Development of a Gender Difference in Voice Onset Time

Fredrik Karlsson, Elisabeth Zetterholm & Kirk P. H. Sullivan

Department of Philosophy & Linguistics, Umeå University, Sweden

Abstract

This paper investigates the effect of gender on voice onset time distribution at three stages of speech development. Two subject groups consisting of children, aged approximately 3 and 9 years, were compared to adult speakers regarding voice onset time of initial plosives. The results showed significant gender effects in the aspirated plosives in the young subjects that were not present in the plosives produced by adults. It is hypothesised that the effect of gender at the earlier stages of develpment may be due to the differences in airflow intensity and variability.

1. Introduction

In a cross-language study of categories formed by voicing and aspiration, Lisker and Abramson (1964) argued that a single acoustic measurement of the time interval between the release of the consonant and onset of vocal folds vibration contained sufficient information for quantification of the difference between voiced and voiceless plosives. They argued that the measurement incorporated aspects of two other acoustic differences between phonologically voiced and voiceless plosives in English, i.e. aspiration length and overall amplitude. Lisker and Abramson named this measurement *Voice Onset Time* (VOT) and presented data showing a pattern of seemingly discrete groups formed by VOT measurements for plosives from 11 languages with 2–4 categories, matching the groups formed by the phonological features [±voice] and [±asp]. Subsequent perception tests have shown that the VOT measurement agrees well with listener's responses (Lisker, Lieberman, Erickson, Dechovitz, and Mandler 1977; Kluender, Lotto, and Jenison 1995).

For Swedish, which has been described as a language with two phonological voicing categories, voiced plosives ([b],[d] and [g]) have been shown to have a VOT value of <40 ms and voiceless plosives ([p],[t] and [k]) a VOT value of > 40 ms (Fant 1973). Presence of aspiration has been shown to increase VOT of voiceless plosives.

Any feature of speech that is language specific and distinctive has to be acquired by a developing child in order to properly communicate the appropriate contrasts. For voicing, this implies acquisition of adult-like proficiency in production and perception of the acoustic features quantified by the VOT measurement. A number of studies have shown that this is a gradual process which may be influenced by increasingly adult-like articulators, improvements in articulatory proficiency (Kewley-Port and Preston 1974) as well as an evolving model of the language specific requirements for specific speech sounds. Among the trends described in the literature are gradual progression from a unimodal to a bimodal distribution of VOT caused by a gradual increase in perceptually voiceless productions (Kewley-Port and Preston 1974). These processes have all been hypothesised to be caused by articulatory factors such as co-occurrence of articulatory movements (Kewley-Port and Preston 1974). The development of voiced/voiceless distribution modality has been argued to reach adult-like maturity between the ages of 8–11 years (Kent 1976).

1.1. Gender bias in the distribution of VOT

Gender differences in voice onset time have received relatively little attention in the literature. This is surprising considering the differences in the articulators between male and female speakers that have been established in previous research. For instance, data summarised by Titze (1994) shows that the average vocal fold membranous length is 6 mm shorter in female adults compared to male adults. The shorter membranous length in turn increases the possibility of a more rapid closure gesture, which is shown by the higher average f_0 value in female speech compared to male speech. Accordingly, if voice onset time is influenced by the abduction speed of the vocal folds (Kewley-Port and Preston 1974), male and female plosives would be unequally affected by this factor, creating a gender bias.

This hypothesised effect of gender on VOT was investigated by Swartz (1992). Using VOT measurements obtained from the waveform of productions made by adult male and female native speakers of American English, Swartz showed a significant difference in VOT due to gender and also that this difference did not correlate with the higher speaking rate of men compared to women.

1.2. The implication of gender bias in the study of child speech development

Whiteside and Marshall (1998) investigated VOT measurements of labial and alveolar plosives produced by a gender-balanced group of thirty British English speaking children, aged 7, 9 and 11 years. The results showed a difference due to the factor gender for /p/, but not for /b/, /t/ and /d/. Furthermore, there was a significant gender and age interaction effect for /p/ and /t/, where VOT showed a marked fall between the ages 9 and 11.

