Article:

Whiteside, S.P. and Marshall, J. (2001) Developmental trends in voice onset time: some evidence for sex differences. Phonetica, 58 (3). pp. 196-210. ISSN 1423-0321

https://doi.org/10.1159/000056199

Published paper
Whiteside, S.P. and Marshall, J. (2001) *Developmental trends in voice onset time: some evidence for sex differences.* Phonetica, 58 (3). pp. 196-210.

Phonetica, 58, 196-210.

DEVELOPMENTAL TRENDS IN VOICE ONSET TIME: SOME EVIDENCE FOR SEX DIFFERENCES

S. P. Whiteside & J. Marshall

Department of Human Communication Sciences, University of Sheffield,
Sheffield S10 2TA, United Kingdom.

Send correspondence to S. P. Whiteside, Department of Human Communication Sciences, University of Sheffield, Sheffield S10 2TA or e-mail (s.whiteside@sheffield.ac.uk).

Phonetica, 58, 196-210.

Abstract

This study reports on an investigation into the voice onset time (VOT) patterns of the plosives /p b t d/ in a group of thirty children aged 7 (n=10), 9 (n=10) and 11 (n=10) years. Equal numbers of girls and boys participated in the study. Each child named a series of letter objects to elicit /p b t d/ in a syllable onset position with a fixed vowel context. VOT data were examined for age, sex and plosive differences, and the following working hypotheses were made. Firstly, that there would be sex differences in the VOT patterns of preadolescent children. Secondly, that the sex differences in VOT patterns would be linked to age and development, and that these would eventually become marked by the age of 11 years, by which time adult-like VOT values should have been achieved. Finally, that the extent of sex and age differences would be dependent upon the plosive being investigated. Results indicated patterns of decrease with age in the VOT values of /p b/ for the boys, with some evidence of increases in the VOT values of /t/ for the girls. In addition, "voiced" and "voiceless" cognates showed a more marked bimodal distribution in the girls' VOT patterns. This bimodal distribution was investigated by examining the degree of difference between the VOT values of voiced and voiceless cognate pairs /p b/ and /t d/, and examining the effects of age, sex and cognate pair. These results indicated that more marked sex differences in the "voiced"/"voiceless" contrast emerged between the data of the 9- and 11-year olds, a pattern, which was more marked for the alveolar plosives. These preliminary results confirmed all three working hypotheses. These findings are presented and discussed both within a developmental and sociophonetic framework.

Phonetica, 58, 196-210.

Introduction

In English, plosive consonants may be categorised as phonologically "voiced" or "voiceless". A number of consonants are produced with the same manner and place of articulation but differ only along the dimension of voicing. Here, voicing is the primary phonetic dimension used to distinguish these minimal pairs, for example with the plosive minimal pairs /p b/ /t d/ /k g/, /p t k/ are "voiceless" and /b d g/ are "voiced". A range of physiological, and acoustic differences, have been identified between the two voicing categories. Acoustic cues which have been noted to differentiate the two voicing categories include voice onset time (Lisker & Abramson, 1964; Lisker & Abramson, 1967; Abramson & Lisker, 1973), amplitude of the burst (Repp, 1979; Jongman & Blumstein, 1985), fundamental frequency characteristics (Haggard, Ambler & Callow, 1970), closure durations (or stop gap) preceding the release of plosive (Lisker, 1957), and preceding vowel duration in the case of non-initial plosives (Chen, 1970).

There is therefore, a ‘many-to-one mapping’ of acoustic cues to voicing (Lisker, Liberman, Erickson, Dechovitz & Mandler, 1977). However, of these cues, VOT is considered to be the primary cue for the voicing distinction as it has been identified as ‘the single most effective measure’ (Lisker & Abramson, 1971) in the perception and production of word initial prevocalic plosives (Lisker & Abramson, 1964; Lisker & Abramson, 1967). Furthermore, this cue is robust and maintained in the spontaneous speech of adult speakers (Abramson & Lisker, 1995; Krull, 1991). The number and range of studies that have been carried out on this temporal feature illustrate the importance of VOT. VOT has been investigated with respect to age differences (Eguchi & Hirsh, 1969; Macken & Barton, 1979; Barton & Macken, 1980) and sex differences (Swartz, 1992; Smith, 1978b; Ryalls, Zipprer & Baldauff, 1997) for example.

Phonetica, 58, 196-210.

VOT values exhibit intrinsic variations, relating to both their place of articulation and voicing category. VOT values increase as the place of articulation moves from anterior to more posterior, that is, alveolar plosives will be produced with a longer VOT value than the corresponding labial plosive (Lisker & Abramson, 1964; Lisker & Abramson, 1967). "Voiceless" plosives in English are produced with a wider distribution of VOT values than the "voiced" plosive as the "voiceless" plosive is the less stable member of the minimal pair. This instability could be explained by the increased complexity in timing of separately innervated supraglottal and glottal articulators which has been demonstrated through fibreoptic (Sawashima, Abramson, Cooper & Lisker, 1970) and EMG studies (Hirose & Gay, 1972). However, the degree of variability and stability needs to be constrained by the language in question, and whether a two- (e.g. American and British English, Cantonese, Dutch), three- (e.g. Thai), or four-category (e.g. Hindi) system of plosives is operating (see Lisker & Abramson, 1964 for examples of two-, three- and four-category languages).

The role of the development of motor control in the acquisition of adult-like VOT production needs to be considered. Stages in the acquisition of the voicing contrast have been discussed in previous studies which may go some way to explaining developmental patterns of VOT values with respect to a child's developmental level. The actual age at which these stages are reached is controversial because children often present different developmental VOT patterns, and reach different stages of VOT development at varying chronological ages. However, the stages of development can be summarised from the literature as follows.

