Development of voicing contrast: A comparison of voice onset time in stop perception

[1]Savithri, S. R., [2]Pushpavathi, M. & [3]Sujatha V. Shastry

Abstract

The study investigated development of voicing contrast in 4-7 year old Kannada speaking children. Six children each in the age range of 4-5, 5-6, and 6-7 years participated in the study. Voice onset time (VOT) continuum of /g-k/, /d -t/, and /b – p/ were synthesized. Subjects listened to the tokens and identified the stops on a binary forced-choice format. The results indicated that VOT was a strong cue in perception of voicing in Kannada.

Key words: Voice onset time, continuum, binary forced-choice method, 50% cross over, boundary width

Voice onset time (VOT) is the time difference between the release of a complete articulatory constriction and the onset of phonation (Lisker & Abramson, 1964). Voiced stop consonants are characterized by lead VOT; that is the voicing precedes the release of the articulator. Unvoiced stop consonants are characterized by lag VOT; that is, the voicing follows the articulatory release. Several studies have been conducted on VOT as a cue for perception of voicing. The results of these studies indicate that there is a developmental trend in VOT (Zlatin & Koenegskenecht, 1976) and that a poor performance in disordered population compared to normal controls (Elliot, Hammer, & Scholl 1989; Thibodau & Sussman, 1979). All these studies are in English. Therefore, these data can't be generalized to other languages owing to the differences between stops among languages. English has a two-way classification of stops (voiced – unvoiced) where as Kannada has a four-way classification (unvoiced unaspirated, unvoiced aspirated, voiced unaspirated and voiced murmured). Therefore, the present study investigated the development of voicing contrast in Kannada speaking children.

Method

Stimuli: Three meaningful Kannada words (gaḍi, dad̠a, bad̠i) with velar /g/, dental /d̠/, and bilabial /b/ in the initial position as uttered by a 22 year old Kannada speaking normal female subject were audio recorded and digitized and stored on to the computer memory. Using the waveform display program of the SSL (Voice and Speech Systems, Bangalore) lead VOT was truncated in steps of 2 pitch pulses till burst was reached. Following this, silence in steps of 10 ms was added till the lag VOT approximated adult values. Thus synthetic tokens for VOT continuum of /g-k/ (-96 to +30), /d̠ -t̠/ (-70 to +30), and /b – p/ (-89 to +30) were generated. The tokens were randomized and iterated 10 times with an inter token interval of 3 seconds and interest interval of 5 seconds. A total of 330 tokens formed the material.

Subjects: Six Kannada speaking normal children each in the age ranges of 4-5, 5-6, and 6-7 years and 6 adults participated in the study. All the subjects had normal speech and hearing.

Procedure: Subjects were tested individually. Stimuli were binaurally presented through heard phones at comfortable listening levels. Subjects were instructed to listen to the token carefully and identify the token as having voiced / unvoiced stops. Subjects pointed to toys representing words with voiced / unvoiced stops. Prior to the testing subjects were provided with a training session where they pointed to the toys on listening to natural tokens. The experimenter recorded subject's response on a binary forced–choice format. Percent response of subjects for each stop was calculated and an identification curve was plotted from which 50% cross over point (point where the 50% of percent was for voiced / unvoiced stop), lower limit (point which had 75 % response for voiced stop), upper limit (point which had 75 % response for voiced stop), and phoneme boundary width (time difference between upper and lower limit) were calculated.

[1]Professor and Head, Department of Speech-Language Sciences, AIISH, Mysore 570 006, e-mail: savithri_2k@yahoo.com, [2]Reader and HOD-Clinical services, AIISH Mysore 570 006 [3] Clinical Assistant, Department of Speech Language Pathology, AIISH, , Mysore 570 006

Results

In general, the 50% cross over from voiced to unvoiced stop occurred in the lead VOT region. The shift occurred earlier in velar followed by bilabial and dental stops. $-year old children showed earlier cross over compared to 7-year old children except for bilabial stop. Also, children in the age range of 4-5 years showed earlier cross over compared to adults. Table 1 shows the 50% cross over points.

