onset cues, DAD/VOT/F1 cutback. However, this result fell just short of statistical significance. These findings indicate that the continua lacking spectral cues to the voicing contrast in the vowel onset required a shorter VOT to be perceived as /dæd/ than stimuli which contained these cues. Others have reported similar results with normal-hearing [2] and some hearing-impaired children [5]. Overall, however, VOT appears to prevail over cues in the vowel onset for effect on initial stop voicing perception. Indeed, categorisation of the stimuli was achieved despite conflicting spectral cues in the TAD/VOT and DAD/VOT conditions. In a recent study, Revoile et al. [6] found that the insertion of aspiration between voiced stop transients and subsequent vowels in spoken stimuli yielded a near complete reversal in perception from voiced to voiceless.

Note that between the natural and synthetic stimuli, conditions with analogous cues (i.e. TAD/VOT, and also DAD/VOT/cutback) yielded similar results. However, Table I reveals that the synthetic stimuli effected greater standard deviations for phoneme boundary means than those found for the natural stimuli. Also, among the continua for natural stimuli, one of the condition with conflicting voicing cues (F1 transition in the presence of long VOTs), DAD/VOT, produced larger standard deviations than found for the other two conditions. We may speculate that the greater variability in results for the synthetic condition and natural with conflicting cues was due to a more artificial quality inherent to these stimuli. Statistically, however, the spoken stimuli and synthetic stimuli generated to resemble the spoken stimuli produced similar perceptual effects for distinction of the voicing contrast in initial alveolar stops, at least for these normal-hearing listeners.

Identification function gradients

A significant difference in identification function gradient was obtained between the two listener groups [$F(1,13)=15.32$, $p=.002$]. Table II reveals that the gradients for the adult group are steeper than those for the children in each condition. This outcome suggests that the children were more tentative in their phoneme distinctions of /d/ vs /t/ than were the adults. This effect is also reported by Simon and Fourcin [2] who found that age-related development in the ability to label voicing contrasts was mirrored by an increase in identification function gradient. An interaction was found between listener group and test condition [$F(4,52)=3.22$, $p=0.2$]. Examination of the means shows that this tendency is largely due to the considerably steeper slope observed for the adults in the natural TAD/VOT condition. This 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 [$F(4,16)=2.37, p=.1$] than is seen for the children's group [$F(4,36)=.69, p=.6$]. 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 short 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

[1] Lisker,L. & Abramson,A.S. A cross-language study of voicing in initial stops: acoustic measurements. Word, 20, 384-422, 1964

[2] Simon,C. & Fourcin,A.J. Cross-language study of speech pattern learning. J.Acoust.Soc.Am., 63, 925-935, 1978

[3] Parady,S., Dorman,M., Whaley,P. & Raphael,E. Identification and discrimination of a synthesised voicing contrast by normal and sensorineural hearing-impaired children. J.Acoust.Soc.Am., 69, 783-789, 1981

[4] Johnson,D., Whaley,P. & Dorman,M.F. Processing of cues for stop consonant voicing by young hearing-impaired listeners. J.Speech Hear.Res., 27, 112-118, 1984

[5] Hazan,V. & Fourcin, A. Microprocessor - controlled speech pattern audiometry: preliminary results. Audiology, 24, 325-335, 1985

[6] Revoile S., Pickett, J.M., Holden-Pitt,L.D., Talkin,D., Brandt,F.D. Burst and transition cues to voicing perception for spoken initial stops by impaired and normal-hearing listeners. J.Speech Hear.Res, 30, 3-12, 1987

[7] Klatt,D.H. Software for a cascade/parallel formant synthesizer. J.Acoust.Soc.Am., 67, 971-995, 1980

[8] Hazan,V. Speech pattern audiometric assessment of hearing-impaired children. Doctoral Dissertation. University College London, 1986

[9] Bock,K.D. & Jones,L.V. "The Measurement and Prediction of Judgment and Choice". Holden-Day, San Francisco, 1968

	TEST CONDITIONS				
	Natural stimuli			Synthetic Stimuli	
Phoneme Boundary	TAD/VOT	DAD/VOT/ Vowel Cutback	DAD/VOT	TAD/VOT	DAD/VOT/ F1 Cutback
Children n=10	28.7 (2.2)	32.4 (1.3)	32.3 (3.2)	29.6 (5.2)	31.9 (3.4)
Adults n=5	26.1 (1.7)	31.5 (3.1)	32.7 (3.1)	30.1 (4.5)	33.3 (5.6)
Total n=15	27.8 (2.3)	32.1 (2.0)	32.4 (3.1)	29.8 (4.9)	32.3 (4.1)

Table I : Mean phoneme boundary values (in milliseconds)

ID Function Gradient (Slope)					
Children n=10	-.92 (.52)	-.82 (.23)	-1.03 (.50)	-.88 (.36)	-1.04 (.60)
Adults n=5	-2.34 (.74)	-1.52 (.33)	-1.53 (.68)	-1.28 (.53)	-1.31 (.74)
Total N=15	-1.40 (.90)	-1.05 (.43)	-1.20 (.59)	-1.01 (.45)	-1.13 (.63)

Table II : Mean identification function gradients