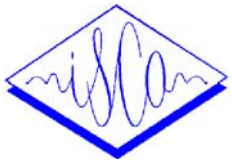

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

CROSS-LANGUAGE ANALYSIS OF VOICE ONSET TIME IN STUTTERED SPEECH

N. Rezaei-Aghbash, S. P. Whiteside & P. A. Cudd
Department of Human Communication Sciences,
University of Sheffield, Sheffield S10 2TA, England, UK.
s.whiteside@sheffield.ac.uk

ABSTRACT

This study compares the VOT measurements between Persian and English stutterers. The speech samples of five adult stutterers (four males and one female) of each language were recorded during a reading task. 98 stuttering events in words beginning with plosives and 34 fluent plosives (i.e., final fluent iterations following the stuttered portions of speech) were identified in the data of both groups. The VOT measurements were obtained from both stuttered as well as fluent productions of speech for voiceless plosives /p, t, k/ and voiced plosives /b, d, ɡ/ respectively.

Keywords: stuttering, voice onset time, cross-language study

10.21437/Eurospeech.1999-259

1. INTRODUCTION

Voice onset time (VOT) is a key temporal feature of speech. Voice onset time is defined as the interval between the release of the plosive (defined by the burst) and the start of the glottal pulsing of the following vowel or diphthong [5, 7]. Compared to normal speech, less attention has been paid to dysfluent speech, and almost none, to cross-language comparisons of stuttered speech.

Voice onset time manifests itself either as a voicing lag (VOT+V) or a voicing lead (VOT-V). Voiceless plosives on the whole are produced with a voicing lag and voiced plosives are produced with voicing lead or short voicing lag. However, it should be noted that, in stuttered speech we may find that the above mentioned characteristics do not necessarily apply. That is, during stuttering, voiceless plosives may be produced with a very short lag as is usually found in voiced plosives. Conversely, voiced plosives may be produced with a long lag which is characteristic of voiceless plosives.

It has been reported that mean VOT values in stuttered speech increase from bilabial to alveolar to velar [8]. Compared with normal speakers it has also been found that VOT values in the speech of stutterers are longer than those found in the speech of normal fluent speakers [1, 2, 6]. For example, Healey and Gutkin [2], examined stutterers' and nonstutterers' fluent voice onset time (VOT) in the speech of ten adult male stutterers and ten adult male non-stutterers. Speakers produced three voiceless plosive consonants /p, t, k/ and three voiced
plosive consonants /b, d, ɡ/ with the vowels /i/, /a/, and /u/ in a carrier phrase, (e.g., Pea is a word; Do is a word). The results showed that, the stutterers had longer mean VOT values than normally fluent speakers did.

This preliminary study investigates the VOT patterns in two groups of stutterers from two different languages, namely Persian and English.

2. METHOD

2.1 Subjects

2.1.1 English subjects

Four males and one female participated in the study. They ranged in age from 18 to 36 years with a mean of 27.6 years. The subjects were all native speakers of English from West Yorkshire, England who were classified as mild to severe stutterers.

The criteria for rating the severity of stuttering was based on SPM (Syllable Spoken per Minute) scores, [4]. According to this criteria, SPM score for mild stuttering is between 128-167 SPM, moderate stuttering is between 107-134 SPM, and severe stuttering is between 45-122 SPM. It was according to this measure, that the speakers were classified as mild to severe stutterers.

In addition, a severity calculation from 100 words was also performed [3]. A severity rating of 88.7 dysfluencies per 100 words (with a range of 14 to 209) was found for the English subjects.

2.1.2 Persian Subjects

Four males and one female participated in the study. Their age range was 17 to 35 years, with a mean of 24.6 years. The speakers were all native speakers of Persian from Tehran, Iran. The severity of their stuttering according to the SPM score [4] was classified as mild to severe. In addition, a severity rating of 95.8 dysfluencies per 100 words [3] was found with a range of 35 to 262.

