TRANSFORMATION OF VOICE ONSET TIME IN HEARING AID PROCESSED SPEECH AND ITS RELATION WITH NUMBER OF CHANNELS IN HEARING AIDS

[1]Ajish K. Abraham

Abstract

Modification of temporal features of a speech signal as it passes through the processing stages in a digital hearing aid may affect the speech recognition and speech perception. Transformation of voice onset time (VOT) may be one factor which affects the intelligibility of hearing aid processed speech and thereby affecting the factor of clarity of sound. The contribution of VOT transformation in the degradation in speech quality with increase in the number of channels of a hearing aid needs to be investigated. The objectives of the study are 1) to acoustically analyze and extract VOT from the speech signal at the input and output of a 2, 4, 6 and 16 channel digital hearing aid, 2) to estimate the transformation of VOT in a 2, 4, 6 and 16 channel hearing aid processed speech and 3) to investigate the effect of number of channels of hearing aids on this transformation. The stimulus consisted of three meaningful voices cognate word pairs, such as BEES / PEAS, DIME / TIME and GOAT / COAT produced by a male adult (native Kannada) speaker. VOT values were found out for the UN processed stimuli and for the processed stimuli from four behind the Ear hearing aids, by using speech analysis software and then the corresponding changes in VOT was noted. Perceptual evaluation was done by six judges with normal hearing. Results showed a reduction in lead VOT by 16% in hearing aid processed speech which would result in degradation of speech intelligibility by about 11% and a reduction in lag VOT by 18.4% leading to a decrease in percentage intelligibility by 19.4%. No significant difference between the percentage reduction in lag VOT and lead VOT was observed across processed speech of hearing aids with different channels.

Keywords: *Voice Onset Time, lag VOT, and lead VOT, Percentage intelligibility, Just Noticeable Difference*

Hearing aids, the auditory rehabilitation aid used by majority of the people with hearing impairment, have undergone a technological transformation since the mid – 1990s. Followed by the emergence of the first digital hearing aid in the late 1990s, the pace of technology developments got faster and led to many innovations. Thus the present day hearing aids have become complex digital signal processors, with a wide range of programmable and adaptive features.

A survey (Kochkin, 2005) of more than1, 500 users of hearing instruments measured only 71% overall satisfaction with instruments that are 0 – 5 years old, and 78% with 1 year old instruments. Some of the factors that relate most with customer satisfaction in this survey were clarity of sound (p=0.72) and natural sounding (p=0.68), where 'p' denotes correlation with overall hearing instrument satisfaction. These two factors which are ranked as 2nd and 5th respectively can be attributed to speech discrimination and percentage intelligibility. Nataraja, Savithri, Sreedevi and Sangeetha, (2000) estimated the changes in speech intelligibility as a result of modifying the temporal, spectral and prosodic parameters in the speech of the children with hearing impairment. Results indicated that modification of duration of the temporal parameters resulted in maximum improvement in intelligibility. Hence, modification of duration of temporal parameters of a speech signal as it passes through the processing stages in a digital hearing aid may affect the speech recognition and speech perception and this requires to be investigated.

Momentary blockage of the vocal tract due to an articulatory occlusion is a characteristic of stop consonants /p b t d k g/. Duration of the blockage usually varies between 50 to 100 ms. The blockage is then released with a burst of air. After the burst and a brief interval, next is the onset of vocal fold vibrations (voicing) for the vowel which follows. Voice onset time (VOT) is specified as the time interval between two events that is, between the release of a complete articulatory occlusion and the onset of quasi-periodical vocal fold vibration.

VOT is a relevant perceptual cue to distinguish between voiced and voiceless and between aspirated and unaspirated stops. VOT measures are commonly used in research and clinical

[1]Reader in Electronics & Acoustics, All India Institute of Speech and Hearing (AIISH), Mysore-06, Email: ajish68@aiishmysore.in

practice, and have been used in the developmental maturation of neuromotor coordination, (DiSimoni, 1974) speech production of hearing-impaired talkers (Monsen, 1976) and assessment and treatment of individuals who stutter (Borden, Baer & Kenney, 1985). VOT is considered a major cue for differentiation of prevocalic stops along the voicing dimension (Lisker and Abramson, 1964, 1971, Abramson and Lisker, 1965).

Eguchi & Hirsh (1969) and Kent (1976) have opined that the distinctive acoustic cue VOT is helpful to assess the general process of motor skill acquisition. Research of speech development in children has given due importance to VOT as an important temporal feature. Zlatin (1974) reported that VOT is a primary cue for differentiation of homorganic stop consonants. Hansen, Gray and Kim (2010) reported that, the important temporal feature VOT is often overlooked in perception as well as recognition of speech.

