Voice onset time in Swedish children and adults

Inger Lundeborg Hammarström, Maria Larsson, Sara Wiman and Anita McAllister

Linköping University Post Print

N.B.: When citing this work, cite the original article.

Original Publication:

Inger Lundeborg Hammarström, Maria Larsson, Sara Wiman and Anita McAllister, Voice onset time in Swedish children and adults, 2012, Logopedics, Phoniatrics, Vocology, (37), 3, 117-122.

<http://dx.doi.org/10.3109/14015439.2012.664654>

Copyright: Informa Healthcare

<http://informahealthcare.com/>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-85082>

Abstract

Voice onset time (VOT) is a temporal acoustic parameter, which reflects the timing of speech motor control. The objective of the work was to obtain normative VOT-data in Swedish children. Thus, 150 children aged 8-11 years old and 36 adults were audio-recorded when producing the plosives in minimal pairs. Measures were made using waveforms and spectrograms. Results show that Swedish children developed adult-like VOT-values between 9 and 10 years. By the age of ten, prevoicing was also found fully adult like in extent. The results indicate that all Swedish adults do not produce voiced plosives with prevoicing. No evident gender differences were found. The obtained VOT-values can be used as normative data when assessing children with speech and language disorders.

Introduction

Speech is one of human's most complex motor behaviours and requires spatial and temporal control of more than 70 muscles (Cheng, Murdoch, Goozee and Scott, 2007) ranging from the respiratory muscles through the vocal apparatus to the oral cavity. Thus the developing child needs many years to reach adult mastery of fluent speech. Most Swedish children master the realization of phonemic contrasts of the Swedish language by the age of seven years but continue to develop the precision of the phonetic realizations for another couple of years (Nettelbladt, 2007). Different speech sounds require different degrees of motor coordination and control. Plosives are among those which require close coordination between the larynx and the lips, tongue and jaws (Auzou, Özsancak, Morris, Jan, Eustache and Hannequin, 2000). In Swedish there are six plosives, /p t k/ which are classified as voiceless, and /b d g/ as their respective voiced equivalents. When producing the voiceless plosives in word-initial position, the vocal folds are in an open position during the oral closure phase the velum is closed and the intraoral pressure increases rapidly. When the oral contact is released, a rapid

airflow results via the mouth. This sudden change in airflow creates an acoustic pressure transient burst of acoustic energy. The vocal folds then begin to approximate to begin the glottal vibration of the following vowel. When the voiced plosives are produced the vocal folds are relatively closed already at the release of the oral closure (Hertegard and Gauffin, 1995). There are conflicting results in studies regarding whether the Swedish voiced plosives exhibit prevoicing, meaning that voicing begins before the release. Some reports that investigated subjects did not have prevoicing (Keating, Linker and Huffman, 1983), whereas others report that prevoicing does occur for all or almost all subjects (Karlsson, Zetterholm and Sullivan, 2004; Helgason and Ringen, 2008). In the clinical setting the central tool for the classification of the voicing aspect of plosives is auditory-perceptual evaluation. However, acoustic analysis can provide insights into error patterns that go beyond perceptual judgments (Ballard and Robin, 2002). The acoustic cue which is considered most reliable for the distinction between voiceless and voiced plosives is voice onset time (VOT) (Auzou, Ozsancak et al., 2000; Whiteside and Marshall, 2001; Helgason and Ringen, 2008). VOT is defined as the time between the onset of the release of the oral closure and the onset of vocal fold vibration and is measured in milliseconds (Lisker and Abramson, 1964; Auzou, Ozsancak et al., 2000; Whiteside and Marshall, 2001; Grigos, Saxman and Gordon, 2005; Helgason and Ringen 2008). Since the vocal fold activity begins after the release of the oral constriction in voiceless plosives, a positive value of the VOT is obtained. If aspiration is present, which is the case for /p t k/ in word initial position in Swedish, the VOT-value increases (Karlsson, Zetterholm et al., 2004). Typical adult values for the voiceless plosives are between 49 and 78 milliseconds and between -91 and -61 milliseconds for the voiced plosives. In babbling voiceless aspirated sequences are rare (Davies and MacNeilage, 1995; Whalen, , Levitt, and Goldstein, 2007). This may indicate that younger children are less able

to realize the temporal requirements of the abduction/adduction gestures for the production of these sounds.

