THE EFFECT OF AGE ON ACOUSTIC CUE WEIGHTING IN THE PERCEPTION OF INITIAL STOP VOICING CONTRAST IN HEBREW

Taitelbaum-Swead Riki, Hildesheimer Minka, Kishon-Rabin Liat

Department of Communication Disorders, Sackler Faculty of Medicine,Tel-Aviv University
taite@post.tau.ac.il

ABSTRACT

Listeners use several acoustic cues with different relative weighting, for perceiving a phonetic contrast. The weighting depends on variables such as age and language. The purpose of this study was to examine the relative weighting of some acoustic cues to the perception of initial voicing of plosives in different age groups. Three groups of children and one group of adults, Hebrew speakers were evaluated. The relative weighting of three cues were tested: Voice Onset Time (VOT), initial burst, and the transitions to the following vowel. A similar hierarchy of cues was found in all groups, with VOT the primary one, although not in the same weighting. Adults give more weight to transitions and burst compared to children. These findings indicate the ability of the auditory system to focus on the important cues already at young age. However, there appears to be a continued shift in the weighting during development.

Keywords: voicing perception, VOT, acoustic cues

1. INTRODUCTION

Several components of the speech signal appear to provide cues to the perception of a phonologic contrast. The listeners do not pay equal emphasis to all acoustic cues, but rather seem to rely more heavily on some acoustic cues compared to others [1].For example, the perception of voicing of plosives in initial position, seems to be mediated mainly by Voice Onset Time (VOT) in many languages [2,3]. There are, however, secondary acoustic cues such as first formant transition, initial burst and fundamental frequency that contribute to the perception of voiced plosives [4,5,6]. Furthermore, studies suggested that the weight assigned to each of these cues for auditory perception may depend on many variables such as the listener's language [3], the phonetic context [7] and the linguistic experience[8,9,10].In Hebrew, for example, the VOT boundaries in the perception of initial voicing were found to be more negative than those found in English [11,12]. This suggests that the acoustic cues and their relative weighting might be different in Hebrew from those reported in other languages. Furthermore, the relative effect of the different cues may be age related.

Thus, the purpose of the present study was to examine the relative weighting of three acoustic cues (VOT, burst, transitions) to the perception of initial voicing in Hebrew in different age groups.

2. METHODS

2.1.1 Subjects

A total of four groups of subjects participated in the study, one group of adults, and three groups of children. Fourteen female adult subjects, 22-24 years of age (average of 23 years old), natives of the Hebrew language participated. All had pure-tone air-conduction thresholds less then 15 dB HL bilaterally at octave frequencies from 250 - 4,000 Hz.

Three groups of children aged 4-5 years (n=11), 6-7 years (n=11) and 9-10 years (n=11) were evaluated. Only children who met the following criteria were included in this study: native speakers of the Hebrew language, without previous exposure to another language, normal motor, speech, language, hearing and cognitive development (based on a detailed questionnaire), and normal articulation. All children passed a hearing screening test, which consisted of pure tone of 0.5, 1, 2 4, and 6 KHz, present at 25dBHL via headphones [13]. Two children from the youngest group were eventually dropped from the study due to lack of attention to the task.

2.1.2 Test Material Recording

Stimuli consisted of meaningful words /bil/-/pil/ (name of male, elephant), that differ in the voicing of the initial stop consonant and were suitable for pictorial representation. For this study natural stimuli were chosen because they resemble speech perception in real- life situations. A female speaker native of the Hebrew language was selected from a pool of five speakers. She was chosen based on the clarity of her speech judged by experienced listeners, and by the ability to distinguish speech components both spectrally and temporally. All stimuli were similar in their length and their relative amplitude values of their three major segments: voicing lead, burst and the vowel. The utterances were digitally recorded in a sound treated room via a JVC MV40 microphone and using the Kay-CSL 4030 software. Sampling rate was 22,050 Hz and quantization level set at 16 bits. Spectrograms of /bil/ -/pil/ pair are shown in Figure 1.