Proceedings of the 10th Australian International Conference on Speech Science & Technology
Macquarie University, Sydney, December 8 to 10, 2004.

Accepted after full paper review

Furthermore, Whiteside and Marshall (1998) found a main effect of the factor age for bilabial /b/, where 7-year old children produced shorter VOT compared to the other age groups, and alveolar /d/, where the group consisting of 11-year olds produced significantly lower VOT values.

However, Whiteside and Marshall (1998) did not report which, if any, of the investigated children had entered into the voice change period. One may argue that the entry into this period might shift the conditions for voicing, creating a possible confounding factor across groups in the statistical analysis. The interpretation of the results presented by Whiteside and Marshall (1998) is, therefore, complicated by the possibility of the alternative grouping factor 'voice change', which may be present in more than one group but possibly not all, due to the small inter-group interval in age.

This paper examines the effect of gender in VOT productions in three distinct, gender-balanced groups with clear separation in age. The three age groups investigated are 1) young children who have recently established a voicing contrast, 2) older children with the same approximate age as the subjects studied by Whiteside and Marshall (1998), but who have not yet gone through a voice change, and 3) adults.

2. Method

2.1. Subjects

Three gender balanced groups consisting of a total of 24 subjects were recruited for the recordings. The first ('Young') group consisted of four males and four females who had just started producing a voicing contrast. The subjects were recruited from an ongoing study regarding acoustic, sub-phonetic aspects of consonant cluster development. The subjects were therefore accustomed to the recording situation. The mean chronological age of the subjects in this group was 34.8 months with a standard deviation of 1.96 months. The second ('Prechange') group consisted of four males and four females children who had not yet reached the voice change period. The chronological ages of this group were eight (1 female) and nine years (3 males and 4 females). The third group ('Adults') consisted of four adult male and four adult female subjects. The mean chronological age of this group was 28.9 years with a standard deviation of 4.05.

The subjects had no prior knowledge of the aim of the study. The adult subjects reported having no known hearing deficiency at the time of recording. For the groups consisting of children, these reports were given by their care-givers.

All subjects were recruited from the Umeå area.

2.2. Speech material

The word list used in the experiment consisted of nine monosyllabic words with an initial voiced, voiceless aspirated or voiceless-unaspirated-plosive ([sC]) cluster. The full list of target words was [bɑːk], [pʰɑːk], [spɑːk], [doː], [tʰoː], [stoː], [gɑːl], [kʰɑːl] and [skɑːl].

The criteria used in selecting the list were a) that bilabial, dental and velar places of plosive articulation were included, and b) that the size of the word list was small enough to obtain at least five repetitions of each word from the youngest children. The number of productions of each word was subsequently increased to 6 in order to decrease the impact of discarded productions on the statistical power of the design. The number of possible productions for each subject was therefore 9 words x 6 repetition = 54 items. The number of possible productions for the entire study was 1296.

The target words themselves were chosen so that each word would be either familiar to the youngest group of children (group 1) or possible to introduce to the children so that they, at the time of recording, had acquired a familiarity with the word. The novel word [pɑːk] was introduced as the name of a cartoon character.

2.3. Procedure

Recording sessions were conducted in a sound-treated room using a DAT recorder (48 kHz sampling frequency). The productions were elicited using a computer screen presenting a slide-show of hand-drawn pictures depicting the word. The transition from one slide to the next was accompanied by a randomised visual effect, such as the next slide appearing from one of the corners of the screen in order to introduce a novelty effect and, thereby, increase the subjects interest in the pictures and the production task.

The transition intervals were controlled from outside the recording studio, ensuring that the productions were made at a relaxed pace. At no time did the subjects know the number of words remaining; this reduced the effect of the different production quality at the end of a list.