For the voiced plosives the vocal folds vibrate before (prevoicing) or simultaneously with the burst and a negative or low VOT-value is obtained. The acquisition of adult-like VOT-patterns is gradual in the developing child. Several studies of English children display similar results regarding when this development is completed which seems to be around the age of 11 years, when stable adult-like VOT-values are observed (Nitttrouer 1993; Whiteside and Marshall 2001; Whiteside, Dobbin and Henry, 2003). Larsson and Wiman, (2010) measured the VOT of 83 Swedish children aged 3, 4 and 5 years, and a developmental trend was found, but no child had achieved adult-like VOT patterns. The 3-year olds had the longest mean VOT. Also the 3-year old children lacked prevoicing (figure 1).

Figure 1. Prevalence of prevoicing in 3 to 5 year old children. Data from Larsson and Wiman, 2010.

VOT is thought to reflect the co-ordination and timing between laryngeal and oral articulation. In verbal dyspraxia, a speech disorder denoted by speech distortions particularly influencing timing and coordination, the effects of disrupted interarticulatory co-ordination is

evident. Thus the coordination and timing of more than one articulatory sub-system is required. Since voiced speech involves more than one sub-system this phenomenon is regarded as more complicated than co-articulation (Lubker, McAllister and Lindblom, 1977; Löfqvist, 1980). The planning of motor commands in speech is thought to be articulator and not muscle specific (van der Merve, 2009). The few reports of VOT in individuals with verbal apraxia that are published are of single adult subjects indicating trouble with VOT production (Auzou, Ozsancak et al., 2000). In order to be able to include VOT-measures when dealing clinically with verbal dyspraxia, reliable norm values are necessary. No extensive study of VOT in typically developed Swedish children and adults have been conducted and there is a need for norm values. The aim of the present study is to investigate and compare VOT patterns among Swedish children (8, 9, 10 and 11 years) and adults to study the development of VOT and to obtain Swedish normative data.

Method

Subjects

One-hundred and fifty children in four age groups; 7.9-8.8 years (mean age 8 y 3 m), 8.9-9.8 years (mean age 9 y 2 m) , 9.9-10.8 (mean age 10 y 1m), 10.9-11.8 (mean age 11 y 3 m) and 36 adults (mean age 32 y 9 m) participated in the study (Table I). The children were recruited from schools in a town with approximately 42 000 inhabitants in the south east of Sweden. The adults, 18 males and 18 females aged 22-59 years (mean age 32 y 9 m), were recruited at workplaces and sport clubs from the same town as the children and from another town with about 130 000 inhabitants in the same geographical area (Table I). Children and adults with known cognitive deficits, phonological problems, hearing loss and/or non-native Swedish speakers were excluded as well as children undergoing mutational voice change. The study

was carried out in accordance with the ethical principles for medical research of the Helsinki declaration as revised in 2008 (WMA Declaration of Helsinki –Ethical Principles for Medical Research Involving Human Subjects).

Table I Number of participants and mean age (SD) in each group

	♂+♀		♀		♂	
	n	mean age (SD)	n	mean age (SD)	n	mean age (SD)
<i>8 years</i>	32	8.03 (3.4m)	16	8.4 (3.8m)	16	8.2 (2.8m)
<i>9 years</i>	39	9.02 (3.3m)	23	9.02 (3.4m)	16	9.1 (3.1m)
<i>10 years</i>	43	10.01 (3.4m)	20	10.00 (3.4)	23	10.1 (3.5)
<i>11 years</i>	36	11.03 (3.6m)	15	11.02 (3.4m)	21	11.3 (3.7m)
<i>Adults</i>	36	32.09 (11y4m)	18	33.09 (12y6m)	18	31.8 (10y3m)

Speech stimuli

The subjects produced two representations of all plosives occurring in Swedish ; /p/, /b/, /t/ /d/ /k/ and /g/ using minimal pairs, pil-bil (arrow-car), tennis- Dennis (tennis-Dennis) and kula-gula (ball-yellow). The first production was elicited in the form of sentence completion in order to establish the chosen words. In the second production the words were elicited by picture naming, one by one. None of the words in the minimal pairs were presented immediately following each other. The speech samples were recorded on a Marantz PMD 660 Professional Recorder with a sampling frequency of 44.1 kHz and an Audio-Technica MB 3k unidirectional microphone on a stand directed to the speaker at a distance of approximately 30 cm (11.8 inches).

