

Perceptual and Acoustic Analysis of Word Initial Voicing Contrasts Across Speaker Age

Kelly W. Lucky, Megan M. Hodge and Anne H. B. Putnam Rochet

Department of Speech Pathology and Audiology
University of Alberta, Edmonton, Alberta T6G 2G4

1. Introduction

There are several acoustic cues, including voice onset time (VOT), first formant frequency (F1), fundamental frequency (F0) and burst amplitude (BA) that influence how a listener perceives a plosive with respect to the feature of voicing. For a given speaker, these cues may be complementary, providing the listener with redundant information to identify the contrast accurately or they may be contradictory, leading to perceptual uncertainty. While previous investigators have studied the development of VOT in children, limited information is available regarding children's production of these multiple cues and how these influence adult listeners' perceptions of children's voicing contrasts [1]. This study explored the relationship among four acoustic characteristics of word-initial voiced and voiceless alveolar plosives produced by four age groups of normal speakers of Western Canadian English and was designed to answer the following questions:

1. Are there differences among speaker age groups in the frequency of correct and incorrect or ambiguous identifications of voiced and voiceless plosives by adult listeners?
2. Are there age group differences in VOT, F1, F0 and BA for correctly identified voiced and voiceless plosives?
3. What combination of these four acoustic cues best predicts listeners' identification of word-initial voiced and voiceless alveolar plosives?

2. Method

Ten female and 10 male subjects in each of four age groups (2.5-3.0 years; 4.5-5.0 years; 10-11 years; and adults) (N=80) were audio tape recorded as they produced five spontaneous repetitions of the minimal pair monosyllabic words "tot" and "dot". All subjects passed age appropriate hearing and speech screening tests. Computer files were made for each recorded test word using CSpeech. The recordings were low-pass filtered at 10.5 kHz, and sampled at 26 kHz. PERCEPT software operating with CSRE 3.0 was used to present the computer files of the recorded utterances to listeners for perceptual identification and to record and score their responses. Five speech-language pathologists judged the word initial plosive as /t/, /d/, or ambiguous with respect to voicing, for each token recorded from the subjects. Words were presented in blocks according to speaker age. The 10 words spoken by each of the 20 subjects in an age group were randomized across speaker and initial consonant.

CSpeech was used to obtain measures of VOT, F1 and F0 at vowel onset, and BA, for one token for each

voicing condition per subject, randomly selected from the subject's perceptually validated tokens. Perceptually validated tokens were those where at least four of five judges correctly identified the voicing feature of the initial consonant. This provided 80 tokens for analysis for each voicing category. F0 and F1 values were transformed to log Hz so that differences between voicing conditions within an age group could also be compared across age groups (different vocal tract sizes).

3. Results

The youngest age group (2.5-3.0 years) had significantly fewer correct and significantly greater incorrect and ambiguous identifications, compared to the older groups. The youngest age group also had significantly fewer perceptually validated tokens than the older age groups. While all speakers produced some perceptually distinct voicing contrasts, listeners' reliability in judging the contrasts increased with speaker age.

For all age groups, VOT was longer, F1 onset frequency was higher, and burst amplitude was greater for /t/ than /d/, while F0 onset frequency did not differ significantly. No significant age effects or age by voicing condition interactions were found for VOT, F0 or BA. Variability on the acoustic measures decreased as age increased. A discriminant function analysis was conducted using one perceptually validated token for each voicing category from half of the speakers in each age group. To make meaningful spectral comparison across ages groups (who varied in vocal tract size), the F1-F0 difference in log Hz was entered as the "spectral variable". This analysis revealed that VOT was the primary cue for the perception of voicing while F1-F0 information at the onset of the vowel, and to a lesser extent burst amplitude, provided secondary cues. The ability of the function to discriminate between /t/ and /d/ was statistically significant (chi square (3) = 109.51, p=0.000). The classification function for the tokens that were used to derive the discriminant function prediction equation correctly classified 96.25% of these tokens. The classification for the cross-validation procedure which used a perceptually validated token from the remaining half of the speakers in each age group correctly classified 98.75% of these tokens.

4. References

[1] Forrest, K. & Rockman, B. (1988). Acoustic and perceptual analysis of word-initial stops in phonologically disordered children. *Journal of Speech and Hearing Research*, 31, 449-459.