Stimuli	4-5	5-6	6-7	Average	Adults
g-k	-17	-30	-10	-19	-19
d-t	-28	-7	-9	-14	-18
b-p	-12	-14	-26	-17	-5

Table 1: 50% cross over (ms) in all groups

Lower limit was observed in the lead VOT region for all age groups. Lower limit occurred earlier in children in the age range of 5-6 years compared to other age groups. In general, it was highest for dental place of articulation followed by velar and bilabial place in both children and adults. Table 2 shows the lower limit in all age groups.

Stimuli	4-5	5-6	6-7	Average	Adults
g-k	-44	-50	-18	-37	-37
d-t	-32	-34	-63	-43	-43
b-p	-5	-65	-4	-25	-36

Table 2: Lower limit (ms) of phoneme boundary width in all groups.

Upper limit of the boundary width occurred in lead VOT region in children in most stops except for g-k (6-7 years) and b-p (4-5 years). No development trend was seen in children. Table 3 shows the upper limit of phoneme boundary width in all age groups.

Stimuli	4-5	5-6	6-7	Average	Adults
g-k	-2	-5	+2	-1.8	-2
d-t	-5	-16	-22	-15	+8
b-p	+18	-5	-7	-2	-3

Table 3: Upper limit (ms) of phoneme boundary width in all groups

The boundary widths were higher in velar followed by dental and bilabial in children. The boundary width decreased from 4 years to 7 years (except d-t) though not linearly. Also children had higher boundary width compared to adults. Table 4 shows the boundary widths in all groups of subjects.

Stimuli	4-5	5-6	6-7	Average	Adults
g-k	42	45	16	34	25
d-t	27	19	41	29	26
b-p	13	60	3	25	13

Table 4: Phoneme boundary width (ms) in all groups

Discussion

The results indicated a 50% cross over in the lead VOT region with an average of –17 ms in children and –5 ms in adults. A comparison of 50% cross over points as obtained by various authors is in table 5.

Author (year)	Language	Stimuli	50 % cross over
Yeni Komshian et al. (1967)	English	/g/	+35
Simon (1974)	English		+ 15-20
Zlatin et. al. (1975)	English		+25
Flege & Eefting (1986)	English		+36.2
	Spanish		19.9
Williams (1977)	English		+25
	Spanish		-4
Lisker & Abramson (1967)	Thai		-20
Sathya (1996)	Telugu	g, d, b	-10 (children) -20 (adults)
Present study (1996)	Kannada	g, d, b	-17 (children) -5 (adults)

Table 5: 50% cross over as obtained by various authors.

Williams (1977) speculated that Spanish listeners give greater weight to prevoicing as a cue to voicedness than English listeners and greater weight to the presence of an audible release burst and the lack of low frequency energy immediately following it, as cues to voiceless. The finding of Williams (1977) that the category boundary between /b/ and /p/ occurred along a VOT continuum occurred around –4 ms for Puertoricans who were monolingual Spanish speakers suggest that the phonetic processing of speech may be slowly attuned to the acoustic properties of stops found in a particular language (Aslin & Pisoni, 1980). Flege & Eefting (1986) imply that cross language research suggest that speakers of different languages may learn to perceive stops differently because they are exposed to different kinds of stop consonants. Further, English language environment listeners tend to identify both /b/ and /p/ as the phoneme /b/ and the prevoiced/ voiced contrast is physiologically irrelevant in English. This contrast is perceived categorically in other languages – for example Hindi, Spanish and Thai (Burnham, Earnshaw & Clark, 1991).