Measures

The second production of each word was analysed using the Praat software (<http://www.fon.hum.uva.nl/praat/>, version 5.1.31, Paul Boersma and David Weenik, Phonetic Sciences Department, University of Amsterdam) through which waveforms and spectrograms

were generated and displayed. The VOT measurements were made according to the method described by Lisker and Abrahamson (1964) by measuring the distance between the transient burst of the stop to the first positive zero-crossing, introducing periodicity of the following vowel. VOT is expressed in milliseconds (ms). For additional support of the analysis the spectrograms were used and the measurements taken from the release of the plosive (marked by a transient in the spectrogram) to the onset of low-frequency periodical activity.

Figure 2. Percentage of individuals having prevoicing when producing /b d g/.

Statistical analyses

Demographic data were expressed with descriptive statistics. Differences in VOT between the age groups were analysed using one-way ANOVA with LSD post hoc test. Chi-2 test were used in order to analyse differences in occurrence of prevoicing between the age groups. Gender differences were analysed with t-test. Differences in values between the plosives were analysed One-way ANOVA. P-values <0.05 were considered statistically significant. Regarding the voiced plosives, descriptive statistics were used for negative and positive VOT

separately, but when comparing VOT-values of the different age-groups positive and negative VOTs have been analysed together. All measures were performed by two of the authors (M.L. and S.W.) and the values were mutually decided.

Results

Reliability

Based on a reanalysis of a random selection of 10% of the data intra-rater reliability was calculated using Chronbach's alpha. Intra-rater reliability was found to be very good with an alpha value of 0.99.

The voiceless plosives

All age groups had a VOT with a voicing lag (positive VOT) when producing the voiceless plosives. The values increased as the place of articulation moved posteriorly (/p/ < /t/ < /k/), and the values differed significantly. The mean VOT-values for all the voiceless plosives /p t k/ decreased with increasing age among the children and the standard deviations were higher in the two youngest age groups. For /p/ the difference between the age groups 9 and 11 years was statistically significant and for /t/ the differences between the age group 8 years and all the other groups except the 9 year olds were statistically significant. Regarding /k/ the same differences between other age groups as for /t/ were statistically significant and significant differences were also found between the 9 year olds and 10 and 11 year old children (Table II). No sex differences were found.

Table II. Mean (standard deviation) for VOT in ms for /p t k/ in 8-, 9-, 10-, and 11-year olds and adults

<i>age</i>	<i>/p/¹</i>	<i>/t/¹</i>	<i>/k/¹</i>
8 vs 9 years	49.77 (12.67) / 50.44(14.51)	60.79 (13.90)/ 55.84 (13.6)	77.93 (17.58)/ 72.55 (12.05)
8 vs 10 years	49.77 (12.67) / 46.17 (8.09)	60.79 (13.90)/ 55.79 (8.37)*	77.93 (17.58)/ 66.92 (8.78)***
8 vs 11 years	49.77 (12.67) / 44.76 (9.52)	60.79 (13.90)/ 52.79 (9.45) **	77.93 (17.58)/ 66.41 (12.80)***
8 vs Adults	49.77 (12.67) / 46.29 (9.73)	60.79 (13.90)/ 53.82 (9.21)**	77.93 (17.58)/ 70.06 (11.89)**
9 vs 10 years	50.44 (14.51)/ 46.17 (8.09)	55.84 (13.60)/ 55.79 (8.37)	72.55 (12.05)/ 66.92 (8.78)*
9 vs 11years	50.44(14.51)/ 44.76 (9.52)*	55.84 (13.60)/ 52.79 (9.45)	72.55 (12.05)/ 66.41 (12.80)*
9 vs adults	50.44(14.51)/ 46.29 (9.73)	55.84 (13.60)/ 53.82 (9.21)	72.55 (12.05)/ 70.06 (11.89)
10 vs 11 years	46.17 (8.09)/ 44.76 (9.52)	55.79 (8.37)/ 52.79 (9.45)	66.92 (8.78)/ 66.41 (12.80)
10 vs adults	46.17 (8.09)/ 46.29 (9.73)	55.79 (8.37)/ 53.82 (9.21)	66.92 (8.78)/ 70.06 (11.89)
11 vs adults	44.76 (9.52)/ 46.29 (9.73)	52.79 (9.45)/ 53.82 (9.21)	66.41 (12.80)/ 70.06 (11.89)