The study conducted by Nataraja et al., (2000) considered altering the duration of the temporal parameters in the speech of the children with hearing impairment and evaluated the improvement in intelligibility. In single cue condition, in the initial position, burst duration and VOT were transformed; in the medial position, consonant duration, closure duration, and transition duration of the second formant and preceding and following durations were transformed. Results show that among all temporal parameters whose transformation resulted in improvement of intelligibility and percentage discrimination of the speech of children with hearing impairment, VOT was the one with maximum effect. Hence, alteration of VOT may be one factor which affects the intelligibility of hearing aid processed speech and thereby affecting the factor of clarity of sound. Hence, it needs to be investigated whether VOT is getting affected when the speech signal undergo signal processing in hearing aids.

Keidser and Grant (2001) reported no significant difference in speech recognition in one, two or four channel hearing aids. Abhaykumar (2010) compared the speech identification scores with single channel, multichannel and channel free hearing aids in quiet condition. Results indicated that there is no statistically significant difference between the scores. Volker Honmann and Birger Kollmeir (1995) conducted an experiment using an algorithm for amplification with 23 independently compressed frequency channels to find out the effect of multichannel dynamic compression on speech intelligibility. They observed distinct degradation in speech quality and virtually no decrease in speech intelligibility for high signal to noise ratios when tested with normal hearing listeners. Proportional change in VOT with increase in the number of channels of a hearing aid transformation may contribute degradation in speech quality which needs to be investigated.

It is essentially required to increase the overall satisfaction of users of hearing instruments from the measured level of 71 to 78% (Kochkin, 2005). This would be possible, if the factors that relate most with customer satisfaction namely, clarity of sound and natural sounding are enhanced. This study would help to find out whether there are any changes in VOT in a hearing aid processed speech and if so whether these changes have got any impact on these factors with specific reference to increase in the number of channels. Results of this study would provide technical inputs to the designers of signal processing algorithm to make the required changes in the software design of hearing aids to enhance the clarity of sound and natural sounding in the processed speech at the hearing aid output.

The objectives of the study were 1) to acoustically analyze and extract VOT from the speech signal at the input and output of a 2, 4, 6 and 16 channel digital hearing aid, 2) to estimate the transformation of VOT in a 2, 4, 6 and 16 channel hearing aid processed speech and 3) to investigate the effect of number of channels of hearing aids on this transformation.

Method

Stimuli: Three meaningful voice cognate word pairs, such as BEES / PEAS, DIME / TIME and GOAT / COAT were selected for study. This corresponds to the same series of stimuli developed by Lisker and Abramson (1967) for similar studies. The stimuli comprises six plosives in the initial position categorized as unvoiced (velar /k/, dental /t/, and bilabial /p/) and voiced (velar /g/, dental /d/ and bilabial /b/).

Recording: Each of the word in the stimulus list was uttered by an adult (native Kannada) male speaker and was recorded by a condenser microphone (B & K 4189) that was placed at a distance of 15 cm from the mouth of the speaker. The syllables were recorded by the precision sound level meter (B & K 2250) with sound recording software (B & K BZ7226). The recorded samples were digitized at a sampling frequency of 22 kHz and 16 bits / sample and were stored on a PC.

Instrumentation: The signals required for measurement of VOT as well as for analysis are acquired and processed by the set up shown in figure 1. To deliver the recorded stimuli, the corresponding wave file is downloaded to the front end of the analyzer (B & K Pulse 12.0) system. The stimulus is delivered through a leveled response loudspeaker to the hearing aid under test kept in the anechoic test chamber (B & K4232). To rule out any distortion in the delivery path, a dedicated hardware unit (B & K Pulse 12.0) and a power amplifier (B & K 2716C) were included in the stimulus delivery path. Four behind the ear digital hearing aids were used for testing. These hearing aids differed in terms of the number of channels, two, four, six and sixteen, but incorporated similar signal processing strategy. They incorporated wide dynamic range compression. The hearing aids under test were programmed with the first fit algorithm. The stimuli after being processed from all 4 hearing aids were collected by B & K 4189 condenser microphone that is placed inside a 2cc coupler and were recorded by BZ 7226 recording setup. Recorded processed syllables were stored on the PC for further analysis.