In addition to VOT there are three additional acoustic properties that vary in degree across art of the synthetic speech series (Williams, 1980). These variations are restricted to the voicing lag region of the continuum and are as follows:

(a) The presence/ absence of varying duration of aspiration or aperiodic energy in the interval between articulatory release and the onset of voicing. The presence of aspirated formants is an acoustic property that has been demonstrated to provide a positive cue for initial voicelessness to English listeners (Winitz et. al., 1975), (b) The absence of periodic acoustic energy at the level of F1 during periodic excitation of the vocal tract referred to as first formant cut back (Liberman, Delattre, & Cooper, 1958). There is also evidence that the presence or absence of periodic energy in the region of F1 provides a perceptual cue for an initial contrast in voicing for English listeners (Delattre, Liberman, & Cooper, 1955; Liberman et. al., 1958; Lisker, 1975), and (c) Differences in the degree and temporal extent of formant transitions under conditions periodic excitation of the vocal tract. There is some evidence that this acoustic variable may also provide a cue for initial voicing in English (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952; Stevens & Klatt, 1974; Summerfied & Haggard, 1974).

An examination of spectrograms of /b/ and /p/ taken from a Kannada speaker and those provided by Williams (1980) reveals that the stop consonants in Kannada are entirely different from that of English in all the 3 parameters listed above. Neither aspiration in the lag VOT region, nor F1 cut back is present in Kannada stop consonants. These differences in the acoustic properties of stops in English and Kannada might be reflected in the perception also with a 50% cross over in the lag VOT region for English and lead VOT region for Kannada. Also, syllable has been synthesized in the other studies and waveform editing has been used in the present study.

While comparing the discrimination data obtained from adult speakers of Thai and English for synthetic bilabial consonants, Aslin & Pisoni (1980) comment that the relative discriminability in the 20 ms of voicing lag is greater than in the – 20 ms region of voicing lead despite that fact that the slopes of the labeling functions for Thai subjects in these regions are very nearly identical. They propose that the smaller incidence of discrimination of VOT (and TOT) differences in the minus region of voicing lead value is probably due to the generally poorer ability of the auditory system to resolve temporal differences in which a lower frequency component precedes a higher frequency component (for un voiced stop and lower frequency component – voicing – precedes a higher frequency component for – burst release – for voiced stop). Aslin & Pisoni (1980) further commenting on infant studies on VOT suggest that the discrimination of the relative order between the onset of first formant and higher formants is more highly discriminable at certain regions along the VOT stimulus continuum corresponding roughly to the location of the threshold for resolving these differences phycho-physically. In the case of temporal order processing, this falls roughly near the region surrounding +/ -20 ms, a value corresponding to the threshold for temporal order processing (Hirsh, 1959). Further commenting on Pisoni's (1977) experiment on TOT (tone onset time), Aslin & Pisoni say that two distinct regions of high discriminability are present in the discrimination functions. Evidence of discrimination of VOT contrasts that straddle the –20 and + 20 ms regions of the stimulus continuum probably results from general sensory constraints on the mammalian auditory system to resolve small differences in temporal order and not from phonetic categorization.

The results also indicated wider boundary widths in children as compared to adults. Also, the boundary width decreased from 4 to 7 years. This indicates that there is a developmental trend in children on VOT perception. Based on the above findings it could be concluded that VOT is a cue for voicing distinction in Kannada and there is evidence for development of speech perception (VOT).

An audiocassette consisting of synthetic tokens for VOT has been prepared. This can be used with the clinical population to test the speech perception abilities in clinical population. Also, further studies can focus on various Indian languages and clinical population.

Conclusions

Earlier studies on perception of voicing contrast indicated change from voiced to unvoiced percept in the lag VOT region. But, the results of the present study indicated a 50% cross over in the lead VOT region with an average of –17 ms in children and –5 ms in adults. The results indicate that there are language differences in the perception of voicing contrast. This difference may be attributed to the differences in stops in a language.

References

Aslin, R.N. & Pisoni, D. B.1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (Eds.). *Child Phonology, Vol. 2: Perception*. New York: Academic Press.

