LANGUAGE IN INDIA
Strength for Today and Bright Hope for Tomorrow
Volume 10 : 5 May 2010
ISSN 1930-2940

Variation of Voice Onset Time (VOT) in Kannada Language

Manjunath. N., M.Sc. (Speech and Hearing) Candidate
Sneha Mareen Varghese, M.Sc. (Speech and Hearing) Candidate
Narasimhan. S .V., M.Sc. (Speech and Hearing)

Abstract

Speaking rate changes during normal conversation alters the duration of acoustic properties that specify phonetic segments (Summerfield, 1975). One such acoustic property which is affected by speaking rate is Voice Onset Time (VOT). As most of the studies on VOT are in English and therefore, do not provide sufficient data for generalization to other languages, the present study was planned to investigating the changes in VOT for voiceless and voiced stop consonants in Kannada language (a south Indian Dravidian Language) across different speaking rates.

A total of 20 Kannada speaking young adults, age ranging from 18 to 24 years participated in the study. A set of six phrases were constructed, which were meaningful and they had voiceless & voiced sounds in initial position of the words. Subjects were asked to read the phrases at three different speaking rates. The flashlight from a metronome was used to help the participants to maintain the speaking rate and all the utterances were acoustically analyzed for VOT.

Results revealed that, VOT values were higher at slower speaking rates and lower at faster speaking rates. These results are consistent with the earlier studies by Volaitis & Miller (1992). The results are also discussed in relation to the speech motor control with reference to the 'Target undershoot Model' by Lindblom (1963). The changes in VOT across speaking rates can also be attributed to the changes primarily in dynamic specification of gesture of articulators.

Introduction

Voice onset time (VOT) is the interval between the release of a stop consonant occlusion and the onset of the vocal fold vibration and is measured from acoustic displays as the time between the release burst and the first quasi periodicity in the acoustic signal (Lisker and Abramson, 1964; Keating, 1984; Klatt, 1975). There are three different types of Voice onset time. First, Zero VOT, it is the onset of vocal fold vibration coincides (approximately) with the plosive release. And the second one is Positive VOT, in this there is delay in the onset of vocal fold after the plosive release (Lag VOT). The third one is, the negative VOT, in which the onset of vocal fold vibration precedes the plosive release (Lead VOT).

VOT has been found to be an effective means to distinguish between voicing categories in oral stops. Studies have consistently shown the voiceless plosives /p/, /t/, /k/ to have long positive VOT'S in the range of 30 to 100 msec (English) and 1 to 45 msec (Kannada). In voiced plosives /b/, /d/, /g/ have negative VOT's in the range of -100 to 0 msec (English) and -126 to -60msec (Kannada). Thus the value of VOT is a good indicator for voiced and voiceless stops and also plays a large role in perceptual discrimination of phonemes of the same place of articulation. However, earlier investigations have shown that VOT is affected by language (Lisker & Abramson, 1964, 1967; Ravanan, 1993), age (Menyuk & Klatt, 1975) and gender (Ravishankar, 1981).

In recent years, more and more studies have focused on the relation between VOT and other possible correlates. The proposed correlates can be divided into two categories, speaker related and non-speaker related. The most widely studied speaker related factors are gender (Whiteside and Irving, 1998, 2001; Ryalls, 1997, 2002), age (Ryalls, 1997), Speaking rate (Kessinger and Blumstein, 1998; Volatis and Miller,1992), lung volume (Ryalls, 2002), other speaker related background are ethnic background (Ryalls et al 1997), dialectal background (Scmidt and Flege, 1996; Syndral,1996). The non-speaker related factors include place of articulation, word frequency, phonetic context (Whiteside et al 2004, Neiman et al, 1983).

Invariant properties of articulatory movements have been a dominant issue in studies related to speech motor control. A study of invariant properties has thrown light on motor control processes underlying speech production. A number of studies have indicated differences in articulatory movement across a variety of context and other influences (Kent, 1970; Fujimura, 1986). One such influence is the rate of speech. Speaking rate is typically defined as the number of output syllables or words per unit time. Speaking rate is thought to affect most, though not soley, coarticulation and spectral temporal reduction (Lindblom, 1963). In turn speaking rate changes the VOT values (Miller et al, 1986; Summerfield, 1975).