2.4. Data inclusion criteria

The criteria for inclusion in the analysis were 1) that the production should be made in response to a presented picture and, 2) that the production should not co-occur with another sound which prohibited making an accurate measure of plosive release or onset of voicing. Thirty productions were removed from the speech sample due to failure to meet these criteria.

2.5. Markup procedure

For each production, the release of the plosive was marked at the last zero crossing in the waveform before a transient. The onset of voicing was marked at the last zero crossing before the onset of periodicity in the waveform (figure 1).

From these annotations, a voice onset time value was calculated as the difference in time between voicing onset and release of the plosive.

3. Results

The voice onset time measurements are displayed in the form of 'Box and whisker plots' in figures 2 and 3.

Accepted after full paper review

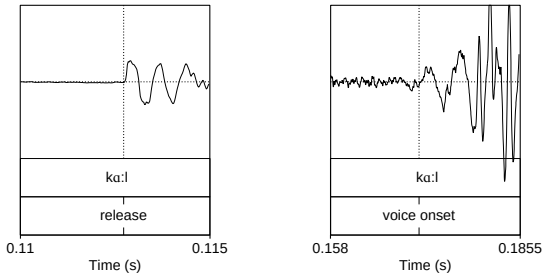

Figure 1: Example markup of plosive release (left picture) and voicing onset (right picture) for the initial plosive in the target word [kɑːl].

In each plot, the subsample median is indicated by a black dot in the centre of the box. The top and bottom of the box (the "hinges") indicate the first (bottom line) and third (upper line) quantile. Whiskers extend 1.5 times the length of the corresponding hinges. Data points outside the whiskers are viewed as outliers and are indicated by unfilled circles.

Previous studies have reported VOT distributions which significantly deviated from a normal distribution (Swartz 1992; Koenig 2000; Koenig 2001). Therefore, a Shapiro-Wilk's test for normality was applied in order to test the normality assumption of the collected data. Outliers indicated by unfilled circles in figures 2 and 3 were excluded from the data set tested in order to investigate the shape of the general distribution without the effect of extreme values. The results indicated that normality in the data could not be assumed ($P < 0.001$).

Due to the non-normality in the cell data, the non-parametric Wilcoxon signed rank test was performed comparing VOT measurements in the cells created by the factors gender, group, place of articulation and voicing type. The statistical testing failed to show a significant main effect of these variables ($P>0.05$).

Figures 2 and 3 do, however, indicate that the relationship of the gender-grouped data differs between the cells created by the factors group and voicing-type. Cell by cell Wilcoxon tests showed a significant effect of gender in aspirated plosives uttered by 'Young' speakers ($W=2763;p=0.002$) and 'Prechange' speakers ($W=3129;p=0.035$). A significant effect of gender was also found in voiced plosives uttered by adults ($W=1858;p=0.01$) but not in any of the other cells created by the factors group and voicing type.

4. Discussion

Previous research done in the field of voicing has established that the acoustic voice onset time (VOT) measurement facilitates acoustic categorisation of plosives into 'voiced' and 'voiceless' (Lisker and Abramson 1964) and that this categorisation corresponds well with subjects responses in perception tests (Lisker, Lieberman, Erickson, Dechovitz, and Mandler 1977; Kluender, Lotto, and Jenison 1995). Research into factors influencing voice onset time measurements have shown that VOT values within a voicing category may vary due to subject's gender (Swartz 1992) and age (Whiteside and Marshall 1998). Evidence for an interaction effect between age and gender has also been provided by Whiteside and Marshall (1998), but it is unclear whether a confounding factor of voice change entry was present in the data.

This study investigated the extent to which there is an interaction effect of age and gender on voice onset time. The age groups used in the study were widely separated as to remove potential confounding factors, such as voice change entry, from the subjects groups. The results show an overall tendency for female speakers of the 'Young' and 'Prechange' groups to have longer VOT in aspirated plosives than the male speakers. This effect of gender was shown to be significant for the younger subject groups across place of articulation, but was not present in the speech of the adult speakers or in the plosives uttered by children in a consonant cluster.