2.2 Speech Material and Recordings

2.2.1 English Data

All subjects read an identical reading passage; "Watch out for the weather". The speech samples of 4 adult males stutterers were recorded using an Olympus colour video camera (Model VX-303), while a microphone was placed approximately 25 centimetres from the speakers. For the female subject, the recording was made directly onto audiotape in a quiet room. The video recorded samples of the male subjects were then transferred onto the audiotape (Maxell CrO2 90 min.), using a Panasonic Model AG-6200 video recorder. In order to reduce the noise level in the samples, the audio output of the samples was fed into a soundcraft 6000 mixer and the frequency range adjusted as follows: a) -9 dB low pass filter at 10 kHz shelving, b) +15 dB at 2.5 kHz, c) -15 dB high pass filter at 60 Hz shelving, and d) 100 Hz shelving filter at 12 dB/octave. The output filtered audio samples were recorded on a Denon Model DRM-740 cassette deck with Dolby B noise reduction onto audiotape.

2.2.2 Persian Data

All subjects read an identical passage detailing the Iranian Association for Stutterers. To record their speech, a lapel microphone (Sony ECM-144) was placed approximately 15 centimetres from their mouth, while they were sitting in a semi-sound treated room, using a Panasonic Stereo radio/cassette recorder Model RXFT 530 onto audiotape.

2.3 VOT Measurements

All speech samples were played back using a SONY Professional Walkman and digitised (sampling rate 10kHz) using the KAY CSL. All acoustic analysis was carried out using the KAY CSL. Stuttering events occurring on words beginning with plosives and fluent plosives (i.e., final fluent iterations following the stuttered portions of speech) were isolated for both speaker groups. VOT measurements were obtained for both stuttered as well as fluent productions of the voiceless plosives /p, t, k/ and voiced plosives /b, d, ɡ/ respectively. The stuttering events were identified by listening to recorded speech samples. This process was guided by the use of orthographic and phonetic transcriptions. This procedure was repeated 20 times in order to determine the incidence of stuttering events. In addition, in order to identify the stuttering events, especially in cases of multiple repetitions, the speech sample of each individual stutterer was also examined using the CSL. With this procedure every single stuttering event could be identified more accurately. The stuttering events were then marked on the phonetic text of each subject. During stuttered speech, some plosives are repeated several times by some stutterers in order to produce the target word. In these cases, voice onset time was measured for all of the repeated units (i.e., the dysfluent portion/s) as far as was possible. Here, the repeated units consisted of at least one consonant and one vowel (i.e., syllable), and the final iteration (as the fluent portion/s) of the target word/syllable. In the case of 'a block' where stutterers fail to produce the target word/syllable fluently, the 'blocked' plosives were measured without considering the fluent counterpart. The total number of stuttered plosives for the English and Persian speakers was 47 and 51 respectively, and 10 and 24 fluent plosives for the two groups respectively.

Wide-band spectrograms were used to make VOT measurements. The onset of the plosive burst was located on the spectrogram and marked. A cursor was then placed at the voiced onset of the first formant frequency (F1) in the spectrogram. An automatic reading between the burst and voiced onset of F1 was provided by the CSL. Negative VOT values were excluded, but these amounted to only 1.5% of the data.

3. RESULTS

3.1 English Stutterers

The results of VOT measurements of stuttered and fluent plosives for the English speakers are shown in Tables 1 and 2 respectively, and are compared in Figure 1.

The results of VOT measurements of stuttered and fluent plosives for the Persian speakers are shown in Tables 3 and 4 respectively, and are compared in Figure 2.

VOT (milliseconds)						
Plosive	/p/	/t/	/k/	/b/	/d/	/g/
Mean	22.0	67.3	43.5	13.5	32.1	32
SD.	-	28.5	12.0	11.2	40.0	-

Table 1. Mean and standard deviation (SD) of voice onset time values for the stuttered plosives - English speakers.

VOT (milliseconds)						
Plosive	/p/	/t/	/k/	/b/	/d/	/g/
Mean	15	60	37	7.4	26.5	-
SD.	-	-	-	14.4	20.5	-

Table 2. Mean and standard deviation (SD) of voice onset time values for the final iterations (fluent sounds/syllables) following the stuttered portions - English speakers.