Stage 1 - From the latter half of the first year to around the age of 18 months, some children display no distinction in the production of VOT values between the "voiced" and "voiceless" adult forms. In these cases, most stop productions fall within the range of 0-30 msec (Kewley-Port & Preston, 1974; Macken & Barton, 1979)

Stage 2 - a distinction develops with "voiceless" stops produced with longer VOTs, although they are still perceived as "voiced" (Macken & Barton, 1979; Barton & Macken, 1980). For some children, this stage occurs around 18 months of age, but can extend up to the age of 28 months (Macken & Barton, 1979). This is an example of covert contrast production where the differences, although produced, are too subtle to be perceived by listeners (Edwards, Gibbon & Fourakis, 1997);

Stage 3 - with further development, overshoot of adult VOT values is noted in the production of "voiceless" stops which are later retracted back to more adult-like values around the age of 4 years (Barton & Macken, 1980). The observation that 4-year old children display an overshoot of adult VOT values, therefore contrasts with the interpretation of VOT data from 2- and 6- year olds which suggests that the development of the VOT is a continual movement towards adult values (Zlatin & Koenigsknecht, 1976);

Stage 4 - by the age of 6 years, a bimodal distribution of VOT values is produced for the contrastive "voiced" versus "voiceless" stops although some overlap continues to be observed (Kent, 1976);

Stage 5 - from the age of six years, stops are produced with adult-like VOTs with non-overlapping bimodal distribution (Gilbert & Purves, 1977). What is worthy of note at this point however, is that variability in VOT productions continues well beyond the normal period of phonological acquisition, and they generally reach an adult-like minimum level around eight years of age (Eguchi & Hirsh, 1969; Sander, 1972; Tingley & Allen, 1975; Zlatin & Koenigsknecht, 1976).

Different perspectives on the underlying factors affecting acquisition of voicing have been proposed including perception and motor development. In order for any voicing contrasts to appear, the child must have developed to a point where the perceptual system is able to discriminate salient phonetic features, control has been gained over the timing of supraglottal and glottal articulations, and there is a need to make semantic distinctions between minimal pairs which differ only with respect to voicing (Zlatin & Koenigsknecht, 1976).

In the early stages of the acquisition of the voicing contrast, infants do not appear to be limited by their perceptual abilities as they are able to perceive categorical differences in VOT before they are able to reproduce them (Wolf, 1973; Zlatin & Koenigsknecht, 1975; Aslin et al, 1983). However, there is a link between perceptual identification categories and VOT values used in production, and the VOT difference needed to distinguish "voiced" versus "voiceless" shows a reduction with age until adult-like perception is achieved at around six years (Lisker & Abramson, 1967; Zlatin & Koenigsknecht, 1975). This age effect in perception may be due to peripheral or central auditory processing, or it may be due to the speaker's own changes in speech production (Flege & Eefting, 1986).

In addition to perceptual development, we need to consider the role of the development of motor control in the acquisition of adult-like VOT productions. In the production of a stop consonant, three physiological mechanisms are required: (i) articulations to allow stop closure and release; (ii) articulations to isolate the nasal cavity at the velum; and (iii) articulations to initiate vocal fold vibration (Kewley-Port & Preston, 1974). In order for contrastive voicing production the speaker must be able to control these three separately innervated mechanisms. As with all other areas of motor development, control of laryngeal timing continues to be refined through childhood (Tingley & Allen, 1975; Zlatin & Koenigsknecht, 1976; Ohde, 1985). Three main stages of the development of motor speech mechanisms could therefore be proposed to account for VOT values and their variability:

Stage 1. In the early stages of speech motor development, all stops are produced within the same voicing category as articulatory gestures for "voiced" stops are less complicated than those for voiceless (Kewley-Port & Preston, 1974);

Stage 2. Between four and seven years of age the child refines motor patterns of speech. As children have less refined neuromotor capabilities than adults, they have an inability to exert adult-like control over speech productions which is realised in their speech as longer durations and greater variability (Smith, 1978a);

Stage 3. By eleven years, variability in fine and gross motor skills reaches adult-like minimum values (Eguchi & Hirsh, 1969; Tingley & Allen, 1975).

There have been, so far, relatively few investigations into sex differences in VOT production. The results from studies are somewhat contradictory with reports of longer VOT values in males, longer VOTs in females, and in some cases no differences (Sweeting & Baken, 1982). Smith (1978b) for example reported longer VOTs for initial /d g/ in adult males. However, more studies have noted longer VOTs in adult females for the plosives /t d/ (Swartz, 1992), /p d/ (Whiteside & Irving, 1997) and across all plosives (Ryalls et al., 1997). In general though, the overall trend indicates that adult females produce longer VOT values than adult males. Underlying causes of this sex difference have not been fully established but suggestions have included physiological or anatomical differences (Swartz, 1992).

Given the evidence, which suggests that the VOT patterns of adults show sex-linked differences, the issue of whether sex-linked differences in VOT occur in children's speech needs to be addressed. Sex differences in vocal parameters such as formant frequencies and fundamental frequency have been attributed in part to sociophonetic or stylistic conventions

(Mattingly, 1966; Van Bezooijen, 1995). In addition to the extent of sex differences in the pronunciation of vowels, there is some sociophonetic evidence for allophonic sex differences in the consonants of speakers from Sheffield, England (Stoddart et al., 1999). Examples of the sex-linked differences in the realisation of consonants include the tendency for males to use the glottal allophone ([ʔ]) for non-initial /t/ more frequently than females, and the tendency for males to realise /θ,ð/ as [f,v] in both medial and final positions (e.g. [bɹɒvə] - *brother*, [mɑːf] - *mouth*). In addition to these documented sociophonetic sex differences, there is some acoustic evidence for sex differences in the VOT patterns of adults from the same geographical region (Whiteside & Irving, 1997). This suggests the possibility that sex differences in VOT patterns may, in part, be influenced by sex-linked sociophonetic factors. Furthermore, if these sex differences exist in adults, the question is whether they also exist in children's speech. Further still, if these sex differences do exist in children's speech, when do they begin to emerge as children approach adulthood? There is some evidence to suggest that adult-like VOT values are achieved around the age of 11 years (Tingley & Allen, 1975). On this basis, one could predict that if any sex-linked differences in VOT patterns exist, they would begin to emerge during preadolescence and become established by the age of 11 years. Also, given the range of sex differences reported for the VOT patterns of different plosives (see above), one would predict that any age- and sex-linked differences might be affected by the plosive in question.

The aim of this preliminary study was therefore to investigate both sex and age differences in the production of VOT of /p b t d/ in a fixed phonetic context for thirty male and female British English preadolescent children from Sheffield, across the ages of 7, 9 and 11 years. On the basis of evidence from previous research, the following hypotheses were therefore made.

Firstly, there would be sex differences in the VOT patterns of preadolescent children. Secondly, the sex differences in VOT patterns would be linked to age and development, and would eventually become marked by the age of 11 years, by which time adult-like VOT values should have been achieved. Finally, the extent of sex and age differences would be dependent upon the plosive being investigated. Therefore, in addition to the effects of sex, age and plosive on VOT patterns, interactions between these factors were also the subject of investigation in this study.