Burnham, D.K. Earnshaw, C.J. & Clark, J.E. (1991). Development of categorical identification of native and non-native bilabial stops in infants, children and adults. *Journal of Child Language,* 18, 231-260.

Cooper, F.S., Delattre, P.C., Liberman, A.M., Borst, J.M. & Gerstman, L.J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America,* 24, 597-606.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transition cues for consonants. *Journal of the Acoustical Society of America,* 27,769-773
.
Elliot, L.L., Hammer, M.A. & Scholl, M.E. (1989). Fine grained auditory discrimination in normal children and children with language learning problems. *Journal of Speech and Hearing Research,* 32, 112-119.

Flege, J.E & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants, *Phonetica*, 43, 155-171.

Hirsh, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America,* 31 (6), 759-767.

Liberman, A.M., Delattre, P.C. & Cooper, F.S. (1958). Cited by M.A. Zlatin in Voicng contrast: Perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America,* 56, 981-994.

Lisker, L (1975). Is it VOT or a first formant transition detector? *Journal of the Acoustical Society of America*, 57 (6), 1547-1551.

Lisker, L. & Abramson, A.S. (1964). A cross language study of voicing in intial stops: Acoustic measurements. *Word*, 20, 384-422.

Lisker, L. & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384-422.

Lisker, L. & Abramson, A. (1967). Cited by M. A. Zlatin in Voicng contrast: Perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America,* 56, 981-994.

Pisoni, D.B. (1977). Identification and discrimination of the relative onset of two component tones: Implication for the perception of voicing in stops. *Journal of the Acoustical Society of America,* 61, 1352- 1361.

Sathya. K. (1996). Personal communication.

Simon, C. (1974). Some aspects of development of speech production and perception in children. *Speech communication seminar, 7-15, Stockholm*. London: John Wiley and Sons.

Stevens, K.N. & Klatt, D. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America,* 55, 643-659.

Summerfied, A.Q. & Haggard, M.P. (1974). Perceptual processing of multiple cues and contexts: Effects of following vowel upon stop consonant voicing. *Journal of Phonetics*, 2, 279-295.

Thibodau, L.M. & Sussman, H.M. (1979). Performance on a test of categorical perception of speech in normal and communication disordered children. *Journal of Phonetics*, 7, 375-391.

Williams, L. (1977a). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception and Psychophysics*, 21, 289-297.

Williams, L. (1977b). The voicing contrast in Spanish. *Journal of Phonetics*, 5, 169-184.

Williams, L. (1980). Phonetic variation as a function of second language learning. In G. H. Yeni-Komshian, J.F. Kavanagh & C.A. Ferguson (Eds.). *Child Phonology, Vol. 2: Perception*. New York: Academic Press.

Winitz et. al. (1975). Cited by Williams, L. (1980). Phonetic variation as a function of second language learning. In G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (Eds.). *Child Phonology, Vol. 2: Perception*. New York: Academic Press.

Yeni Komshian, G., Preston, M. & Cullen, J. (1967). A brief report on a study of imitation of synthetic speech stimuli by children. Haskins Laboratories: Status Report on Speech Research, SR-11, 48-49.

Zlatin, M. A. & Koenegskenecht, R. (1975). Development of the voicing contrast: Perception of stop consonants. *Journal of Speech and Hearing Research*, 18, 541-553.

Zlatin, M.A. & Koenegskenecht, R. (1976). Development of the voicing contrast: A comparison of voice onset time in stop consonants. *Journal of Speech and Hearing Research*, 19, 93-111.

Acknowledgements

This research was made possible by grants received from the Department of Science and Technology, Govt. of India (No. SP/ SO/ No 2/ 91). The authors acknowledge the financial support of DST. The authors are thankful to the Director, All India Institute of Speech and Hearing, Mysore for providing the infrastructure to carry out the project.