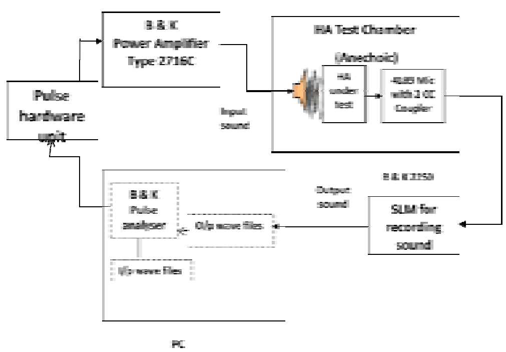

Figure 1: *Instrumentation for signal acquisition, measurement and analyses*

Test environment: Measurement as well as stimulus (both processed and unprocessed) recording was carried out in a sound proof room built as per ANSI S3.1 - 1991 standards.

Subjects: Six post graduate students of AIISH, Mysore with normal hearing, participated as judges in the study. The judge did not know in advance, the word pair presented for judgment. Participants were asked to assess whether two sounds were the same or different in a fixed AX discrimination task. For each input output stimulus pair, a score of 1 was given for "different" and 0 for "same". Maximum score for this test was 'four' for each category of stimuli because four numbers of hearing aids were included in the study.

Procedure: Hearing aid processed speech from all four hearing aids with six input stimuli to each of them were recorded. The input along with its output recordings were played back to six judges in random order. The judges were asked to perceive whether there is any difference between the input and output sample.

Acoustic Analyses: B & K Pulse Time Edit & Analysis software was used to measure the VOT values from the waveform of the processed / unprocessed stimuli. The interval between the release of stop occlusion and the onset of the following vowel was considered as VOT (Figure 2). The use of a waveform in the measurement of VOT values was found to be more accurate and reliable (Francis & Ciocca, 2003). Time measurement was directly done by the Time Edit and analysis software (B & K). In cases of difficulty in locating the points of measurement on a time scale, spectral analysis was employed (B & K Pulse 12.0).

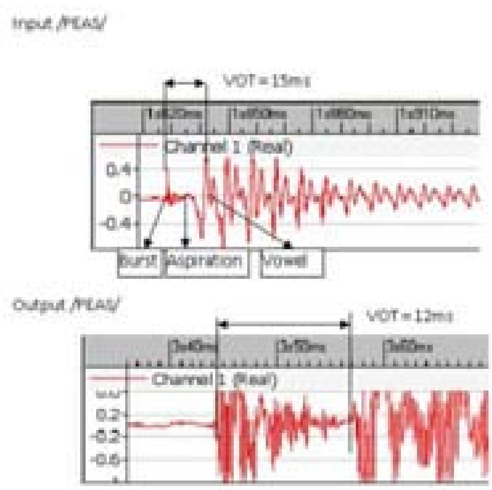

Figure 2: *Illustration of measurement of VOT*

Statistical analysis: SPSS (Version 18) software was used to carry out the statistical analysis. Friedman test was administered to find the effect of number of channels on transformation of VOT.

Reliability measurements: Reliability between judges who served as subjects for the study was tested using reliability coefficient Cronbach's alpha and it showed sufficient reliability between judges (0.97).

Results and Discussion

Transformation in Lead VOT: Table 1 shows a reduction in lead VOT for all categories of hearing aids. Table 3 shows that percentage reduction in lead VOT was in the range of 14 to

20 with a mean of 17.67 in 2 channel hearing aid. In 4 channel hearing aid, the range was between 13 and 18 with a mean of 14.67. The range was from 13 to 16 with a mean of 15.67 in six channel hearing aid and was from 13 to 20 with a mean of 15.67 in sixteen channel hearing aid. Nataraja et al., (2000) reported that lead VOT in the speech of normal hearing group was significantly longer than that in hearing impaired children. They also reported that % improvement in speech intelligibility on enhancing the lead of VOT by 50% on an average in the speech of hearing impaired children was 33%. Considering this, it can be inferred that a reduction in lead VOT by 16% (table 3) in hearing aid processed speech would result in degradation of speech intelligibility by about 11%.