Method

Subjects

Five males and five females in each of three age groups: i) 7 years (mean age 6;9) ii) 9 years (mean age 8;6) and iii) 11 years (mean age 10;7), were selected from a primary school in Sheffield. The mean and standard deviations for age (given in months), height (given in centimetres) and weight (given in kilograms) are given in Table 1 by sex. All subjects: i) had a similar height and weight to the same sex members of their age group; ii) spoke with a similar regional accent iii) had lived in Sheffield all their lives and within 3 to 5 miles of the school; iv) had age-appropriate intelligence levels as judged by the class teacher; v) were monolingual speakers of English; and vi) had no speech, language or hearing problems; vii) volunteered to participate in the study with parental consent.

<PLACE TABLE 1 AROUND HERE>

Speech stimuli

All subjects produced five repetitions of each of the plosives /p b t d/ in an intervocalic, syllable initial position within the phrases of ‘silver /pi/’, ‘silver /bi/’, ‘silver /ti/’and ‘silver /di/’. No data were obtained for the velar place of articulation in an attempt to maintain a maximum control of phonetic and linguistic variables. That is to say, although /ki/ ('key') is a familiar word of English, its phonetically matched "voiced" cognate /gi/ ('ghee') is not. It was therefore deemed unacceptable to use these syllables as a minimal pair. The target plosives were elicited in syllable initial prevocalic position for two reasons. Firstly, as this is the most easily measurable position for VOT and, secondly, to allow comparison with other studies where the data was also obtained within a similar context. Vowel context was standardised in all cases with the plosive preceded by schwa (/ə/) and followed by /i/. All plosives were elicited in the same manner by asking the children to name silver (aluminium foil) shaped letters. Prior to recording, the task was explained to subjects to familiarise them with the materials and procedure. A set of practice materials was used to increase their confidence and reduce any possible hesitancy that might confound the results. In cases of inaccurate articulation the subject was asked to repeat the test item to ensure five samples of each plosive were available to be analysed. Six hundred target plosive tokens were elicited in total.

Recording and analysis

The speech data were recorded in a quiet room. All subjects were recorded on the same day to reduce the possibility of environmental differences altering the recording. Data was recorded directly onto a DAT (Digital AudioTape) recorder (Sony, model TCD-D3).

Speech samples from each subject were digitized onto a Kay Elemetrics Computerised Lab (CSL) model 4300 using a sampling rate of 10 kHz. From this digital information, speech pressure waveforms and spectrograms were generated and displayed. VOT measurements were made directly from the speech pressure waveforms by measuring the distance between the release of the plosive to the onset of voicing (the first cycle of periodic oscillation in the speech pressure waveform). The point of closure release was taken as the transient burst of the plosive's release. In the cases where measures of VOT needed validation, Fast Fourier transform wideband (146 Hz) spectrograms were used. In these cases where both the speech waveform and the spectrogram were referred to for validation, the VOT measurement was taken from the same data source. In the cases where VOT was unclear, e.g. plosives being released with affrication or the presence of background noise, the speech sample was discarded. The total number of omitted data points was 10, which amounted to 1.67% of the total data sample. In light of the different types of VOT patterns that are possible (Swartz, 1992), a decision was taken to treat negative VOT values separately from those with zero or positive values. There was only one speech sample (for /b/) of a 9-year old boy that displayed pre-voicing or negative VOT, which was excluded from analysis for statistical reasons.[1] The total number of data samples that were therefore excluded from analysis amounted to 1.8% of the data sample.

To ensure consistency in the VOT measurements, a test of intra-rater reliability was carried out. For this, 20% of the data was reanalysed one week following the completion of the initial analyses. The data were selected by randomly choosing one subject from each of the 6 age x sex groups. A Pearson’s product-moment correlation was used to calculate the level of intra-rater

[1] The lack of instances of pre-voicing in the children's data could be attributed to the object-naming paradigm. This would need further investigation by comparing VOT patterns elicited via object naming, picture naming, and repetition paradigms. In addition, speech samples collected during unscripted communication settings would also serve as interesting comparisons.

reliability. A significant coefficient (r=0.99, p<0.001) demonstrated a high level of intra-rater reliability.

Results

The mean VOT values for the plosives /p b t d/ are given in Table 2 and illustrated in Figures 1 to 4, according to age and sex. In order to test for effects of age, plosive and sex on VOT, the data were subjected to a three way (by age, plosive and sex) ANOVA, where the following results were indicated. Significant age differences (F (2, 565)=5.9, p<.005) were found with post-hoc Scheffé tests indicating significant contrasts (p<.05) between the VOT values of the 7 and 9 year old groups, and between those of the 9 and 11 year olds. In both significant comparisons, the 9-year-olds had longer VOT values. There were also significant plosive differences (F (3, 565)=792.9, p<.0001), which were characterised by systematic differences between the "voiced" and "voiceless" plosives, and differences in VOT according to place of articulation. The order of magnitude of VOT according to the voicing and place dimensions was as expected, and was as follows: b < d < p < t. The significant sex differences (F (1, 565)=12.0, p<.005) were characterised by the girls having longer VOT values than the boys, a finding that confirmed one of our hypotheses.

In addition, significant age x sex (F (2, 565)=16.7, p<.0001), plosive x sex (F (3, 565)=9.8, p<.0001), and age x plosive x sex (F (6, 565)=4.0, p<.005) interactions were found. Age x sex interactions confirmed the emergence of sex-linked differences and were characterised by a general increase in VOT values between age 7 and 9 years, and a marked decrease between age 9 and 11 years in the boys' data. This contrasted with slight decreases and increases in VOT values for the same age comparisons for the girls. By the age of 11 years, there were marked sex differences with the girls having longer VOT values than the boys. This is illustrated by the fact that the mean overall VOT value for all four plosives (/pbtd/) was 57.3 ms for the girls compared to 44.0 ms for the boys. The significant plosive x sex interactions confirmed that sex differences varied according to the plosive in question, and were accounted for by the girls having longer VOT values than the boys for /pbt/. Although there was evidence of sex differences for /b/ (Figure 2), these were more marked for /p/, with the girls and boys having an overall VOT mean of 80.1 ms and 62.4 ms, respectively. In addition, the sex differences were accentuated in the data of the 11-year olds (Figure 1). The significant age x plosive x sex interactions are explained for example by the following VOT patterns. The girls showed an increase in VOT values for /t/ from age 7 through to 11 years (Figure 3). A pattern which contrasted with the data of the boys' data which showed an increase between age 7 and 9 years, but a marked decrease in VOT values between age 9 and 11 years (Figure 3). In the case of /d/, decreases in VOT values were observed for the girls from age 7 years through to 11 years (Figure 4), whereas the boys showed an increase between the ages of 7 and 9 years, and a decrease thereafter (Figure 4). No significant age x plosive interactions were found. The net effect of sex, age and plosive differences indicated the emergence of a more marked bimodal distribution between the girls' "voiced" and "voiceless" plosives. This therefore suggested both an age- and sex-linked pattern of "voiced"/"voiceless" contrasts for both the bilabial and alveolar plosives.