2.1.3 Stimuli Generation

Voice-onset time values, duration of the burst and duration of the vowels for each recording stimulus, were measured using both the Praat version 4.2 software (for spectral parameters) and Sound Forge 4.5 (for temporal parameters). Four different continua of each pair were constructed. For all continua VOT varied from -40 msec to +40 msec, every 10 msec. In other words, each continuum consisted of 9 stimuli. Using the Sound Forge wave editor, we created a "pool" of five segments: (1) the burst extracted from the voiced /bil/; (2) the burst extracted from the voiceless /pil/; (3) the vowel extracted from the voiced /bil/ (transition + steady state); (4) the vowel extracted from the voiceless /pil/ (transition + steady state); and (5) the voicing lead extracted from the voiced /bi/ of approximately -190 msec. Only a 40 msec portion of the voicing lead was used. The VOT continua were created by combining segments from this pool. In the first VOT continuum, segments 1 & 3 were combined to create the consonant -vowel (CV) combination named bi_b. In the second continuum, segments 2 & 3 were combined to create the CV combination named pi_b. In the third continuum, segments 2 & 4 were combined and in the fourth continuum segments 1 & 4 were combined to create the CV combination named pi_p and bi_p, respectively. In order to create the VOT continua, the voicing lead stimulus was used, from which 10 msec segments were deleted successively. To create a positive VOT, silence was added between the end of the burst and the onset of the vowel transition. It is important to note that at the point of concatenation, amplitude was increased and decreased gradually to ensure that all stimuli sounded naturally with no additional confounding acoustic cues.

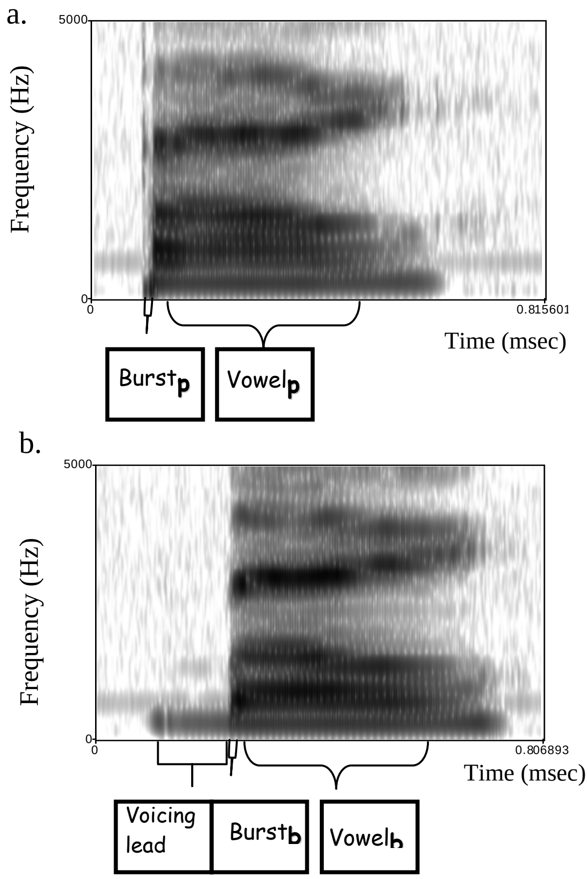

Figure 1: Temporal- spatial representations (spectrograms) of the /pil/ stimuli (figure 1a) and of the /bil/ stimuli (figure 1b).

2.1.4 Procedure

Each subject was asked to listen to a set of contrasts /bil/-/pil/, including four continua. Subjects were tested in a sound treated room. Stimuli were presented binaurally through TDH-50 headphones at a most comfortable level. Each VOT continuum was tested separately. Stimuli were presented in random order, directly from the computer. The collection of responses was via software control. The identification task was based on two alternative forced choice response procedures. Each of the 9 stimuli (from –40 msec to +40 msec) was repeated 10 times in each continuum (a total of 90 stimuli/continuum). For the youngest group of children (4-5 years), the paradigm was shorter; with 9 stimuli (from -40 msec to +40 msec) each repeated only five times (a total of 45 stimuli/continuum) due to limited attention span. It should be noted that in a separate study which evaluated short (45 stimuli /continuum) compared to long (90 stimuli /continuum) paradigms, no differences between the paradigms were found in ten adults. All subjects were introduced to the target words. Responses were obtained from adults by pressing one of two keys (option 1- bil, and option 2- pil). No feedback was provided. In the children groups, pictures of the two objects were presented via a computer screen and the children were asked to repeat the word they heard and to point on the relevant picture on the computer screen. Prior to testing, the children were trained on six trials of randomized endpoint stimuli to familiarize them with the nature of the task. Feedback on the correct responses was provided only during the training session**.**