Miller et al (1986) measured VOT and overall syllable duration for /bi/ and /pi/, produced across a wide range of speaking rates. They found that as rate was slowed such that syllable duration increased, VOT values become systematically longer. As a consequence, the VOT value that optimally differentiated the VOT distributions for /b/ and /p/ also moved towards longer values. They concluded that the phonetic category boundary along a VOT series shift toward longer VOT values as the syllables within the series become longer (Green & Miller, 1985; Summerfield, 1981). The follow up study by Volaitis & Miller (1992) also measured the VOT for both voiceless (p, t, k) and voiced (b, d, g) stop consonants for syllable initial consonant at different speaking rates. Their results revealed that VOT systematically increased with an increase in syllable duration i.e., speaking rate. Their results replicate the findings of Miller (1986) for labial consonants and extend them to alveolar and velar consonants.

Thus, VOT is found to be negatively correlated with speaking rate and the correlation is highly significant, especially for voiceless stop consonants (Kessinger and Blumstein, 1998; Volatis and Miller, 1992). However, no significant changes occurred in the negative or short-lag VOT's of voiced plosives (Volatis and Miller, 1992). As speaker slows down the speaking rate, all the phonetic segments will be stretched and therefore they should all show an increase in duration.

Many of the previous studies that have examined male-female differences in VOT have not controlled for speech tempo, and simply asked the participants to read the stimuli at a comfortable or natural rate (Ryalls et al., 1997; Swartz, 1978; Sweeting & Baken, 1982; Whiteside & Irving, 1997, 1998; Wadnerkar et al, 2006). Allen et al. (2003) recognized this factor and developed an algorithm for determining the magnitude of the effect of individual speech tempo differences. They determined that approximately 10% of the variability in VOT resulted from individual differences even when speakers were directed to speak slowly or quickly. Hence a follow up study was done (Allen et al, 2003) controlling for speech tempo and it was found that there was no difference existing between male and female participants. Thus, they concluded that previous findings ascribed to sex-based differences may, in fact, be the result of differences in speech tempo.

Robb et al (2005) compared the VOT's and syllable durations in the phrases produced by the participants in their study. They found that the females produced significantly longer VOT's for voiceless stops but no significant differences between the males and females in syllable duration. Thus they concluded that the differences that they observed were sex-based rather than speech tempo-based. From the Robb et al. findings, future studies comparing males and female VOT's in phrases and sentences need to either fix speech tempo or provide an acoustic measure that will allow the researcher to separate the effects of speech tempo and VOT. It should be noted that previous studies comparing the speech tempo of males and females have produced differing results (Byrd, 1994; Lutz & Mallard, 1986; Robb, Maclagan, & Chen, 2004; Walker, 1998).

The literature reviewed suggests that studies are mostly in English and therefore, do not provide sufficient data for generalization to other languages. There is a need for studies with larger number of subjects across various age groups in different languages. Such information can further our understanding of speech timing. In this context the present study aimed to investigate the changes in VOT for voiced and voiceless stop consonants in Kannada language (a south Indian Dravidian Language) with change in speaking rates and also to note gender differences for VOT across different speaking rates.

Method

Subjects

A total of 20 young adults (10 males and 10 females) participated in the study. The age ranged from 18 to 24 years (mean age 22 years) and all the subjects were native Kannada speakers. All participants were non-smokers and reported no history of neurological, vascular or motor impairment which would affect articulation, phonation and/or respiration. None of the participants reported that they had received any kind of formal vocal training. A speech language pathologist while recording determined that the speech and language skills of all the participants were within normal limits for young adults.

Test materials

For the purpose of this study a set of six phrases was constructed, which were meaningful and they had voiceless (/p/, /t/, /k/) & voiced (/b/, /d/ & /g/) sounds in initial position of the word. Phrases used were /pa:ka mathu pa:kashale/, /ba:le mathu ba:le kamba/, /ta:ta mathu ta:tabirla/, /da:ku mathu da:kumaara/, / ka:ge mathu ka:gemara/, and /ga:li mathu ga:lipata/.

Procedure

Subjects were asked to read the phrases at comfortable pitch and loudness level at five different rates which was set in the visual metronome. The flashlight from a metronome was used to help the participants to maintain the speaking rate. The tempo of the phrases uttered was controlled by using metronome set at 80bpm, 160bpm and 240 bpm. The subjects were asked to match the tempo of production with that of the metronome flashing light at five different rates. The participant's speech samples were recorded through the Sony cardiode microphone (F-V120), which was connected directly to the computer and tempo was controlled by using the Metronome Plus (2nd edition, 2.0.0.1, M & M systems Germany). Hence these samples were recorded live and the same was loaded onto the PRAAT software (Paul & David 2008, version 5.0.27) and stored on computer hard disk. The recordings were done digitally and sampled at 16K Hz, 12 bit quantization. All the recordings were obtained with the microphone position fixed at a constant distance of 10-12 cm from the corner of the mouth.

