THE EFFECTS OF LIP ROUNDING ON VOICE ONSET TIME PRODUCTION IN CHILDREN ACQUIRING HUNGARIAN

Krisztina Zajdó

Department of Special Education/Speech-Language Pathology
The University of West Hungary, Győr, Hungary
zajdo@hotmail.com

ABSTRACT

The goal of this project was to examine the effects of lip rounding on the production of voice onset time in voiceless bilabial stops in children acquiring Hungarian as their first language. The duration of voice onset time in stops followed by or embedded between high back rounded phonologically long vowels with prominent lip protrusion are considerably longer. VOT durations in stops followed or surrounded by phonologically short front mid rounded vowels with less lip protrusion or high front unrounded vowels with no lip protrusion are shorter. Developmental trends in VOT acquisition are discussed.

Keywords: VOT (voice onset time), lip rounding, acquisition, motor control, bilabial unaspirated voiceless stops.

1. INTRODUCTION

Language communities around the world utilize different voice onset time features when differentiating language-specific stop sounds. Voice onset time is defined as the duration of time between the release of a stop consonant and the onset of voicing [3]. Among the 451 languages currently listed in the UCLA Phonological Segment Inventory Database [1, 2], 375 languages (83.15% of all languages listed) use bilabial voiceless plosives differing in VOT across speaker communities. Recent results suggest that the three major categories of VOT established in cross-linguistic research (long lead, short lag, long lag) may not be fine-grained enough to account for the variation encountered in different language communities [4].

Hungarian uses a voiceless unaspirated plosive /p/ which is characterized by an average voice onset time of 18.51 ms (range: 9.1-28.8, SD: 5.92) in spontaneous speech, and 24.64 ms (range: 13.2-34.8, SD:11.95) in isolated words [5].

Previous results on mothers' child-directed speech [6] show that, when produced word medially in the word /pipi/ 'chick', mothers with children at 2;0, 3;0 and 4;0 years of age produce VOTs of 14.46 ms (SD:1.38), 17.46 ms (SD: 5.98), and 16.26 ms (SD: 3.90), respectively. Thus, mothers' speech modeling behavior remains relatively unchanged during their children's speech acquisition between the ages of 2;0 to 4;0 years. No statistically significant differences were detected in the durational properties of VOT in mothers' speech. Throughout this period of their children's development, mothers model VOT speech timing patterns with considerable consistency. In contrast with data from adults and mothers using 'motherese', the acquisition process of language-specific VOT production in Hungarian-speaking children remains largely unexplored.

It was proposed [7] that, among all VOT types, the voice onset time of the short unaspirated voiceless plosive /p/ is probably the easiest to acquire. Cross-linguistically, this type of stop sound is acquired earliest [8, 9, 10, 11, 12, 13, 14, 15, 16, 17]. To investigate developmental trends in VOT production, the effect of a motorically challenging speech behavior, lip rounding, on VOT production in voiceless unaspirated bilabial plosives in children acquiring Hungarian was explored.

2. HYPOTHESES

Previous results [18, 19] show that between 2;0 and 4;0 years of age, children's production of words with rounded vowels are characterized by lower speech accuracy than those with unrounded vowels. Further, children's production of words with front rounded vowels are perceived to be less accurately produced cross-linguistically [20]. Thus, lip rounding appears to pose a challenge for children to execute motorically in running speech. Data from acoustic analyses [21, 22, 23] also reflect children's difficulty with producing vowels with appropriate lip protrusion, resulting in the production of less peripheral vowels in the acoustic vowel space which are perceived by adults as less accurate.

Children' speech movements have been shown to be slower in velocity than in adults [24]. Specifically, speech movements involving lip kinematics exhibited by 5;0 year olds are 50-70% slower than those of adults [25]. Overall, children's motor control capabilities require considerably more time to plan and execute speech movements than adults. It is hypothesized that slower velocity of movement impacts VOT values measured in children's speech. That is, children generate longer VOT durations than adults, especially when the plosive is followed by or embedded within rounded vowels. Further, it is hypothesized that biological maturation and increased skills in speech motor control are reflected by shorter, more adult-like VOT values in older as opposed to younger children.

3. METHODS

3.1. Participants and setting

Participants included 7 boys at the ages of 2;0, 3;0 and 4;0 years of age (n=21; for all details, see [18]). Subjects were monolingual native speakers of Hungarian with no history of speech, language or hearing problems. Each boy and his mother sat in a sound-attenuated room and played with puppets with the names /pipi/, /pøpø/ and /pu:pu:/ among other toys.