<PLACE TABLE 2 AROUND HERE>

<PLACE FIGS 1, 2, 3, 4 AROUND HERE>

In order to investigate the phonological contrast between the "voiced" and "voiceless" plosives, the mean difference between the VOT values of the "voiceless" bilabial plosive /p/ and its "voiced" cognate /b/, and the "voiceless" alveolar plosive /t/ and its "voiced" cognate /d/ was calculated. The method used was to subtract the "voiced" VOT values from the "voiceless" VOT values. This parameter was termed the VOT "voiceless"-"voiced" difference (VOTVVD). The mean and standard deviation values of this parameter (VOTVVD) are given in Table 3 for each cognate pair (/pb/ and /td/) according to age and sex group. This parameter was then subjected to a three way ANOVA, to test for age, sex and cognate pair effects. Although results indicated no significant age differences, significant differences were found for both sex ($F(1, 277)=22.9$, $p<.0001$), and cognate pair ($F(1, 277)=49.1$, $p<.0001$). The significant sex differences were characterised by the girls having larger values for VOTVVD, a finding that confirmed the more bimodal distribution in the girls' VOT values. The significant cognate pair differences were due to the alveolar plosives having larger values for VOTVVD, than the bilabials.

<PLACE TABLE 3 & FIGURES 5 & 6 AROUND HERE>

In addition, the following significant interactions were found. Significant age x sex interactions ($F(2, 277)=15.0$, $p<.0001$) were characterised by the boys having VOTVVD values that increased between age 7 and 9 years, with decreases in this parameter between the ages of 9 and 11 years (Figures 5 and 6). This contrasted with the girls' patterns, which displayed a reverse pattern for both age intervals (Figures 5 and 6). Significant interactions of age x cognate pair

(F(2, 277)=5.9, p<.005) were characterised by an increase in VOTVVD values across age 7, 9 and 11 years for the alveolar plosives (Figure 6), and a decrease and subsequent increase for the bilabials for the same age comparisons (Figure 5). Significant sex x cognate pair interactions (F(1, 277)=4.9, p<.05) were marked by the girls having larger mean VOTVVD values than the boys for both the bilabial and alveolar plosives. For the girls, the VOTVVD values were 65.8 ms and 76.5 ms for the bilabials and alveolar plosives, respectively. Corresponding values for the boys were 50.4 ms and 76.5 ms. Finally, significant interactions for age x sex x cognate pair (F(2, 277)=4.3, p<.05) were characterised by both the boys and girls showing an increase in VOTVVD values between the ages of 7 and 9 years for the alveolar plosives. However, between the ages of 9 and 11 years the boys showed a decrease in values of VOTVVD compared to an increase in values in the girls' data (Figure 6). The pattern for the bilabial plosives is more complex with the boys showing an increase in VOTVVD values between the ages of 7 and 9 years but a decrease between the ages of 9 and 11 years. This pattern was reversed for the girls (Figure 5).

Discussion

The VOT patterns reported here for the "voiced/voiceless" plosives were as expected with the "voiced" cognates showing smaller VOT values than their "voiceless" cognates. In addition, the "voiced/voiceless" bilabial plosives showed smaller VOT values than the alveolar "voiced/voiceless" plosives. These patterns were not unexpected and agree with the results of the numerous studies that have been reported previously on VOT (e.g. Klatt, 1975; Lisker & Abramson, 1964; Lisker & Abramson, 1967; Ryalls et al., 1997; Smith 1978b).

In addition, the VOT patterns reported here across the 7, 9 and 11-year old boys and girls suggest that there are developmental trends for this parameter. Furthermore, there is evidence to suggest that these trends are sex-linked for the children studied here (Figures 1 to 4), which confirms the first hypothesis which was that there would be sex differences in the patterns of preadolescent children. The sex-linked developmental patterns in the VOT data are most revealing for VOTVVD, which parameterises the phonological contrast between the homorganic "voiced/voiceless" plosives. The developmental trend for this parameter was that the girls showed a decrease in contrast for /p b/ from age 7 to 9 years, with a subsequent increase from age 9 to 11 years. However, there was a clear pattern of increase in the contrast for /t d/ from age 7 years, through age 9 years up to 11 years. The boys showed a different pattern for both /p b/ and /t d/ with VOTVVD increasing between age 7 and 9 years, and decreasing between age 9 and 11 years.

These data suggest therefore, that the phonological contrast between /p/ and /b/, and /t/ and /d/ was more marked for the 11 year old girls compared to their male peers (Figures 5 and 6). This trend is replicated in data for a group of adult speakers from the same geographical region (Whiteside & Irving, 1997). This confirmed the second hypothesis which stated that sex differences in VOT patterns would be linked to age and development and become marked by the age of 11, by which time adult-like values would be established. The reasons underlying changes in the VOTVVD parameter were the net result of the following VOT patterns that were found for the boys and girls between age 7 and 11 years. In the case of the bilabial plosives /p b/, the boys showed increases in VOT values for /b/ but decreases for /p/. These patterns were different to those for the girls who showed little change in the VOT values of /p/ and /b/. For the alveolar plosives /t d/, the boys showed decreases in VOT values for both /t/ and /d/ compared to the girls who showed increases in the VOT values of /t/, but decreases for /d/. These plosive-specific patterns confirmed the third hypothesis - that the extent of sex and age differences would be dependent upon the plosive being investigated.