Two of the major articulatory components differing between initial and medial plosives are degree of oral pressure, and presence or absence of aspiration. In addition, the vocal folds' oscillation involved in voicing production is highly dependent on air pressure; one of the necessary prerequisites for voicing is a sufficient degree of trans-glottal pressure for the vocal folds to keep vibrating. In contrast, a weak airflow through the vocal folds may be insufficient for the Bernoulli effect to cause vocal fold vibration. However, a too forceful airflow forces the vocal folds too far apart. This delays the voicing onset until the glottal opening is restored, through vocal folds tension, to a configuration suitable for the Bernoulli effect. Therefore, the results from airflow measurements across age and gender published by Koenig (2000) should be considered a factor that may underlie the differences in VOT found here.

In her report, Koenig (2000) argued that her data from VhV syllables were consistent with aerodynamic studies showing greater airflow in unvoiced plosives produced by adult male subjects compared to those produced by women and children. Furthermore, the data showed a larger variability in oral airflow pressure in the speech of the 5-year olds. Due to the airflow dependence, airflow variability might be argued to result in a variable timing of voicing onset, which is in agreement with the VOT measurements for voiceless aspirated plosives produced by the children reported here.

The adult speakers did not produce significantly different voice onset times in aspirated plosives due to subject's gender. However, the voice onset times of voiced, unaspirated plosives produced by females were significantly smaller than in the male productions; with the bulk of productions in the pre-voiced range.

Arguably, this might also be an effect of a airflow differences during plosive articulation. A larger degree of oral airflow at plosive release may be due to a larger oral pressure being built up during the preceding closure. This would in effect create a reaction force to the