3. RESULTS

For each subject, the identification scores of /bil/ were computed. Figure 2 illustrates the mean identification functions for each age group and each continua (bi_b,bi_p,pi_p,pi_b) as a function of a VOT continuum. As seen from the figure, for all groups, stimuli with negative VOT values were identified as voiced plosives whereas stimuli with positive VOT values were identified as voiceless. Furthermore, it can be seen that the phonetic boundary (PB) shifted according to the manipulations that were conducted. Phonetic boundaries were more negative when all cues were extracted from the word with voiceless plosive (***pi$_p$***), and positive when all cues were extracted from the word with voiced plosive (***bi$_b$***).However, when conflicting cues were extracted, i.e., the burst from the voiced stop and the transitions from the voiceless stop and vice versa (***bi$_p$*, *pi$_b$***), the values of the boundary were between the two extremes. In addition, all groups showed a gap between the curves which represents the influence of the burst and transitions on the identification of voicing.

The younger group (4-5 years) showed less categorical functions, shallower slopes and

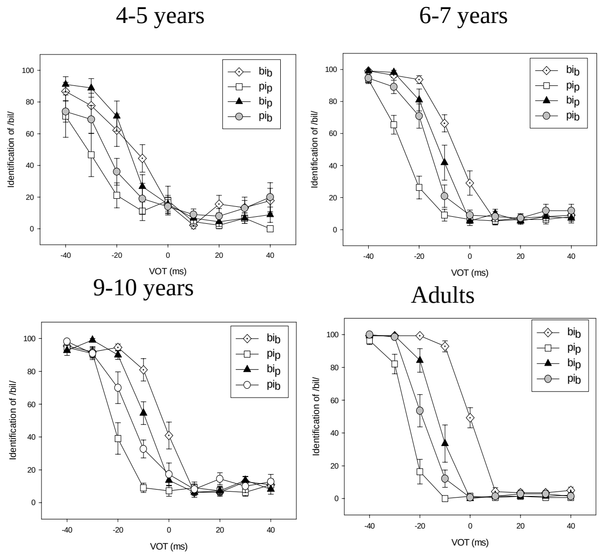

Figure 2: Mean identification functions +SE, for each age group (4-5 years, 6-7 years, 9-10 years and adults), for /bil/-/pil/ stimuli in four different continua (***bi$_b$*,*bi$_p$*,*pi$_p$*,*pi$_b$***).

negative PB values. A one way repeated measures analysis of variance (ANOVA) revealed a significant effect of age ($p<0.01$) on the PB. The 4-5 years age group showed significantly more negative PB values compared to the other groups. Also found was a significant effect of the following acoustic cues: burst ($p<0.01$) and transitions ($p<0.01$) on the PB values.

In order to test the relative effect of each of the acoustic cues on the identification of voicing, a regression analysis with a mixed model was used. Standardized coefficient (beta/SE), p-values and beta values for each of the acoustic cues and each group are shown in Table 1.

Table 1: standardized coefficient (beta/SE) values, p-values and beta values for the VOT, burst and transitions in each age group

/bil/-/pil/		VOT	transitions	burst
adults	beta/SE	-31.8**	-5.2**	-7.7**
	beta	-1.36	-11.5	-16.9
9-10 years	beta/SE	-29.1**	-3.4**	-4.7**
	beta	-1.2	-7.4	-10.9
6-7 years	beta/SE	-28.5**	-3.9**	-5.3**
	beta	-1.2	-8.6	-12.2
4-5 years	beta/SE	-12.9**	-2.6**	-3.5**
	beta	-0.9	-6.9	-10.1

**beta/SE<0.01

The weight of each of the acoustic cues is represented by its coefficient. For comparison between the cues, we used the coefficient divided by the standard error to account for differences in scales. A higher value of the beta/SE, suggests that the listener relies more heavily on this cue. As can be seen from Table 1, all groups relied heavily on the VOT, less on the burst and the least on the transitions. In order to compare between the groups the raw beta was used due to differences in the group size. As shown in the table, adults gave more weight to the burst and the transitions compared to the other groups, whereas, 4-5 years old gave less weight to the VOT compared to the other groups.