The reasons for emerging developmental patterns are likely to involve an interaction of factors that include sex-linked developmental changes in anatomy and physiology and sociophonetic influences. Anatomical and physiological factors affecting the development of vocal characteristics and the sex-linked patterns that emerge are acknowledged and widely documented in the literature (Eguchi & Hirsh, 1969; Hasek et al., 1980; Kent, 1976). VOT is a critical temporal parameter that requires fine coordination between sub-laryngeal, laryngeal and supra-laryngeal systems. On the basis that these systems are still undergoing development up to and beyond puberty, there is no reason to assume that VOT is exempt from the influences of these developmental changes. VOT is therefore likely to reflect some sex-linked developmental patterns in a similar way to those reported on vocal characteristics. Some evidence for the existence of male/female differences in the VOT data of adult speech is documented (Ryalls et al., 1997; Swartz, 1992; Whiteside & Irving, 1997). The data reported here suggest that there is some evidence for the emergence of these sex-linked differences in VOT patterns between the ages of 9 and 11 years.

A pattern that emerged in the boys' data must not be overlooked. This was the observation that largest VOT values were produced by the 9 year old males for the plosives /p t d/, a pattern which was not mirrored for the girls or boys at any other age (see Table 2, Figs. 1, 3 and

Phonetica, 58, 196-210.

4). Possible support for such dramatic changes in VOT production at age nine may come from other developmental acoustic studies. A study on fundamental frequency (F0) by Hasek et al (1980) reported the emergence of a significant difference between the average F0 of boys and girls at around age seven to eight years due to a rapid decline in F0 in males. It is suggested that these sudden changes in male F0 may be accompanied by changes in temporal features of speech, for example, in VOT. Possible explanations for the sudden F0 change include anatomical changes that occur at the onset of puberty. If the anatomical change viewpoint is taken, it seems feasible that sudden increases in the size of the vocal tract and the larynx would influence a number of acoustic parameters, for example, F0, formant frequencies and VOT. The boys might be attempting to stabilise their productions in the face of vocal tract changes, which may lead to adjustments in timing control. With increased age, adaptations would be made, allowing control of timing to stabilise. These patterns in boys at age nine years were not reflected the girls' VOT data. This may be due to the less dramatic vocal tract changes that occur in females during pre-adolescence. This pattern contrasts with those for the boys, which showed irregular and dramatic changes between 7 and 9 years and then between age 9 and 11 years. It is suggested that these dramatic changes may be due to influences in laryngeal and vocal tract development changes, during which the boys have to make adjustments to the timing and phasing of sub-laryngeal, supra-laryngeal and laryngeal systems for the production of voice onset time.

Sociophonetic factors and their influence on sex-linked patterns in vocal characteristics (Linke, 1973; Van Bezooijen, 1995) voice quality (Henton & Bladon, 1985), phonetic realisations (allophonic variation) and speaking style (Byrd, 1992, 1994) have been reported and acknowledged for men and women. These sex-linked allophonic differences have included evidence of women showing more distinctive and peripheral vowel spaces than men (e.g.

Phonetica, 58, 196-210.

Henton, 1985, 1995), releasing sentence-final stops with a greater frequency compared to men, producing glottal stops and flaps less frequently than men (Byrd, 1994). This evidence is resonant of females using more carefully articulated speech, and in particular, adopting this speech style in experimental situations. The production of greater "voiced" versus "voiceless" phonological contrast in the VOT patterns of the 11-year-old boys and girls in this study supports this suggestion and provides some evidence that sociophonetic influences on the development of VOT may emerge with age during preadolescence. In addition, given the recent account of some of the sociophonetic sex differences that occur in the pronunciation of consonants and vowels of speakers from Sheffield (Stoddart et al., 1999), this suggestion does not appear to be unreasonable. Furthermore, the acoustic data here support impressionistic judgements of syllable initial plosives being less aspirated for men compared to women, for this geographical region of the British Isles.

Therefore, in light of the possible sociophonetic and cultural factors and their influence on allophonic variation, comparisons of VOT patterns that are made across different speech communities and accents need to be made with caution. These sociophonetic and cultural factors may go some way in explaining why some studies have either found (Ryalls et al., 1997; Smith, 1978b; Swartz, 1992; Whiteside & Irving, 1997) or not found (Sweeting & Baken, 1982) sex differences in VOT. While the tendency has been for women to have larger VOT values than men (Ryalls et al., 1997; Smith, 1978b; Swartz, 1992; Whiteside & Irving, 1997), there has been some evidence for men to have slightly longer lag VOT values than women for "voiced" plosives (Smith, 1978b). It is also worthy of note here that the men in the aforementioned study also showed evidence of more pre-voicing (voicing lead) and longer voicing leads than their female peers (Smith, 1978b). Swartz (1992) reported the former pattern.

Phonetica, 58, 196-210.

The preliminary data in this study confirm sex differences in VOT for /p b t d/, and their emergence by the age of 11 years. Furthermore they provide evidence of some convergence with the VOT values of young adult men and women (mean age 33.4 years) from the same geographical region of the British Isles (Whiteside & Irving, 1997). One result of this emerging sex-linked difference was that the girls show a more marked "voiced"/"voiceless" contrast than the boys. The factors underlying this development can probably be attributed to a number of factors, which would include motor speech control, and development, anatomical and physiological sex-linked developmental differences, and sociophonetic and cultural factors.

In light of the evidence presented in this study and previous studies (Ryalls et al., 1997; Swartz, 1992; Whiteside & Irving, 1997), the sex-linked development of VOT patterns therefore deserves further investigation. A more comprehensive picture of how sex-linked VOT patterns develop and emerge in children could be established by examining speech data elicited via different paradigms (e.g. picture-naming, repetition, conversational). In addition, these data would need to contain a comprehensive plosive and vowel repertoire. Also, by comparing these emerging sex-linked patterns with those of adults from the same accent and social group, a clearer picture of the extent of the sociophonetic influences on the development of VOT patterns will become clearer. A point that needs addressing however, in any developmental study of VOT (or other acoustic phonetic parameter for that matter), is the fact that even in adulthood, there are developmental changes which occur at different stages across the life span. These changes maybe age-related and/or evidence of the ongoing changes in phonological variation due to sociolinguistic factors. Therefore, careful attention needs to be paid to the age profiles of both male and female adults and children alike, when undertaking any sociophonetic investigation. The fact that Stoddart et al. (1999) found some age-based differences in pronunciation upholds this suggestion, and therefore the need for more balanced sampling of adult age (and sex) groups. This approach would provide data upon which to establish more definitively the effects of age (and sex) on phonetic and phonological variation. More extensive acoustic sociophonetic studies of speakers from this geographical region of the British Isles are clearly needed.

