IS THE VOICED-VOICELESS PHONEMIC BOUNDARY INFLUENCED BY AN INTENSITY LEVEL OF THE PRESENTATION

Elzbieta B. Slawinski and Jane F. MacNeil
Psychology Department, The University of Calgary
2500 University Drive NW Calgary, Alberta T2N 1N4

1. Introduction

Multiple acoustic parameters contribute to the perceptual distinction of many phonetic contrasts. The importance of timing of voice onset relative to the plosive release and as well of other acoustic cues to the voiced/voiceless distinction for stop consonants has been studied in a number of phonetic environments. Many of voiced/voiceless phonetic distinction correlates for stop consonants in initial position have been examined using synthetic speech, including voice onset time (Abramson & Lisker, 1970), the presence of a voiced transition at the onset of voicing (Liberman, Delattre, & Cooper, 1958; Stevens & Klatt, 1974), a pitch change prior to voicing onset (Haggard, Ambler, & Collow, 1970), and the frequency of the first formant of voicing onset (Summerfield & Haggard, 1974; Kluender, 1991; Kluender & Lotto, 1993). All of these studies indicate that the duration of Voice Onset Time (VOT) is a dominant and decisive phonetic correlate of the phonemic (voiced) contrast for stop consonants in a word-initial position. Thus, the difference in the duration of VOT in naturally produced plosive consonants in the initial position of words serves to distinguish voiced and voiceless tokens spoken by native talkers of eleven languages (Lisker & Abramson,1964).

However, it was also found that beside the VOT other acoustical cues contribute to the phonemic boundary expressed by the VOT. For example, a longer VOT is required to judge a phoneme as voiceless when first formant (F1) transition is longer or FI frequency onset is lower (Summerfield & Haggard, 1977). Moreover, an analysis of speech production data demonstrated that voiced and voiceless stop-consonants vary also in the peak intensity and in the duration of the burst of frication noise. The frication noise is of a longer duration and of a higher intensity at the release of a voiceless plosive (Klatt, 1975). Thus, the perceptual categorization of bilabial stop-consonants in word-initial position, while mostly relying on a difference in the VOT, should also depend on an acoustical cue such as loudness of the noise burst. Therefore, the present study examined the influence of the intensity level (Sound Pressure Level) of a stimulus presentation on the voiced/ voiceless phonemic boundary between bilabial stop consonants in the initial position of words.

2. Method

2.1 Stimuli. Ten stimuli differing in duration of VOT and ranging from [ba] to [pa] were synthesized using a parallel/cascade synthesizer KLSYN88a implemented on a Macintosh II computer using a 10 kHz sampling frequency. The duration of 6 ms noise burst was constantly maintained for all stimuli. The first formant (F1) started at 200 Hz and increased in 40 ms to 720 Hz, the second formant (F2) started at 900 Hz and increased in 40 ms to 1240 Hz, the third formant (F3) started at 2000 Hz and achieved a steady state of 2500 Hz in the same time as F1 and F2. The fourth and fifth formants were constantly maintained at 3600 Hz and 4500 Hz, respectively, across the entire stimulus. Each synthetic syllable was 300 ms in duration and during voiced excitation the fundamental frequency was 120 Hz until the last 160 ms when the fundamental frequency fell to 100 Hz. Change from an initial voiced to an initial voiceless stop consonant was achieved both by delaying the onset of energy in F1 relative to higher formants (F2 and F3), and by exciting F2 and F3 with an aspiration noise prior to the onset of a periodic source. Change from an initial voiced to an initial voiceless stop consonant was accomplished by changing VOT value (the timing of periodic source onset relative to burst noise release) from 4 ms to 40 ms in 4 ms steps. Two series of stimuli differing by intensity levels (60 dB SPL and 80 dB SPL) were generated. Thus, these series differed by the intensity (loudness) of the noise burst.