VOT (milliseconds)						
Plosive	/p/	/t/	/k/	/b/	/d/	/g/
Mean	46.7	54.4	83.3	3.9	13.7	21
SD.	8.3	33.8	55.2	4.4	7.6	12.3

Table 3. Mean and standard deviation (SD) of voice onset time values for stuttered plosives - Persian speakers.

VOT (milliseconds)						
Plosive	/p/	/t/	/k/	/b/	/d/	/g/
Mean	65.0	66.5	81.0	4.4	12.7	18.0
SD.	-	22.4	46.4	4.6	4.9	5.7

Table 4. Mean and standard deviation (SD) of voice onset time values for the final iterations (fluent sounds/syllables) following the stuttered portions - Persian speakers.

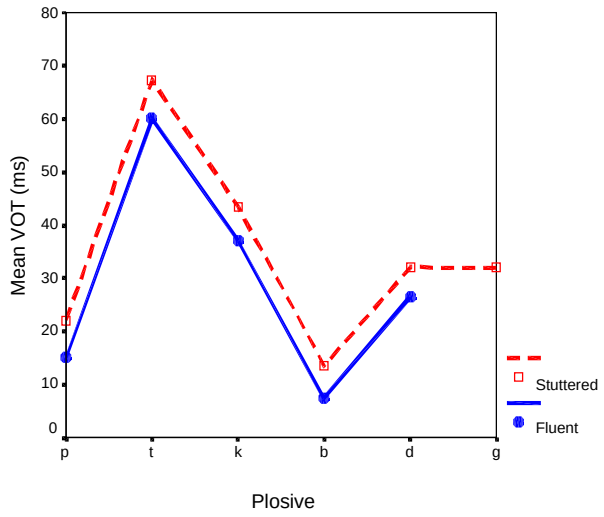

Figure 1. Mean VOT (ms) values for stuttered and fluent plosives - English Speakers.

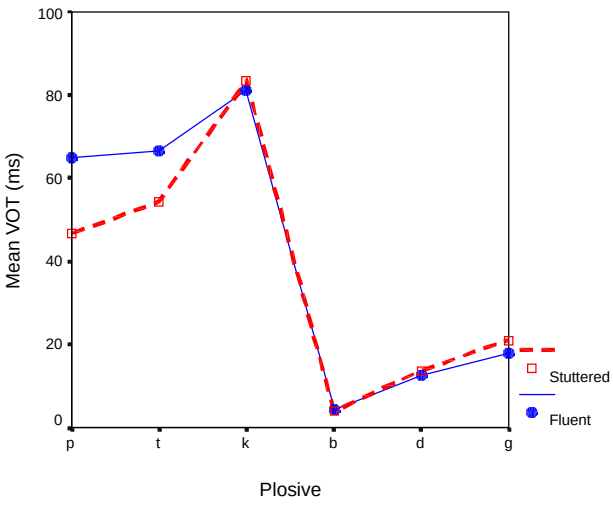

Figure 2. Mean VOT (ms) values for stuttered and fluent plosives - Persian Speakers.

4. DISCUSSION

The data reported here are preliminary and indicate that the English stutterers showed some evidence of lengthened VOT values for stuttered plosives compared to fluent or final fluent iterations. This pattern was replicated across all plosives. This pattern mirrors results from previous studies that report longer VOT values for stutterers compared to non-stutterers [1, 2, 6]. It is suggested that if the fluent or final fluent iterations of the stutterers' productions were compared to fluent productions of a matched group of non-stutterers, a more robust difference might be found between the two speaker groups. In addition it is possible that less clear cut differences were found in this study due to the heterogeneity of the group which included mild to severe stutterers. A more detailed examination of these data in line with the level of severity of stuttering is therefore necessary.