Caregivers were instructed to attempt to elicit the name of each puppet from their child at least 5 times during the production of a self-generated fairy tale. Puppets with names including short vs. long vowels were used in two different sessions. The two recordings were obtained within one week.

3.2. Recording attributes and data selection

The Sound Forge (Version 5.0, Built 117) acoustic recording and analysis computer program was used to record the data with recording attributes set to 32kHz, 16 bit, mono.

Data selection was carried out by a monolingual Hungarian speaker by using the Sound Forge (Version 9.0e, Build 441) program. Tokens with two accurately produced vowels and stops were selected perceptually when listening to the recordings via two high-quality speakers.

From each child, 5 tokens were selected from each of the three categories. A second judge rated 10% of the data set; inter-rater reliability exceeded 90% in each category of tokens. Tokens (n=10) with aspirated stops that are non-phonemic in Hungarian and tokens that were whispered (n=5) were excluded from the study.

3.3. Data analysis

VOT duration was measured by inspecting both the spectrogram and wave form of each token. The Praat program (Version 5.4.04, 32 bit edition [26]) was used for the acoustic analyses.

Markers were placed at the burst and at the beginning of the periodic waveform. VOT values were recorded in milliseconds, rounded to two decimal places.

4. RESULTS

4.1. Results for word-initial voiceless bilabial plosives VOTs

VOT values for the word-initial /p/ sounds are shown in Table 1 below. Each cell represents the average value of 5 productions from all 7 children.

Table 1: Descriptive statistics of the duration of VOT in the word-initial /p/ in /pV₁(:)pV₁(:)/ structured tokens. Values are shown in ms.

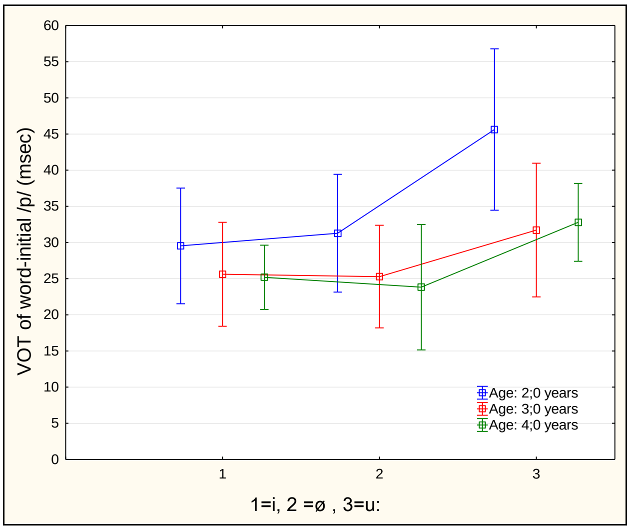
Vowel quality	Mean	Median	Minimum	Maximum	Range	Variance	St. Dev.	St. Error
Data from 2;0 years old boys (N=7)								
i	29.53	33.93	15.55	38.72	23.17	74.87	8.65	3.27
ø	31.28	28.08	20.97	47.29	26.31	77.48	8.80	3.33
u:	45.62	42.10	32.54	66.94	34.40	145.49	12.06	4.56
Data from 3;0 years old boys (N=7)								
i	**25.60**	**26.83**	**14.43**	**36.25**	**21.82**	**60.21**	**7.76**	**2.93**
ø	**25.29**	**26.14**	**14.57**	**35.54**	**20.97**	**58.78**	**7.67**	**2.90**
u:	**31.71**	**36.78**	**16.00**	**41.00**	**25.00**	**99.66**	**9.98**	**3.77**
Data from 4;0 years old boys (N=7)								
i	25.19	22.43	21.17	33.29	12.12	23.15	4.81	1.82
ø	23.82	23.57	13.86	37.50	23.64	87.93	9.39	3.54
u:	32.79	34.57	20.75	37.86	17.11	33.89	5.82	2.20

Results of a two-factor ANOVA test revealed significant main effects of age (F (2,54)=6.2174, p< .005) and the quality of the following vowel (F 2,54)= 9.3782, p< .001). No interaction effect was found (F(4,54)= .75908, p= .55649).

Post-hoc Tukey HSD tests revealed that the results for the 2;0 year olds (mean: 35.48, SD: 12.00) differed significantly both from those of the 3;0 (mean: 27.53, SD: 8.65) and the 4;0 year olds (mean: 27.27, SD: 7.73). Thus, VOT values of older children were shorter, as predicted, possibly indicating more mature speech planning and production skills between ages 2;0 and 4;0. However, data from the two older age groups did not differ at p< .05.