Table 1: *Reduction in VOT (%) for voiced plosives in various places of articulation*

Stimuli	2 channel hearing aid		4 channel hearing aid		6 channel hearing aid		16 channel hearing aid	
	Change in VOT (ms)	% change	Change in VOT (ms)	% change	Change in VOT (ms)	% change	Change in VOT (ms)	% change
Velar	15	20	10	13	12	16	15	20
Dental	15	14	20	18	20	18	15	14
Bilabial	15	19	10	13	10	13	10	13

Table 2: *Reduction in VOT (%) for un-voiced plosives in various places of articulation*

Stimuli	2 channel hearing aid		4 channel hearing aid		6 channel hearing aid		16 channel hearing aid	
	Change in VOT (ms)	% change	Change in VOT (ms)	% change	Change in VOT (ms)	% change	Change in VOT (ms)	% change
Velar	2	7	4	13	4	13	3	10
Dental	3	23	3	23	3	23	3	23
Bilabial	4	27	5	33	3	20	4	27

Transformation in lag VOT: Table 2 shows a reduction in lag VOT for all categories of hearing aids. Table 4 shows that percentage reduction in lag VOT was in the range of 14 to 20 with a mean of 17.67 in 2 channel hearing aid. In 4 channel hearing aid, the range was between 13 and 18 with a mean of 14.67. The range was from 13 to 16 with a mean of 15.67 in six channel hearing aid and was from 13 to 20 with a mean of 15.67 in sixteen channel hearing aid. Along the same lines on the discussion done in the case of lead VOT, Nataraja et al., (2000) reported that lag VOT in the speech of normal hearing group was significantly longer when compared with that in hearing impaired children. Percentage improvement in speech intelligibility is reported to be 45% when the lag VOT was enhanced by 43%. Referring to table 4, mean % reduction in lag VOT for all hearing aids put together is 18.4%. This may lead to a decrease in percentage intelligibility by 19.4% as per the figures arrived at by Nataraja et al., (2000).

Effect of number of channels on transformation in VOT: Table 3 & 4 shows the variation of percentage reduction in VOT across 2 channel, 4 channel, 6 channel and 16 channel hearing aids. Friedman test was administered to ascertain the significance of this. Results reveal that there is no significant difference between the percentage reduction in lag VOT and lead VOT across processed speech of hearing aids with different channels. (χ^2 (3) = 0.913, p>0.05 for lead VOT and χ^2 (3) = 3, p>0.05 for lag VOT). These results are in accordance with the findings of previous studies conducted by Keidser et al., (2001), Abhaykumar (2010) and Volker et al., (1995). These studies reported that speech intelligibility has no relation with the number of channels of hearing aids for inputs with high signal to noise ratio. Present study indicates that there is no effect of number of channels on VOT transformation and hence on speech intelligibility.

Table 3: *Mean, SD and Range of percentage reduction in lead VOT*

	2 channel hearing aid	4 channel hearing aid	6 channel hearing aid	16 channel hearing aid
Mean	17.67	14.67	15.67	15.67
SD	3.21	2.89	2.52	3.79
Range	5	6	5	7

Table 4: *Mean SD and Range of percentage reduction in lag VOT.*

	2 channel hearing aid	4 channel hearing aid	6 channel hearing aid	16 channel hearing aid
Mean	19	23	18.67	20
SD	10.58	10	5.13	8.88
Range	20	16	10	17

Perceptual evaluation: Table 5 shows the score of words perceived differently by judges when they were presented with 4 pairs of input output stimuli, each pair corresponding to input and output of one category of hearing aid. Reliability coefficient Cronbah's alpha showed sufficient reliability between judges (0.97).

Table 5: *Score of words perceived differently by judges* (V=voiced, U=unvoiced)

	Velar		Dental		Bilabial	
Judge	V	U	V	U	V	U
1	2	0	2	1	0	0
2	1	0	2	0	0	1
3	1	0	2	0	0	0
4	2	0	2	1	0	1
5	1	0	2	1	0	1
6	2	0	2	1	0	0
Mean score	1.5 (37.5 %)	0	2 (50 %)	0.66 (16.5 %)	0	0.5 (12.5 %)

The scores obtained by the judges are in accordance with the report of the study conducted by Ajith Kumar (2006). Ajith Kumar reported the values of VOT changes for each category of stop consonants to produce just noticeable difference (JND) for people with normal hearing. For unvoiced velar, the JND has a Mean of 27.9 ms and SD of 11.3 ms and this explains `0' score by the judges, as the change in VOT obtained in our study is in the range of 2 to 4 ms. In voiced velar, the JND has a mean of 32.1 ms and SD of 11.1 ms. As VOT transformation for voiced velar in our case ranges between 12 to 15 ms, the mean score is 37.5%.