Accepted after full paper review

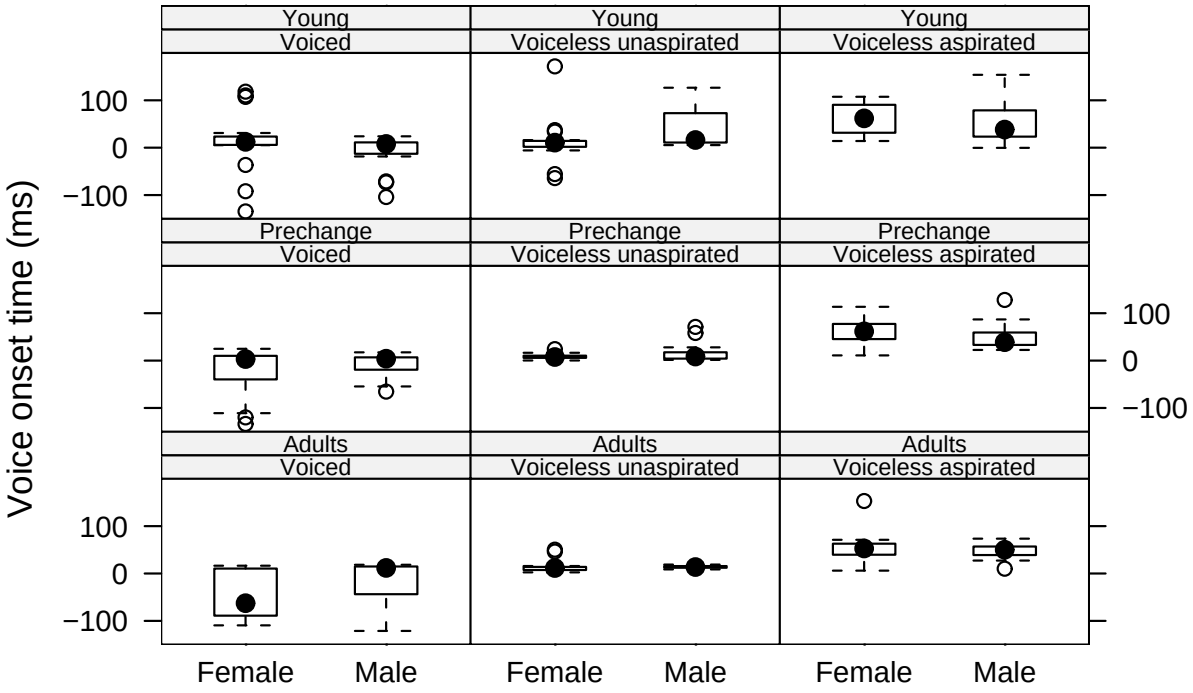

Figure 2: Box-and-whisker plots of Voice Onset Time (in milliseconds) for tokens produced at the labial of articulation divided by subject group and plosive voicing and aspiration category. In each cell, the data are separated according to gender.

pressure from the lungs, which in turn would decrease the likelihood of a rapid voicing onset in the plosives produced with a more powerful airflow. In contrast, female uttered plosives, which were shown by Koenig (2000) to be produced with a relatively weak airflow at plosive release, would influence voicing onset time to a lesser extent, thus increasing the likelihood of a voicing onset relatively soon after the release of the plosive, causing a smaller mean VOT value.

5. Conclusion

The results presented in this report show a significant difference in VOT associated with subject gender for aspirated, voiceless plosives in the groups containing 3-year olds as well as the group consisting of 8-9 year old children. Plosives produced by females had a longer VOT interval than those produced by male speakers.

Furthermore, the analysis showed a significant, opposite, gender difference in VOT for the voiced plosives uttered by adult speakers. In these productions, male speakers produced plosives with a longer VOT interval than female speakers. The results are consistent with data from previous studies investigating gender and age differences in VOT as well as with the distributional properties of the results gathered from studies investigating plosive airflow. It is, therefore, hypothesised that the observed gender differences in VOT in aspirated plosives uttered by children and voiced plosives uttered by adults might be caused by differences in trans-glottal airflow properties.

In order to test this hypothesis, a similar experiment to that reported here should be conducted, using similar subject groups but with an experimental setup that would facilitate measurement of both voice onset time and airflow. This experimental design would facilitate evaluation of the hypothesised correlation between voice onset time and airflow.

References

Fant, G. (1973). *Speech Sounds and Features*. The MIT Press, Cambridge, Massachusetts.

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.

Kewley-Port, D. and M. S. Preston (1974). Early apical stop production: A voice onset time analysis. *Journal of Phonetics 2*, 195–210.

Kluender, K. R., A. J. Lotto, and R. L. Jenison (1995). Perception of voicing for syllable-initial stops at different intensities: Does synchrony capture signal voiceless stop consonants? *Journal of the Acoustical Society of America 97*(4), 2552–2567.

