Identification of bilabial plosives: Integration of VOT and burst intensity information.

Elzbieta B. Slawinski and Nhan Lau Psych. Dept., The University of Calgary
e-mail: eslawins@acs.ucalgary.ca

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, a pitch change prior to voicing onset, the presence of a voiced transition or the frequency of the first formant at the onset of voicing. 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.

Moreover, it was demonstrated that produced voiced and voiceless stop-consonants vary also in the peak's 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. Thus, the perceptual categorization of bilabial stop-consonants in word-initial position, while mostly relies on a difference in the VOT, might also depend on an acoustical cue such as intensity of the noise burst (plosive release). Therefore, a goal of the present study was to examine the influence of the intensity level (Sound Pressure Level) of the initial burst on the categorization of bilabial stop consonants in the initial position of words into voiced and voiceless phonemes.

2. Method

2.1 Stimuli. Ninety nine stimuli differing in the duration of VOT and the intensity of initial frication burst, and ranging from [ba] to [pa] were generated using a synthesizer Monet (Trillium Sound Inc.) implemented on a NEXT computer using a 44 kHz sampling frequency. The duration of noise burst was constantly maintained for all stimuli. The first formant (F1) started at 200 Hz and increased in 40 ms to 760 Hz, the second formant (F2) started at 900 Hz and increased in 40 ms to 1600 Hz, the third formant (F3) started at 2000 Hz and achieved a steady state of 2400 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 175 ms in duration. 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 from 10 ms to 30 ms in 2 ms steps. Moreover, the intensity of the burst varied from 0 dB SPL to 30 dB SPL in 3 dB increments.

2.2 Participants. Twenty eight young adults with normal hearing (<10 dB HL for 250 Hz to 8 kHz range) participated in the experiment. All listeners were phonetically naive and had never participated in a speech perception experiment.

2.3 Procedure. Listeners were tested individually on an identification task in an anechoic room. The stimuli were presented to the listeners via headphones AKG-240 at 70 dBA. Listeners were asked to identify syllables [ba] and [pa] and were instructed to press one of two buttons labeled either "ba" and "pa". Listeners were exposed to randomized 792 stimuli (each of 99 stimuli presented 8 times). Stimuli presentation and a collection of responses were controlled by a NEXT computer.

3. Results and Discussion

The mean percentage of the responses labeled "ba" for each stimulus and each participant were calculated. ANOVA with VOT and burst intensity as factors and the percent of [ba] responses as dependent variable indicated that the main effect of VOT ($F(2, 2755)=1137.32$, $p<0.0001$) and the main effect of intensity ($F(2, 2755)=66.006$, $p<0.0001$) were significant. Thus, both acoustical cues, VOT and burst intensity play roles in a distinction between [b] and [p]. Furthermore, the stimuli differing in the intensity were grouped into three classes: low (0 dB - 9 dB), moderate (12 dB - 21 dB) and high (24 dB - 30 dB). Similarly, stimuli that differ in VOT were categorized into three groups: short (10 ms - 16 ms), medium (18 ms - 24 ms), and long (26 ms - 30 ms). Separate two-factorial ANOVAs on the percent of [ba] responses were conducted for each listener. These analyses revealed that 25% of all listeners (Group 1), while categorizing tokens into [b] and [p], relied mostly on VOT information. However, majority (75%) of listeners (Group 2) in addition to VOT information relied very strongly on burst intensity's information. The 'Group 2' of listeners was more proficient in classifying stimuli, differing in their VOT, into two categories ([ba] and [pa]) than the 'Group 1' (Figure 1). Figure 1 shows that listeners of the 'Group 2' were more likely to assign stimuli into ba-category in a presence of a low burst intensity, and were less likely to identify stimuli as [ba] when the burst intensity was high than were the listeners of the 'Group 1'.

This finding suggests that all listeners are using information about VOT and burst intensity while categorizing initial bilabial stop consonants into voiced and voiceless phonemes, however, they place a different weight to available information.

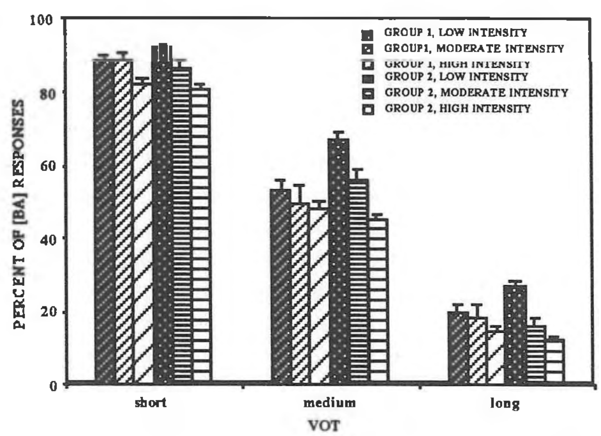

Figure 1. Percent of [ba] responses as a function of VOT for different parameters of the burst intensity and group of listeners. Standard errors are presented as error bars.