In voiced dental, the JND has a mean of 35.1 ms & SD 9.9 ms, whereas our studies showed a range of 15 to 20 ms and hence the mean score is 50%. For unvoiced dental the JND has a mean of 15.5 ms with SD of 8 ms and this explains the score of only 16.5% as our results show 3 ms variation in VOT. For voiced bilabials the score is 0, as our result doesn't reach the vicinity of the JND (Mean 41.5 ms and SD 10.4 ms). Unvoiced bilabials had a VOT change of 3 to 5 ms whereas the JND has a mean of 13.7 ms with SD 5.5 ms and hence a score of 12.5%. Thus the scores obtained in perceptual evaluation tallies with the findings of objective analysis.

Conclusions

The present study measured VOT from the speech signal at the input and output of a 2, 4, 6 and 16 channel hearing aid and estimated the transformation of VOT in hearing aid processed speech. Results indicated a reduction in VOT which may lead to degradation in speech intelligibility which is supported by the results of perceptual evaluation by the judges. The study also investigated the effect of number of channels on the VOT transformation but could not find any significant relation. This indicates a need to investigate the transformation of other temporal parameters of the speech signal across the number of channels to explain the reported degradation of speech quality in proportion to the number of channels. The extent of degradation in VOT can be used as an indicator for hearing aid output quality. Results of this study indicate that, the signal processing algorithm in hearing aids needs to be modified to reduce the degradation in VOT. This step would result in enhancement of clarity of sound and natural sounding at the output of the hearing aid. Ultimately this would increase the overall satisfaction of users of hearing aids.

Acknowledgements

The author thanks Director, AIISH, Mysore for permitting to carry out the study. The author also thanks Dr. M.S. Vasanthalakshmi, Lecturer in Biostatistics, Department of Speech Language Pathology, for the help in statistical analysis. Sincere thanks to Dr. N. Sreedevi, Lecturer, Department of Speech Language Sciences, S.R. Sreenivasa and V.T. Kalaiselvi, Junior Technical Officers, Department of Electronics for the immense support.

References

Abramson, A.S. and Lisker, L. (1965). Voice onset Time in stop consonants: Acoustic Analysis and Synthesis. *5th International Congress of Acoustics, Liege, Belgium*.

Abhay Kumar, R (2010), Speech Identification with single channel, Multichannel and channels free hearing aids. *Student Research at AIISH Mysore,* V: 2006-07, 1-10.

Ajith Kumar U (2006). Perception of some temporal parameters of speech in individuals with auditory dys-synchrony. *Ph.D. Thesis, University of Mysore, Mysore.*

Borden, G.J., Baer, T, Kenney, M.K. (1985), Onset of voicing in stuttered and fluent utterances. *Journal of Speech Hearing Research*, 28, 363-372.

Disimoni, F.G. (1974). Effect of vowel environment on the duration of consonants in the speech of three, six and nine year old children. *Journal of Acoustic Society of America*, 55, 360-361.

Eguchi, S and Hirsh, I.J. (1969). Development of Speech sounds in children. *Acta Otolaryngological supplement*, 257, 1-51.

Francis, A.L., Ciocca, V (2003). Accuracy and variability of acoustic measures of voicing onset. *Journal of Acoustic Society of America*, 113, 1025-1032.

Hanse John H.L., Gray, Sharmistha S., Kim, Wooil (2010). Automatic voice onset time detection for unvoiced stops (/p/, /t/, /k/) with application to accent classification. *Speech Communication*, 52, 777-789.

Kent, R.D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of speech and Hearing Research* 19, 421-447.

Keider, G and Grant, F (2001). The preferred number of channels (one, two or four) in NAL-NLI prescribed wide dynamic range compression (WDRC) devices. *Ear and Hearing*, 22(6), 516-527.

Kochkin, S. (2005). Customer satisfaction with hearing instruments in the digital age. *The Hearing Journal* 58 (9), 30-43.

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

Lisker, L., & Abramson, A.S (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, 10, 1-28.

Lisker, L, and Abramson, A. (1971). Distinctive features and laryngeal control. *Language*, 47, 767-785.

Monsen, R B. (1976). Normal and reduced phonological space: The production of vowels by a deaf adolescent. *Journal of Phonetics*, 4, 189-198.

Nataraja N.P., Savithri S.R, Sreedevi N, and Sangeetha, M. (2000). *Transformation of the speech of the Hearing Impaired,* Project at AIISH, Mysore funded by Department of Science and Technology, New Delhi.

Volker Hohmann and Birger Kollmier (1995). The effect of multichannel dynamic compression on speech intelligibility. *Journal of Acoustic Society of America*, 97, (2), 1191-1195.

Zlatin, M.A. (1974). Voicing contrast: Perceptual and productive voice onset time characteristics of adults. *Journal of Acoustic Society of America*, 56, 981-994.