4. DISCUSSION

Altogether, the results show that (1) VOT seems to be the primary but not the sole cue for the perception of initial voicing of bilabial plosives in Hebrew; (2) a similar hierarchy of the different cues was found in all age groups; (3) a different weighting of the secondary cues was found between the adult and children's groups; (4) while the youngest group already showed categorical perception using all three acoustic cues, shallower slopes as well as more negative PB values were found in this age group.

It is interesting to note that although the bilabial burst is very short (around 10 msec), listeners still relied on this brief temporal-spectral information for the perception of initial voicing. The finding that the listeners relied the least on the transitions to the vowel may be explained by similar first formant transitions for the voiced and the voiceless plosives in the /i/ context.

In perceiving initial voicing of bilabial plosives, in Hebrew, children and adults use the same cues with similar hierarchy. This finding emphasizes the ability of the auditory system to focus on the important cues already at a young age. It is possible that it is an innate ability or acquired in the first years of life. Eimas et al., [14] showed that young infants possess the acoustic basis to discriminate categorically between different stimuli. Nevertheless, the findings that children give less weight to the secondary cues (transitions burst) compared to adults, suggest that they have not completed the maturation process of the auditory system and the linguistic development.

The finding that only children in the youngest group showed more negative PB values compared to all other groups suggests that the category boundaries of initial voicing may develop till the age of six years.

This study provides, additional knowledge regarding the normal development of the acoustic-phonetic stage in speech perception process. Testing this paradigm in special populations such as hearing impaired including those with cochlear implants, may provide insight to the spectral and temporal information transmitted via the CI system.

5. REFERENCES

[1] Dorman, M.F., Studdert-Kennedy, M., Raphael, L.J., 1977.Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context dependent cues. Percept. Psychophys. 22, 109-122.

[2] Van Alphen, Smits, R., 2004. Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. J Phonetics. 32, 455-491.

[3] Abramson, A.S., Lisker, L.,1970. Discriminability along the voicing continuum: cross language tests. Proceedings of the 6th International Congress of Phonetic Sciences. Prague: Academia. 569-573.

[4] Benki, J.R., 2001. Place of articulation and first formant transition pattern both affect perception of voicing in English. J Phonetics. 29, 1-22.

[5] Bernstein, L.E., 1983. Perceptual development for labeling words varying in voice onset time and fundamental frequency. J Phonetics. 11, 383-393.

[6] Repp, B.H., Lin, H.B., 1989. Acoustic properties and perception of stop consonant release transients. J Acoust Soc Am. 85,379-96.

[7] Summerfield, Q., Haggard, M., 1974. Perceptual processing of multiple cues and contexts: effects of following vowel upon stop consonant voicing. J Phonetics. 2, 279-295.

[8] Hazan, V., Barrett, S., 2000 The development of phonemic categorization in children aged 6-12. J Phonetics. 28, 377-396.

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

[10] Mayo, C., Turk, A., 2004. Adult-child differences in acoustic cue weighting are influenced by segmental context: children are not always perceptually biased toward transitions .J Acoust Soc Am. 115,3184-94.

[11] Kishon-Rabin, L., Rotshtein, S., Taitelbaum, R., 2002 Underling mechanism for categorical perception: tone- onset time and voice- onset time evidence of Hebrew voicing. J Basic Clin Physiol Pharmacol. 13,117-134.

[12] Horev, N., Most, T., Pratt, H., 2007. Categorical perception of speech (VOT) and analogous non-speech (FOT) signals: behavioral and electrophysiological correlates. Ear Hear. 28,111-28.

[13] ANSI Specifications for audiometers (ANSI, New York) 1989.

[14] Eimas, P.D., Siqueland, E.R., Jusczyk, P., Vigorito, J., 1971. Speech perception in infants. Science. 171,303-306.