Acknowledgments

Some of the results reported here were presented at the 5th International Conference on Spoken Language Processing in Sydney (1998). Our thanks go to the staff and pupils of St. Mary's School in Walkley, who made this study possible. We also wish to thank two anonymous reviewers for their helpful comments and suggestions on an earlier version of the manuscript

References

Abramson, A.S.; Lisker, L.: Voice-timing perception in Spanish word-initial stops. J. Phonet. *1*: 1 - 8 (1973).

Abramson, A.S.; Lisker, L.: Voice timing for stop classification in conversational English. Proc. 13th ICPhS, vol. 4, pp. 128 - 131 (Stockholm 1995).

Aslin, R.; Pisoni, D.; Jusczyk, P. W.: Auditory development and speech perception in infancy; in Mussen, Handbook of Child Psychology (4th Ed. - Vol. 2 Haith, Campos, Infancy and Developmental Psychobiology), pp. 573 - 687 (Wiley & Sons New York 1983).

Barton D.; Macken, M. A.: An instrumental analysis of the voicing contrast in word-initial stops in the speech of four-year-old English-speaking children. Lang. Speech *23*: 159 - 169 (1980).

Phonetica, 58, 196-210.

Byrd, D.: Preliminary results on speaker-dependent variations in the TIMIT database. J. acoust. Soc. Am. *92*: 593 - 596 (1992).

Byrd, D.: Relations of sex and dialect to reduction. Speech Commun. *15*: 39-54 (1994).

Chen, M.: Vowel length variation as a function of the voicing of the consonant environment. Phonetica *22*: 129-159.

Edwards, J.; Gibbon, F.; Fourakis, M.: On discrete changes in the acquisition of the alveolar/velar stop consonant contrast. Lang. Speech *40*: 203 - 210 (1997).

Eguchi, S.; Hirsh, I.J.: Development of speech sounds in children. Acta oto-lar. 257: suppl., pp. 5 - 51 (1969).

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

Gilbert, J.H.V.; Purves, B.A.: Temporal constraints on consonant clusters in child speech production. J. Child Lang. *4*: 413 - 432 (1977).

Haggard, M.P.; Ambler, S.; Callow, M.: Pitch as a voicing cue. J. acoust. Soc. Am. *47*: 613 - 617 (1970).

Hasek, C.; Singh, S.; Murray, T.: Acoustic attributes of children's voices. J. acoust. Soc. Am. *68*: 1262 - 1265 (1980).

Henton, C.G.: A comparative study of phonetic sex-specific differences across languages; D.Phil. thesis. Oxford University (1985).

Henton, C.G.: Cross-language variation in the vowels of female and male speakers. Proc. 13th ICPhS, vol. 4, pp. 420-423 (Stockholm 1995).

Henton, C.G.; Bladon, R.A.W.: Breathiness in normal female speech, inefficiency versus desirability. Language and Communication *5*: 221 - 227 (1985).

Phonetica, 58, 196-210.

Hirose, H.; Gay, T.: The activity of the intrinsic laryngeal muscles in voicing control. Phonetica *25*: 140 - 164 (1972).

Jongman, A.; Blumstein, S.E.: Acoustic properties for dental and alveolar stop consonants, a cross-language study. J. Phonet. *13*: 235 - 251 (1985).

Kent, R.D.: Anatomical and neuromuscular maturation of the speech mechanism, evidence from acoustic studies. J. Speech Hear. Res. *19*: 421 - 447 (1976).

Kewley-Port, D.; Preston, M.S.: Early apical stop production, a voice onset time analysis. J. Phonet. *2*: 194 - 210 (1974).

Klatt, D.H.: Voice onset time, friction, and aspiration in word-initial consonant clusters. J. Speech Hear. Res. *18*: 686 - 706 (1975).

Krull, D.: VOT in spontaneous speech and in citation form words. PERILUS 12, pp. 101-107 (Stockholm University, Stockholm 1991).

Linke, C.E.: A study of pitch characteristics of female voices and their relationship to vocal effectiveness. Folia Phoniatrica *25*: 173 - 185 (1973).

Lisker, L.: Closure duration and the intervocalic voiced-voiceless distinction in English. Language *33*: 42 - 49 (1957).

Lisker, L.; Abramson, A.S.: A cross-language study of voicing in initial stops, acoustic measurement. Word *20*: 384 - 422 (1964).

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

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

Phonetica, 58, 196-210.

Lisker, L.; Liberman, A.M.; Erickson, D.M.; Dechovitz, D.; Mandler, R.: On pushing the voice onset time (VOT) boundary about. Lang. Speech *20*: 209 - 216 (1977).

Macken, M.S.; Barton, D.: The acquisition of the voicing contrast in English, a study of voice onset time in word-initial stop consonants. J. Child Lang. *7*: 41 - 74 (1979).

Mattingly, I.G.: Speaker variation and vocal tract size. J. acoust. Soc. Am. 39: S1219A (1966).

Ohde. R.N.: Fundamental frequency correlates of stop consonant voicing and vowel quality in the speech of preadolescent children. J. acoust. Soc. Am. *78*: 1554 - 1561 (1985).

Repp, B.H.: Relative amplitude of aspiration noise as a cue for syllable-initial stop consonants. Lang. Speech *22*: 173 - 189 (1979).

Ryalls, J.; Zipprer, A.; Baldauff, P.: A preliminary investigation of the effects of gender and race on voice onset time. J. Speech Hear. Res. *40*: 642 - 645 (1997).

Sander, E.: When are speech sounds learned? J. Speech Hear. Dis. *37*: 55 - 63 (1972).

Sawashima, M.; Abramson, A.S.; Cooper, F.S.; Lisker, L.: Observing laryngeal adjustments during running speech by use of a fibreoptics system. Phonetica *22*, 193 - 201 (1970).

Smith, B.L.: Temporal aspects of English speech production, a developmental perspective. J. Phonet. *6*: 37 - 67 (1978a).

Smith, B.L.: Effects of place of articulation and vowel environment on voiced stop consonant production. Glossa *12*: 163 - 175 (1978b).

Stoddart, J.; Upton, C.; Widdowson, J.D.A.: Sheffield dialect in the 1990s: revisiting the concept of NORMs: in Foulkes, Docherty, Urban Voices, pp. 72-89 (Arnold, London 1999).

Swartz, B.L.: Gender differences in voice onset time. Percept. Motor Skills *75*: 983 - 992 (1992).