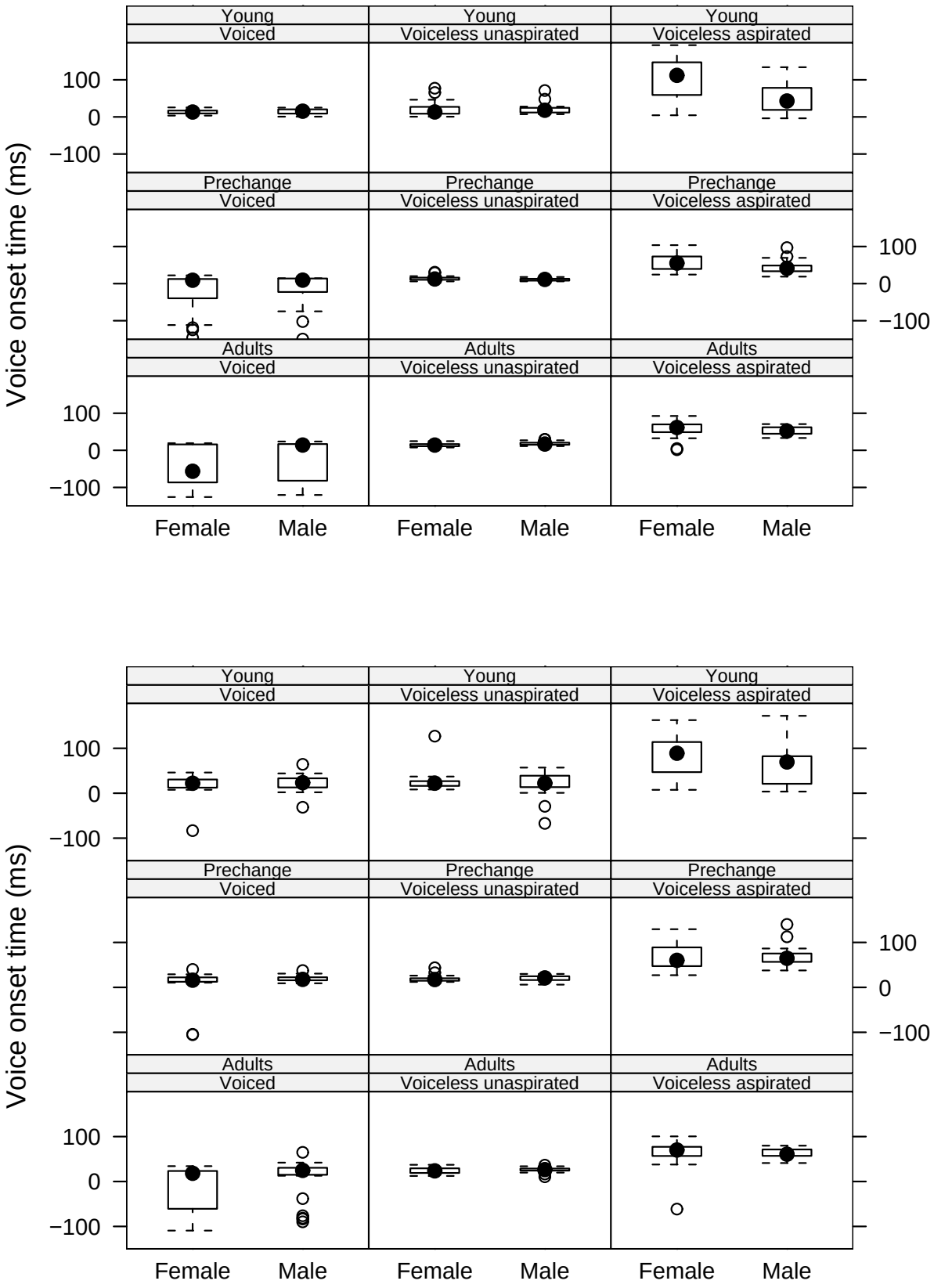

Figure 3: Box-and-whisker plots of Voice Onset Time (in milliseconds) for tokens produced at the dental (top) and velar (bottom) places of articulation divided by subject group and plosive voicing and aspiration category. In each cell, the data are separated according to gender.

Proceedings of the 10th Australian International Conference on Speech Science & Technology
Macquarie University, Sydney, December 8 to 10, 2004.

Accepted after full paper review

Koenig, L. L. (2000). Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language, and Hearing Research 43*, 1211–1228.

Koenig, L. L. (2001). Distributional characteristics of VOT in children's voiceless aspirated stops and interpretation of developmental trends. *Journal of Speech, Language, and Hearing Research 44*, 1058–1068.

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

Lisker, L., A. M. Lieberman, D. M. Erickson, D. Dechovitz, and R. Mandler (1977). On pushing the voice-onset-time boundary about. *Language and Speech 20*, 209–216.

Swartz, B. L. (1992). Gender differencies in voice onset time. *Perceptual and Motor Skills* (75), 983–992.

Titze, I. R. (1994). *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice-Hall.

Whiteside, S. P. and J. Marshall (1998). Voice onset time patterns in 7-, 9-, and 11- year old children. In R. H. Mannell and J. Robert-Ribes (Eds.), *Proceedings of the 5th International Conference on Spoken Language Processing*, Volume 6, pp. 2687–2690. Australian Speech Science and Technology Association, Incorporated.

Accepted after full paper review