Before speech sample collection, a practice session was given for each subject to set the pitch, loudness and tempo. The speech samples were collected first for phrase starting with /p/ stop consonant at 80bpm which was set in the metronome, three trials were recorded. And then tempo was set at 160bpm, the subject was asked to produce the phrase at this speed for stop consonant /p/. Similarly, 240 bpm was set in succession and the subject was asked to produce phrases at this tempo respectively and was recorded. The procedure which was followed for /p/ stop consonant phrase was carried out for other remaining stop consonant phrases- /t/, /k/, /b/, /d/ and /g/.

An allegro rate of 160 beats/min, (approximately, 3.5 syllables/sec) was chosen because it occurs within the normal speaking and reading rate of healthy young adults (Crystal & House, 1988; 1990; Miller, Grosjean & Lomanto 1984; Walker, 1988). This phrase production rate was chosen as it is within the range of typical speaking rates used by young adults and was easily accomplished by all the participants in the study. 160 beats/min was taken as reference rate of speech, so one level below and above this rate at the same interval of 80 bpm was selected to analyze voice onset time i.e., at 80 beats/min and 240 beats/min.

Each participant produced six plosives /p/, /t/, /k/, /b/, /d/ and /g/ in 6 phrases, each three times at five different speaking rates. A total of 1080 tokens (6 plosive phrases × 3 repetitions× 3 speaking rates× 20 participants) were acoustically analyzed for VOT. VOT was measured by visually inspecting the spectrograms in the PRAAT software. VOT was measured by placing a time marker at the onset of the noise burst of each stop and another marker at the onset of steady state vocal fold vibration. Steady state of vocal fold vibration was determined by visualizing the first vertical striation in the second formant of the vowel following the each stop. The PRAAT software display spectrogram, in terms of time denoted in milli seconds along the horizontal axis. This allowed for direct measurement of the time between the markers, and thus VOT. All VOT measures were carried out similarly for all stop consonants (/p/, /t/, /k/, /b/, /d/ and /g/) selected for the study.

Statistical Analysis

The data was subjected to statistical analysis. The mean and standard deviation values of voice onset time at different speaking rates are calculated and tabulated for voiced and voiceless consonants. SPSS (Version 10) was used for statistical analysis. Paired t-test was used on the data to determine the significant difference between the speaking rates and gender.

Results

The data was analyzed to note the effect of different rate of speech on VOT for voiceless and voiced stop consonants. The results are discussed under the following main headings:

1. The voice onset time at different rate of speech for voiceless stop consonants.
2. The voice onset time at different rate of speech for voiced stop consonants.

1. The voice onset time at different rate of speech for voiceless stop consonants:

The mean and standard deviation (SD) values in milliseconds for voiceless stop consonants for males and females across speaking rates are shown in the Table 1. The mean and SD values obtained across different rate of speech indicates as the rate of speech increases the VOT values reduces. More differences was noted at slower rates i.e., 80bpm but the differences were less at faster rate i.e., 160*240bpm. The mean and SD VOT values obtained were longer for males compared to females for all voiceless stops at all rates of speech this difference was consistent.

Voiceless consonants	**/p/**				**/t/**				**/k/**			
Rate of speech	Males		Females		Males		Females		Males		Females	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
80 bpm	27.7	3.2	19.6	2.0	33.5	2.3	18.5	2.0	32.8	7.4	19.5	2.5
160 bpm	21.6	3.5	14.5	1.0	22.8	4.6	12.8	1.9	25.8	1.7	13.3	1.7
240 bpm	17.4	2.2	12.2	1.2	19.2	3.4	11.1	1.7	20.7	3.7	11.1	1.2

Table 1: Showing the mean and standard deviation (SD) values of voice onset time in milliseconds for voiceless stops across males and females for different speaking rates.

As it is also evident from Table 2, statistically significant differences ($p<0.05$) were obtained for voiceless stops and gender across all the speaking rates. The study results are in consonance with the findings of earlier studies by Miller et al (1986) and Volaitis & Miller (1992) which showed significant difference at speaking rates for voiceless stop consonants.

Voiceless stops	Rate of speech		80 bpm		160bpm		240bpm	
			M	F	M	F	M	F
/p/	80	M		S	S	S	S	S
		F			S	S	S	S
	160	M				S	S	S
		F					S	S
	240	M						S
		F						
/t/	80	M		S	S	S	S	S
		F			S	S	S	S
	160	M				S	S	S
		F					S	S
	240	M						S
		F						
/k/	80	M		S	S	S	S	S
		F			S	S	S	S
	160	M				S	S	S
		F					S	S
	240	M						S
		F						

Table 2: Shows the significance across different rate of speech for voiceless stops- /p/, /t/ and /k/ for males and females [M-male, F-Female, S-significant difference ($p<0.05$), NS – No significant difference ($p>0.05$)].