Phonetica, 58, 196-210.

Sweeting, P.M.; Baken, R.J.: Voice onset time in a normal-aged population. J. Speech Hear. Res. *25*: 129-134 (1982).

Tingley, B.M.; Allen, G.D.: Development of speech timing control in children. Child Dev. *46*: 186 - 194 (1975).

Van Bezooijen, R.: Sociocultural aspects of pitch differences between Japanese and Dutch Women. Lang. Speech *38*: 253-265 (1995).

Whiteside, S.P.; Irving, C.J.: Speakers' sex differences in voice onset time, some preliminary findings. Percept. Motor Skills *85*: 459-463 (1997).

Wolf, C.G.: The perception of stop consonants by children. J. exp. Child Psychol. *16*: 318 - 331 (1973).

Zlatin, M.A.; Koenigsknecht, R.A.: Development of the voicing contrast: Perception of stop consonants. J. Speech Hear. Res. *18*: 541-553 (1975).

Zlatin, M.A.; Koenigsknecht, R.A.: Development of the voicing contrast: a comparison of voice onset time in stop perception and production. J. Speech Hear. Res. *19*: 93 - 111 (1976).

Phonetica, 58, 196-210.

Table 1. Mean and standard deviation values (in parentheses) for age (in months), height (in cm) and weight (in kg) for the 30 children who participated in the study

Age Group	Sex	Age in months	Height in cm	Weight in kg
7 years	Males	81.4 (3.4)	124.6 (6.0)	23.2 (2.2)
7 years	Females	81.4 (2.6)	131.0 (5.5)	27.8 (2.7)
9 years	Males	103.2 (2.5)	137.2 (7.9)	30.0 (7.1)
9 years	Females	101.4 (2.6)	133.4 (3.8)	28.0 (4.7)
11 years	Males	127.2 (5.0)	145.6 (9.7)	33.8 (5.3)
11 years	Females	128.0 (2.4)	150.6 (7.1)	36.9 (8.2)

Phonetica, 58, 196-210.

Table 2. Mean and standard deviation voice onset time values (in ms) for /p b t d/ for the 7-, 9- and 11-year olds by sex

Plosive	Age 7 years		Age 9 years		Age 11 years	
	Male	Female	Male	Female	Male	Female
/p/						
Mean	63.0	84.2	76.9	72.3	47.4	83.8
SD	20.9	28.3	30.4	20.6	11.9	23.9
/b/						
Mean	10.9	13.8	15.4	15.9	15.5	14.2
SD	4.5	6.5	4.1	6.5	4.9	4.1
/t/						
Mean	96.3	96.0	109.6	100.4	89.0	108.4
SD	15.3	17.6	32.0	21.4	14.9	25.4
/d/						
Mean	27.5	28.6	33.9	24.4	20.6	22.7
SD	10.3	6.5	18.7	9.2	4.8	7.8

Phonetica, 58, 196-210.

Table 3. Mean and standard deviation values (in ms) for VOT voiceless-voice difference (VOTVVD) for the cognate pairs /p/ and /b/, /t/ and /d/, by age and sex.

Plosive Pairs	Age 7 years		Age 9 years		Age 11 years	
	Male	Female	Male	Female	Male	Female
VOTVVD /p b/						
Mean	52.2	70.6	64.3	56.9	33.9	69.7
SD	18.9	23.3	29.6	17.9	9.1	23.8
VOTVVD /t d/						
Mean	68.6	67.4	74.9	76.0	68.4	85.7
SD	12.8	13.4	19.3	16.3	12.6	20.0

Figure 1. Mean VOT values (in ms) for /p/ by sex and age.

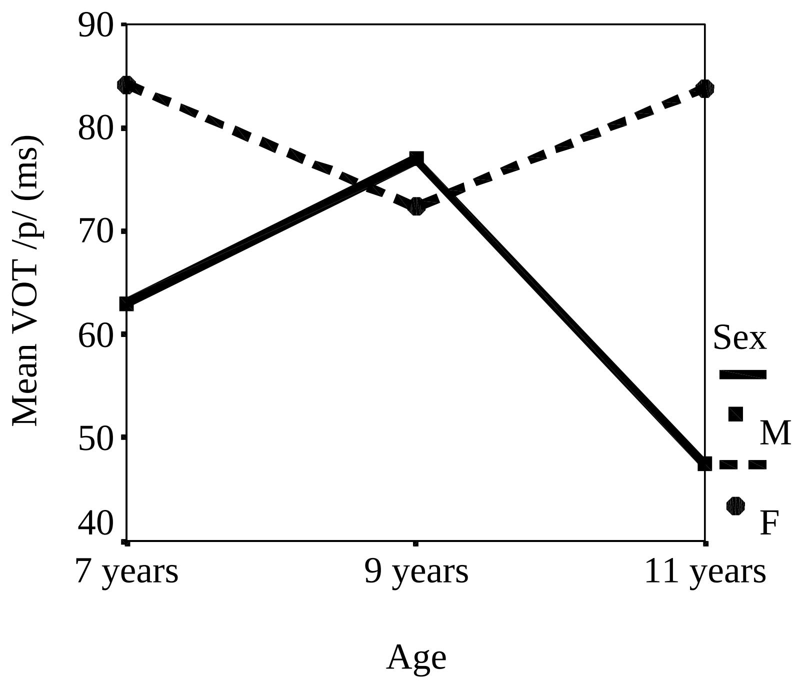

Figure 2. Mean VOT values (in ms) for /b/ by sex and age.

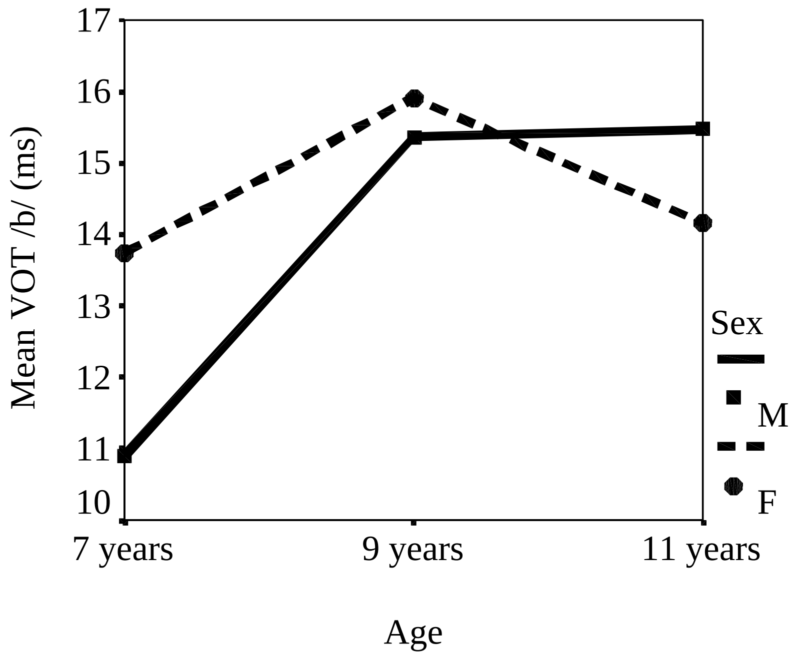

Figure 3. Mean VOT values (in ms) for /t/ by sex and age.