2.2 Subjects. Twelve subjects with normal hearing (<10 dB HL for 250 Hz to 6 kHz range) participated in the experiment. Ten subjects learned English as their first language, and two were bilingual. All subjects except three were phonetically naive and had never participated in a speech perception experiment.

2.3 Procedure. Subjects were tested individually on an identification task in an anechoic room. Two series of stimuli differing by intensity levels (60 dB SPL and 80 dB SPL) were presented via headphones. Subjects were asked to identify syllables [ba] and [pa] and were instructed to press one of two buttons labeled either "ba" and "pa". Subjects were exposed to two randomized series of 100 stimuli (each of 10 stimuli presented 10 times) with a 15 minute break between series. Stimuli presentation and a collection of responses were controlled by a Macintosh II computer.

3. Results and Discussion

The mean percentage of the responses labeled "ba" pooled across subjects for 60 dB and 80 dB listening conditions are plotted in Figure 1.

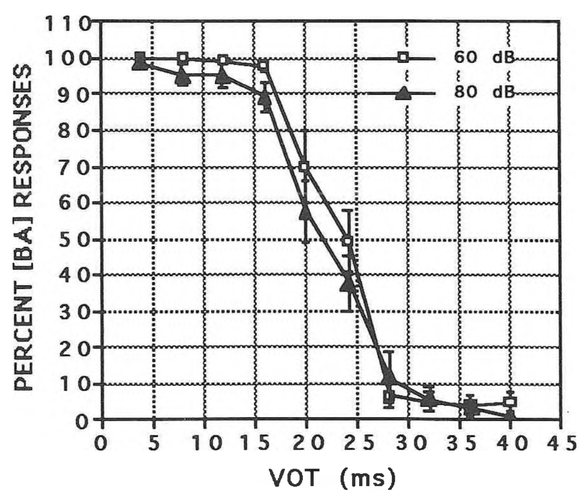

Fig. 1. Identification functions obtained for 60 dB and 80 dB conditions. Standard errors are indicated by errors bars.

In both listening conditions subjects have demonstrated categorical perception. However, it is possible to notice that the phonemic boundary, corresponding to 50 percent of identified [ba], is shifted towards shorter VOT values when stimuli were presented at 80 dB SPL compared to that obtained at 60 dB SPL. Moreover, subjects reported that they heard a more prominent noise burst when stimuli were presented at 80 dB SPL compared to that obtained at 60 dB SPL.

In order to assess the differences in the slopes of the psychometric functions for two listening conditions an interval of uncertainty (difference in VOT corresponding to 75% and 25% of [ba] responses) was measured. The obtained values of 5.3 ms and 7.9 ms for 60 dB and 80 dB respectively were significantly different ($p<.05$) and they suggest that subjects were more uncertain in their judgments when they were exposed to the higher intensity condition (Figure 2).

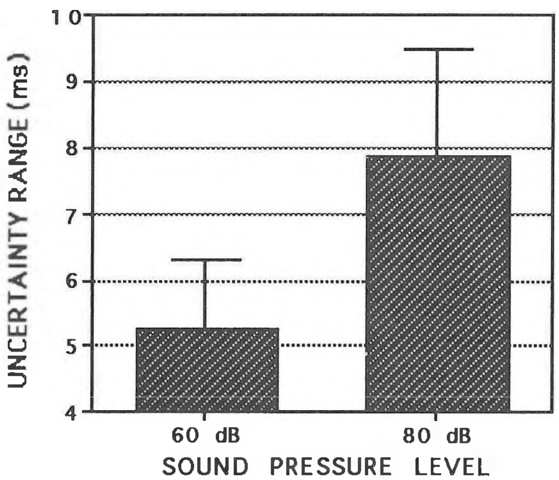

Fig. 2. Mean uncertainty ranges obtained for the 60 dB and 80 dB conditions. Standard errors are indicated by errors bars.

Individual subject phoneme boundaries were calculated using the best fitting logistic function to a continuum of identification probabilities, and subsequently condition-average phoneme boundaries were derived. Figure 3 displays phonemic boundaries obtained for the 60 dB (23.8 ms) and 80 dB (21.4 ms) conditions pooled across subjects.