2. The voice onset time at different rate of speech for voiced stop consonants:

Table 3 shows the mean and SD values in milliseconds for voiced stop consonants for males and females across different rate of speech. The mean values obtained across different rates of speech shows that as rate of speech increases the VOT values reduces. Similar differences were noted at slower rate (80bpm) and at faster rate (240bpm). But the SD values were not consistent across rates.

The mean VOT values obtained were longer for males compared to females for all voiced stops at all rates of speech this difference was consistent and SD values were more for female subjects then the male subjects.

Voiced consonant	**/b/**				**/d/**				**/g/**			
Rate of speech	Males		Females		Males		Females		Males		Females	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
80 bpm	-84.4	6.7	-76.2	7.8	-79.5	8.47	-74.2	12.3	-84.5	4.5	-76.5	6.8
160 bpm	-64.3	8.1	-55.3	8.4	-69.8	10.6	-58.3	7.1	-64.1	6.6	-53.6	6.9
240 bpm	-50.3	3.3	-40.0	8.2	-52.1	8.9	-40.5	9.1	-50.8	8.8	-38.7	10.5

Table 3: Showing the mean and standard deviation (SD) values of voice onset time in milliseconds for voiced stops across males and females for different speaking rates.

Statistically by using paired sample t- test on the data, a significant difference was noted for voiced stop consonants and gender across different rates of speech. From Table 4 it can be seen that significant difference obtained between males and females as well as for voiced consonants across three different rates of speech ($p<0.05$) excepting for consonant /d/ between 80 and 160 bpm . In contrast to the present study Miller et al (1986) and Volaitis & Miller (1992) observed no significant difference for voiced stop consonants where the difference is smaller.

Voiceless stops	**Rate of speech**		**80 bpm**		**160bpm**		**240bpm**	
			M	**F**	**M**	**F**	**M**	**F**
/b/	80	M		S	S	S	S	S
		F			S	S	S	S
	160	M				S	S	S
		F					S	S
	240	M						S
		F						
/d/	80	M		S	NS	S	S	S
		F			S	S	S	S
	160	M				S	S	S
		F					S	S
	240	M						S
		F						
/g/	80	M		S	S	S	S	S
		F			S	S	S	S

	160	M				S	S	S
		F					S	S
	240	M						S
		F						

Table 4: Shows the significance across different rate of speech for voiceless stops- /b/, /d/ and /g/ for males and females [S-significant difference ($p<0.05$), NS – No significant difference ($p>0.05$)].

Discussion

As evident from results, there was a statistical significant difference for voiceless and voiced stop consonants across different speaking rates. The study results were similar to the findings of earlier studies by Miller et al (1986) and Volaitis & Miller (1992) atleast for voiceless stop consonants. Their results revealed that voiceless stop consonants appear to change considerably more with rate than Voiced counterparts. But contrastingly in the present study the results were noted for both voiceless and voiced stop consonants. The changes across rates may be explained on the basis of the hypothesis of multiple sub movements of the articulators (Adams, Weismer & Kent, 1993). According to these investigators, a single movement is constituted by elements of movement subroutines and the blending of these movements. Subroutines in the spatial and temporal dimension determine the actual or total movement. This in turn contributes to precise articulation. The precise in articulation in turn gets reflected in terms of changes in the VOT.

It may be inferred that the movement subroutines executed in the spatial and temporal discussions of speech are more accurate at 160bpm and 240bpm when compared to that of 80bpm. This is reflected in the duration of VOT. There was a statistical significant difference between the 160bpm and 240bpm. Even the mean values were statistical significant the deviations were less at faster rates compared to that of VOT's at slower rates of 80bpm. Thus the values are less clustered at faster rates. This probably suggests that the force dynamics of the articulators are tuned in such a way that it is more or less executed in a subconscious, automatic manner when compared to that of the slower rates of speech. A change from a normal speaking rate (160bpm) of speech may bring about a relatively more conscious control of movement subroutines, which in turn gets reflected in form of changes in force dynamics of the articulators. Thus the effect of longer VOT values at slower rates (80bpm) may probably be explained using this reasoning.

Also to understand this difference, the results were inspected with reference to the 'Target undershoot Model' by Lindblom (1963). This model predicts more spectral, temporal reduction in fast rate of speech than in normal rate of speech. Hence the articulators are said to miss the ideal target position by 'undershoot'. This will explain the reducing VOT values moving from slow to fast rate of speech.