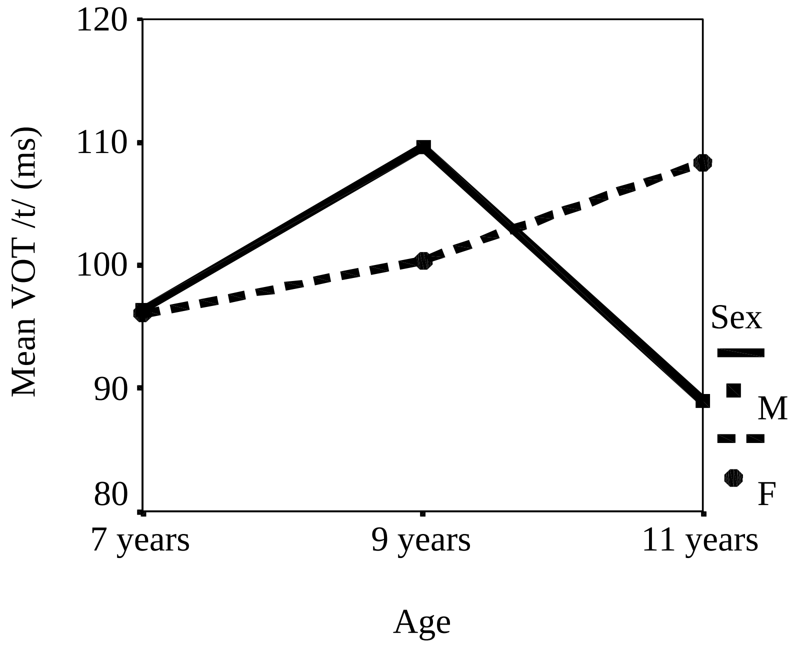

Figure 4. Mean VOT values (in ms) for /d/ by sex and age.

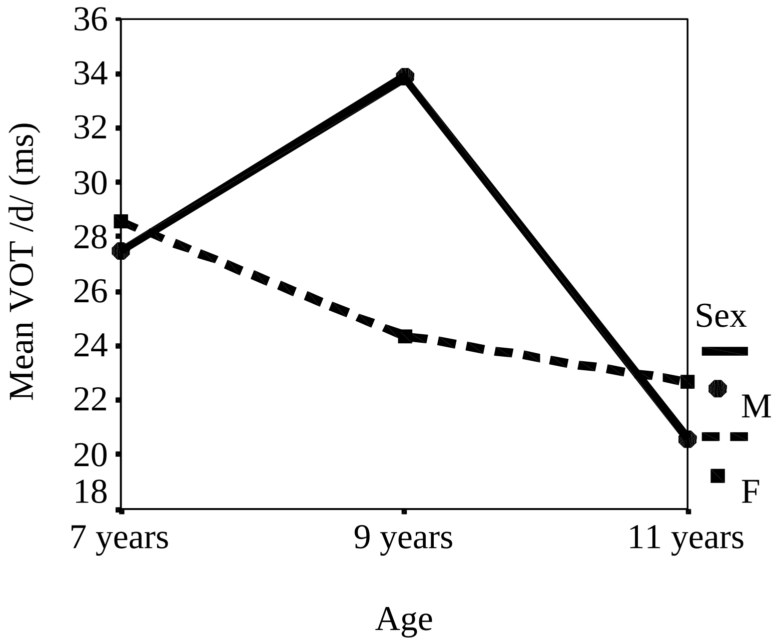

Figure 5. Mean values for VOT voiceless-voice difference (VOTVVD - in ms) for the cognate pair /p/ and /b/ by age and sex.

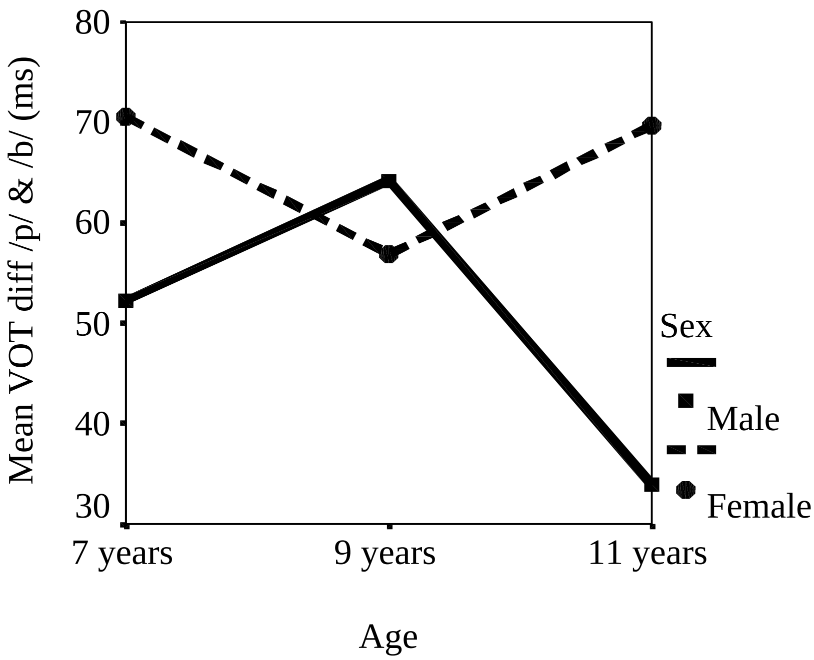

Figure 6. Mean values for VOT voiceless-voice difference (VOTVVD - in ms) for the cognate pair /t/ and /d/, by age and sex.