The difference between these phonemic boundaries of 2.4 ms was found to be statistically significant ($p<.01$). Subjects were more inclined to identify stimuli as [pa] when they were exposed to the noise burst of higher intensity (loudness). Thus, it seems that loudness of the noise burst influences a judgment of the phonemic boundary between [ba] and [pa] categories. Thus, in order to assign a speech sound to the [p] category the shorter VOT was required in the presence of the louder noise burst. This result corresponds to Klatt's findings (1975) that the produced [pa] is characterized by the louder noise burst than that observed in the produced [ba]. Moreover, a study of Kobayashi and Honda (1991) has demonstrated that speakers with an electric larynx in order to make a distinction between voiced and voiceless plosives produce longer duration and higher amplitude of frication noise for the voiceless stops than for the voiced cognates. These researchers have also found that duration and amplitude of the noise burst were highly correlated with the perception of "voicelessness." Therefore, the magnitude of perceived loudness of a noise burst influences the categorization of the [ba-pa] phonemic contrast. Such impact of the noise burst is probably due to the effect of forward masking which in turn depends on the intensity of the noise burst.

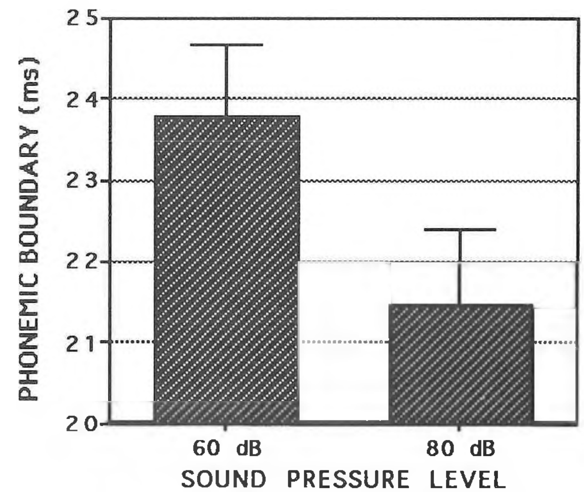

Fig. 3 . Mean phonemic boundaries obtained for the 60 dB and 80 dB conditions. Standard errors are indicated by errors bars.

The difference in the phonemic boundary observed between two intensity conditions may reflect a trading relation between VOT and loudness of a noise burst. Thus, in order to enhance perception of [b] and [p], the perceptual effects of changing one acoustic cues (VOT) may be offset by changing the other cue (loudness of a noise burst) in the opposing direction.

References

1. Abramson, A., and Lisker, L. (1970). Discrimination along the voicing continuum: Cross-language tests. in Proceedings of the 6th International Congress of Phonetic Science, Prague, 1967, 569-573.
2. Haggard, M., Ambler, S., and Callow, M. (1970). Pitch as a voicing cue. JASA, 47, 613-617.
3. Klatt, D.H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. Journal of Speech and Hearing Research, 18, 686-706.
4. Kluender, K.R. (1991). Effects of first formant onset properties on voicing judgments result from processes not specific to humans. JASA, 90, 83-96.
5. Kluender, K.R., and Lotto, A.J. (1993). Effects of first formant onset frequency on [-voice] judgments result from auditory processes not specific to humans. JASA, 95, 1044-1052.
6. Kobayashi, N., and Honda, K. (1991). Production and perception of the consonantal voicing distinction in speech without VOT: An acoustic study using electric larynx speech. JASA, 90, S2362.
7. Liberman, A.M., Delattre, P., and Cooper, F. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. Language and Speech, 1, 153-167.
8. Lisker, L., and Abramson, A.S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. Word, 20, 384-422.
9. Stevens, K.N., and Klatt, D.H. (1974). The role of formant transitions in the voiced-voiceless distinction for stops. JASA, 55, 653-659.
10. Summerfield, Q., and Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. JASA, 62, 435-448.