¹ Mean±SD *p<0.05 **p<0.01 ***p<0.001

The voiced plosives

In all age groups individuals with no prevoicing in their production of voiced plosives were found, but the proportion of these individuals decreased with increasing age (Figure 2). With more anterior place of articulation a higher percentage of individuals with prevoicing were found (/b/ 68%, /d/ 58% and /g/ 50.5%), se figure 3

Figure 3. Mean negative VOT-values in ms for the voiced plosives across age-groups.

The mean negative VOT-values for /b/ ranged from -120.88 (8 year old group) to -81.67 (10 year old group). For /d/ the range was from -90.69 (8 year old group) to -71.62 (11 year old group). For /g/ the range was from -72.62 (8 year old group) to -54.59 (10 year old group), see figure 4.

Note the increased mean negative VOT in the adults compared to the 10 and 11 year old groups. The statistical analyses of the age-group differences showed that the two youngest age groups differed significantly from the older age groups and adults. This was especially evident for /d/ (Table III). The older age-groups had adult-like values. With the exception of a

longer negative VOT for /g/ in the boys of aged 11 years ($p < 0.05$) no sex differences were found.

Discussion

This results of this study show that Swedish children have developed adult-like VOT-values somewhere between 9 and 10 years. This is in line with the findings in studies of English-speaking children (Nittrouer 1993; Whiteside and Marshall 2001; Whiteside, Dobbin et al. 2003).

All subjects had a marked difference in VOT between the unvoiced and voiced plosives with longer VOTs for the unvoiced plosives. The range of mean VOT-values for the different age-groups was between +44.76 and +77.93 milliseconds for the unvoiced plosives. This range is above what is considered to be necessary for the perception of the distinction between unvoiced and voiced plosives (Zlatin and Koenigsknecht, 1976). The range for the voiced plosives was between -72.73 and +1.56 milliseconds. In the present material a relatively large proportion of the adults produced the voiced plosives without prevoicing (36% for d and g

Table III. Comparison of merged VOT- values in ms (both positive and negative) between all combinations of age groups 8, 9, 10, 11 and adults.

Age groups	Both positive and negative VOT ¹		
	/b/	/d/	/g/
8 vs 9 y	-38.57±67.63/ -39.77±58.47	-19.58±53.91/ -12.15±49.2	-9.54±48.31/ +1.56±45.16/
8 vs 10 y	-38.57±67.63/ -62.34±44.23	-19.58±53.91/ -47.07±50.28*	-9.54±48.31/ -21.14±41.47*
8 vs 11 y	-38.57±67.63/ -72.73±38.16**	-19.58±53.91/ -61.91±34.34***	-9.54±48.31/ -37.24±43.51***
8 vs adults	-38.57±67.63/ -69.43±57.6*	-19.58±53.91/ -45.37±56.33*	-9.54±48.31/ -35.23±49.95
9 vs 10 y	-39.77±58.47/ -62.34±44.23	-12.15±49.02/ -47.07±50.28**	+1.56±45.16/ -21.14±41.47
9 vs 11 y	-39.77±58.47/ -72.73±38.16**	-12.15±49.02/ -61.91±34.34***	+1.56±45.16/ -37.24±43.51
9 vs adults	-39.77±58.47/ -69.43±57.6*	-12.15±49.02/ -45.37±56.33**	+1.56±45.16/ -35.23±49.95
10 vs 11 y	-62.34±44.23/ -72.73±38.16	-47.07±50.28/ -61.91±34.34	-21.14±41.47/ -37.24±43.51
10 vs adults	-62.34±44.23/ -69.43±57.6	-47.07±50.28/ -45.37±56.33	-21.14±41.47/ -35.23±49.95
11 vs adults	-72.73±38.16/ -69.43±57.6	-61.91±34.34/ -45.37±56.33	-37.24±43.51/ -35.23±49.95