Another important finding which was obtained from the results was a statistical significant difference between genders for both voiceless and voiced stop consonants when rate of speech was controlled. In previous studies, that have examined male and female differences in VOT have not controlled speech rate (Ryalls et al., 1997; Swartz, 1978; Sweeting & Baken, 1982; Whiteside & Irving, 1997, 1998; Wadnerkar et al, 2006). Thus when speech rate was controlled there was gender differences for VOT.

The analysis showed that VOT for productions of voiced and voiceless plosives, male speaker produced voiced and voiceless stop consonants with a longer VOT interval than female speakers. The results were consistent with data from previous studies by Allen, Miller and De Steno (1985). Contrastingly, results were obtained from Robb et al (2005) found that the female produced significantly longer VOT's for voiceless stops. Thus they concluded that the differences that they observed were sex-based rather than speech tempo-based.

Conclusions

Rate of speech is one of the factors which contribute to the variance in the articulatory movement. The western and European studies have probed into acoustical changes in VOT with speaking rate. From the results of the present study, it can be concluded that changes in acoustical duration of VOT for speaking rate result from changes primarily in dynamic specification of gesture of articulators. Hence there is reduction in the VOT from 80bpm to 240bpm for both voiced and voiceless stop consonants for Kannada language and consistent with the previous studies and also gender differences were noted for VOT when speaking rate is controlled.

Clinically the researchers need to watch for the variations in VOT of the individuals occurring due to the influence of speaking rate. There is some need to note the gender differences for VOT; one should control rate of speech. The results need to be supplemented with kinematic analysis which will enrich our knowledge of speech motor control. Also, the obtained data can be employed in developing algorithm for speech synthesis and speaker identification. Further studies can be done with increased large number of participants using other speaking rates.

References

Adams, S. G., Weismer, G & Kent, R. D. (1993). Speaking rate and speech movement velocity profiles. *Journal of Speech and Hearing Research*, 36; 41-54.

Allen,J.S., Miller,J., & De Steno,D.(2003). Individual talker differences in voice onset time. *Journal of the Acoustical Society of America*.113, 544-552.

Gay, T., Ushijima,T., Hirose, H, & Cooper, F.S.(1974). Effect of speaking rate on labial consonant-vowel articulation. *Journal of Phonetics*,2;47-63.

Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, 26, 117-128.

Kessinger,R., & Blumstein,S.(1997). Effects of speaking rate on voice onset time in Thai, French, and English. *Journal of Phonetics*, 25, 143-168.

Kidd, G. R. (1989). Articulatory-rate context effects in phoneme identification. *Journal of Experimental Psychology: Human Perception & Performance*, 15, 736-748.

Luce, P., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of the Acoustical Society of America*, 78, 1949-1957.

Miller, J. L., & Baer, T. (1983). Some effects of speaking rate on the production of /b/ and /w/. *Journal of the Acoustical Society of America*, 73, 1751-1755.

Miller,J.L., Green, K.P., & Reeves ,A,. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, 43, 106-115.

Munhell, K. G & Lofgvist, A.(1992). Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics*, 20;111-126.

Ravanan, M.P. (1996). Cross linguistic study of some temporal parameters. Dissertation submitted in part fulfilment of Masters degree in speech and hearing to the university of Mysore.

Ravishankar, K.S.(1981).VOT in children. Dissertation submitted in part fulfilment of Masters degree in Speech and Hearing to the University of Mysore.

Robb,M., Gilbert,H., & Lerman , J.(2005). Influence of gender and environmental setting on VOT. *Folia Phoniatrica and Logopaedica*, 57, 125-133.

Summerfield.GA., & Haggard.M.P., (1972). Speech rate effects in the perception of voicing. Speech synthesis & Perception, psychological lab, University of Cambridge, England.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, 92, 723-735.

Wayland, S. C., Miller, J. L., & Volaitis, L. E. (1994). The influence of sentential speaking rate on the internal structure of phonetic categories. *Journal of the Acoustical Society of America,* 89, 2694-2701.

Manjunath.N., M.Sc. (Speech and Hearing) Candidate
Department of Speech-Language Pathology
J.S.S. Institute of Speech and Hearing
Mysore
Karnataka, India
man_n94@yahoo.com

Sneha Mareen Varghese, M.Sc. (Speech and Hearing) Candidate
Department of Speech-Language Pathology
J.S.S. Institute of Speech and Hearing
Mysore
Karnataka, India
snehamv2002@yahoo.com

Narasimhan S.V., M.Sc. (Speech & Hearing) **Corresponding Author**
Department of Speech-Language Pathology
J.S.S. Institute of Speech and Hearing
Mysore, Karnataka, India
narasimhanslp@gmail.com