In contrast to the English subjects, the Persian subjects showed some evidence of larger VOT values for the fluent productions of the voiceless plosives /p/ and /t/ when compared to the stuttered productions. The differences between the fluent and stuttered productions of /k b d g/ are less clear. Again the lack of differences maybe due to the heterogeneity of the Persian speakers which like the English group, also included mild to severe stutterers. A more detailed examination of these data in line with the severity of stuttering is also therefore necessary.

The VOT values of voiceless plosives for different places of articulation in English group does not fully replicate those found in previous studies where VOT has been found to be longer as the place of articulation moves from the front to the back of the vocal tract [6, 8]. Although the alveolar plosives had longer VOT values than the bilabials for both the fluent and stuttered plosive events, VOT values for the velar productions were either similar to, or lower than the alveolar values (see Figure 1). It is worth stating at this stage however, that there were no examples of fluent iterations for /g/.

In contrast to the English group the VOT patterns for the Persian group were consistent with previous findings [8]. Both voiced and voiceless plosives showed an increase from bilabial to alveolar to velar for both the stuttered and fluent voiceless plosives.

Inter-group comparisons of both the stuttered and fluent VOT data indicated that the Persian group had larger VOT values for the voiceless plosives /p k/ compared to the English group. Conversely, the voiced plosives for the English group were more aspirated and had larger VOT values compared to those for the Persian group.

The data here are preliminary and suggest that a more detailed examination of the data is needed to compare the VOT patterns of mild to severe stutterers. In addition, further investigation is needed to compare the VOT values of plosive productions of stutterers (both fluent and non-fluent) with those of normal speakers. This work is currently in progress for a group of both English and Persian speakers. The preliminary data reported in this paper together with the data collection in progress would provide more material to make more detailed cross-language comparisons of VOT in the production of both fluent and stuttered speech in English and Persian.

5. ACKNOWLEDGMENTS

We wish to thank the speakers for their participation in this study. We also wish to extend our thanks to the following parties for all their help and assistance: Dr Trudy Stewart, the Rehabilitation Faculty in the Department of Speech and Language Therapy in the Iran University of Medical Sciences, and the Iranian Association of Self-help for Stutterers. (IASS).

6. REFERENCES

[1] Borden, G. J., Kim, D. H. , and Spiegler, K. (1987). Acoustics of Plosive Consonant-Vowel Relationships During Fluent and Stuttered Utterances. *Journal of Fluency Disorders 12, 175-185.*

[2] Healey, E. C., and Gutkin, B. (1984). Analysis of Stutterers' Voice Onset Times and Fundamental Frequency Contours During Fluency. *Journal of Speech and Hearing Research*, 27, 219-225.

[3] Howell, P., and Williams, M. (1992). Acoustic analysis and perception of vowels in children's and teenagers' stuttered speech. *Journal of the Acoustical Society of America*, 91(3), 1697 - 1706.

[4] Ingham R. J. Cordes A. K., Ingham J. C., and Gow M. L. (1995). Identifying the Onset and Offset of Stuttering Events. *Journal of Speech and Hearing Research*, 38, 315- 426.

[5] Lisker, L., and Abramson , A. S. (1964). A Cross-Language Study of Voicing in Initial Plosives: Acoustic Measurements. *Word*: *Journal of Linguistic Circle of New York*, vol. 20, pp. 384- 422.

[6] Metz, D. E., Conture, E. G., and Caruso, A. (1979). Voice Onset Time, Frication, and Aspiration During Stutterers' Fluent Speech. *Journal of Speech and Hearing Research, 22, 627-648.*

[7] Onslow M., Van Doorn J., and Newsman D. (1992). Variability of acoustic segment durations after prolonged- speech treatment for stuttering. *Journal of Speech and Hearing Research*, 35, 529-536.

[8] Viswanath, N., and Joullian, A. (1994). Consequences of the Relation Between VOT and Types of Fragments in Part-Word Repetition. *Proceedings of the 1st World Congress on Fluency Disorders*, Vol. 1, pp. 60-63. August 8-12, 1994, Munich, Germany.