¹ Mean±SD *p<0.05 **p<0.01 ***p<0.001

and 25% for b). This is in accordance with previous studies of Swedish adult speakers where Keating and colleagues (1983) found that speakers did not have prevoicing, whereas Karlsson and colleagues (2004) and Helgasson and Ringen (2008) found prevoicing in all or almost all subjects. This may indicate that there are two VOT-categories for the Swedish voiced plosives. Could these different findings be attributed to regional dialects?

There was also a clear developmental trend in the studied age-groups both regarding length of VOT for the voiceless plosives and for incidence of prevoicing in the voiced plosives. This

is in line with the results in the previous study of younger children using the same speech material and methodology (Larsson and Wiman, 2010).

At the age of 11 years-old almost 90% of the children have acquired prevoicing for the voiced plosives /b/ and /d/ which is even higher than in the adult population (67.6%). Perhaps this is a reflection of the same phenomena as clinicians often have noted in newly acquired speech skills in children. These children often have a tendency to produce the new skill in an

exaggerated way, such as the tremulant /r/ often being both emphasized and added to almost all words. In studied children the 8 years-old children showed a lowest incidence of prevoicing compared to the older age groups. Also, when compared to 3-5 year old children in a previous study the the 3 year old children did not exhibit this feature at all and the 4 and 5 years old children only to a minor degree (Larsson and Wiman, 2010). So why do the younger children fail to produce the prevoicing observed in a majority of the children 10 and 11 year-olds? Numerous studies of infant speech perception have shown that infants as young as 2 months of age can discriminate syllables beginning with a short-lag versus a long-lag VOTs (Eimas, Siqueland, Jusczyk and Vigorito, 1971; Werker and Tees, 1999). This indicates that the present result is not due to children's perceptual capacities. Instead, the inability to produce the prevoicing feature seems to represent production and coordination constraints related to interarticulatory co-ordination lasting well into the school years. This is a co-articulation phenomenon but involving more than one articulatory sub-system. Voiced speech requires coordination of more than one sub-system. This study indicates that the mastering of this complex activity is not achieved until around 10 years of age. The drop in number of individuals with prevoicing at 11 years-old to adults may be attributed to different dialects. The number of years living in the region was not controlled for in the present study. In order to control for dialectal influences a multicenter study is required. It would also be interesting to follow the development of VOT in individuals into adult life.

There are however some differences related to the different sounds. For the voiceless plosives it seems that with a more anterior place of articulation adult like values are faster developed. The same tendency was seen in the development of prevoicing. In adults the highest proportion of prevoicing was found for the bilabial plosive /b/ which also had the highest

negative VOT value. This is in accordance with results in the study by Helgason and Ringen (Helgason and Ringen 2008).

The variability of VOT-values in the different age groups decreased with increased age.

Several studies of other languages show the same tendency with higher standard deviations in children when compared to adults (Koenig, 2001; Okalidou, Petinou, Theodorou and Karasimou, 2010).

No evident sex differences were seen neither for the children nor for the adults. In other Swedish VOT-studies no sex differences have been observed regarding the voiceless plosives, but when producing voiced plosives, prevoicing was seen to a greater extent in males (Karlsson, Zetterholm et al. 2004; Helgason and Ringen 2008). These studies did not include more than a few subjects whereas the present study included 186 participants.

Conclusion

Typically developing Swedish children develop adult-like VOT values for the voiceless plosives at the age of 9 years and for the voiced plosives at the age of 10 years. The data also indicate that there are two VOT-categories for the Swedish voiced plosives; either with prevoicing, which is most common or with a short voice lag.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

- Auzou, P., Özsancak, C., Morris, R.J., Jan, M., Eustache, F. and Hannequin, D. (2000), Voice onset time in aphasia, apraxia of speech and dysarthria: a review. *Clinical Linguistics and Phonetics* 14(2): 131-150.
- Ballard, K. J. and Robin, D. A. (2002), Assessment of AOS for treatment planning. *Seminars in Speech and Language* 23(4): 281-292.

Cheng, H. Y., Murdoch, B. E., Goozee, J.V. and Scott, D. (2007), Physiologic development of tongue-jaw coordination from childhood to adulthood. *Journal of Speech, Language, and Hearing Research* 50(2): 352-360.

Davies, B and MacNeilage, P. (1995), The articulatory basis of babbling. *Journal of Speech and Hearing Research* 38: 1199-1211

Eimas, P.D., Siqueland, E.R., Jusczyk, P. and Vigorito, J. (1971), Speech perception in infants. *Science* 22: 303-306.

Grigos, M.I., Saxman, J.H. and Gordon, A.M. (2005), Speech motor development during aquisition of the voicing contrast. *Journal of Speech, Language and Hearing Research* 48: 739-752.

Helgason, P. and C. Ringen (2008), Voicing and aspiration in Swedish stops. *Journal of Phonetics* 36: 607-628.

Hertegard, S. and J. Gauffin (1995), Glottal area and vibratory patterns studied with simultaneous stroboscopy, flow glottography, and electroglottography. *Journal of Speech and Hearing Research* 38(1): 85-100.

Karlsson, F., E. Zetterholm, and Sullivan, K.P.H. (2004), Development of a gender difference in voice onset time. *The Journal of the Acoustical Society of America* 116: 1179-1183.

Keating, P., Linker, W and Huffman, M. (1983), Patterns in allophone distribution for voiced and unvoiced stops. *Journal of Phonetics* 11: 277-290.

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* 43: 1211-1228.

Larsson, M. and Wiman S. (2010), Voice onset time hos svenska förskolebarn -Ett utvecklingsperspektiv (Voice onset time in Swedish preschool children - a developmental perspective). Kandidatuppsats i logopedi (bachelor thesis), Linköping University
<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-55537>

Lisker, L. and Abramson, A. S. (1964), A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20(4): 384-422.

Lubker, J., McAllister, R. and Lindblom, B. (1977), Vowel fundamental frequency and tongue height. *Journal of Acoustical Society of America* 62: S16-S17

Löfqvist, A. (1980), Interarticulator programming in stop production. *Journal of Phonetics* 8:475-490

van der Merve A. (2009), A theoretical framework for the characterization of pathological speech sensorimotor control. In ed McNeil M R. *Clinical Management of sensorimotor speech disorders (pp.9-13)* Thieme Publisher Inc. New York

Nettelbladt, U. (2007). Fonologisk utveckling (Phonological development)..In U. Nettelbladt and E.-K. Salameh (Ed.), *Språkutveckling och språkstörning hos barn (Language development and language impairment in children (p. 79)* Lund, Studentlitteratur: 79-80.

Nittrouer, S. (1993), The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech and Hearing Research* 36: 959-972.

Okalidou, A., Petinou, K., Theodorou, E. and Karasimou, E. (2010), Development of voice onset time in standard Greek and Cypriot-Greek-speaking preschoolers. *Clinical Linguistics and Phonetics* 24(7): 503-519.

Werker, J.F. and Tees, R.C. (1999), Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology* 50: 509-535

Whalen, D. H., Levitt, A. G. and Goldstein, L. M. (2007), VOT in the babbling of French- and English-learning infants *Journal of Phonetics* 35 (3): 341-352

Whiteside, S. P., Dobbin, R and Henry, L. (2003), Patterns of variability in voice onset time: A developmental study of motor speech skills. *Neuroscience Letters* 347: 29-32.

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

World Medical Association Declaration of Helsinki, 2008 (Cited 2011-10-31) Available at <http://www.wma.net/en/30publications/10policies/b3/>.

Zlatin, M.A. and Koenigsknecht, R.A. (1976), Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech and Hearing Research* 19:93-111