Post-hoc Tukey HSD tests indicated that the VOT values followed by the vowels /i/ and /ø/ did not differ from each other at p< .05, but that the VOTs of voiceless bilabial plosives followed by the vowel /u:/ (mean: 36.71, SD: 11.21) differed both from those followed by /i/ (mean: 26.77, SD: 7.18) and /ø/(mean: 26.80, SD: 8.84). These results are shown in Figure 1.

Figure 1: Mean plot of VOT values for word-initial /p/ grouped by following vowel quality categorized by age groups (in ms)



Overall, the VOT values of the word-initial /p/ indicate that children in the two older age groups produce stops with shorter VOT durations. Additionally, there are no significant differences between the VOT durations of stops followed by the front unrounded and rounded vowels. Further, it is noteworthy that the production of stops followed by longer VOTs in bilabials followed by phonologically long vowels is also shown in adult speech. Similarly, the trend to produce longer VOTs when the stop is followed by phonologically long vowels, even though exact VOT duration values were not reported, is also found in adult speech [5].

It is also notable that the most sizeable decrease in VOTs is shown in plosives followed by the back rounded vowel. In the case of this vowel, lip rounding is much more prominent (e.g., lips are much more protruded) than during the production of the front rounded vowel. Thus, motorically, the execution of the back rounded vowel appears to require more time, increasing the duration of VOT in the preceding stop sound. The average VOT duration is shorter by more than 10 milliseconds in the two older age groups. The shorter duration shows the effects of biological maturation and motor practice, resulting in the faster planning and execution required to produce the sound sequence that involves protruding the lips for the first vowel, then pulling it back to produce the bilabial, and then protruding the lips again to produce the second vowel.

4.2. Results for word-medial voiceless bilabial plosives VOTs

VOT values for the word-medial /p/ sounds are shown in Table 2 below. Each cell represents the average value of 5 productions from all 7 children.

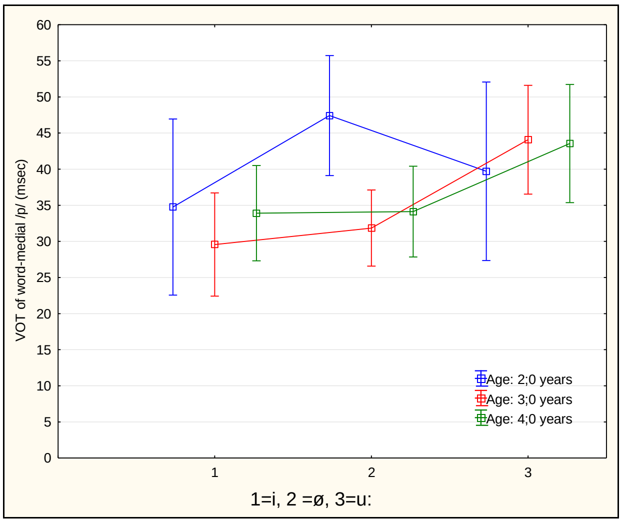
Table 2: Descriptive statistics of the duration of VOT in the word-medial /p/ in /pV_1(:)pV_1(:)/ structured tokens. Values are shown in ms.

Vowel quality	Mean	Median	Minimum	Maximum	Range	Variance	St. Dev.	St. Error
Data from 2;0 years old boys (N=7)								
i	34.75	32.86	14.00	50.53	36.53	173.94	13.19	4.98
ø	47.42	45.71	33.87	59.99	26.13	80.74	8.99	3.40
u:	39.71	35.71	23.59	59.28	35.69	178.71	13.37	5.05
Data from 3;0 years old boys (N=7)								
i	**29.57**	**28.00**	**20.29**	**44.25**	**23.96**	**59.73**	**7.73**	**2.92**
ø	**31.85**	**33.48**	**23.71**	**37.67**	**13.95**	**32.59**	**5.71**	**2.16**
u:	**44.08**	**45.85**	**30.50**	**53.33**	**22.83**	**66.27**	**8.14**	**3.08**
Data from 4;0 years old boys (N=7)								
i	33.90	35.57	23.14	42.43	19.29	50.84	7.13	2.70
ø	34.14	32.40	26.00	44.67	18.67	46.09	6.79	2.57
u:	43.55	44.00	30.50	60.00	29.50	78.19	8.84	3.34

Results of a two-factor ANOVA test revealed a significant main effect of the preceding and following vowel quality (F(2,54)=5.8038, p< .01). For this position, no effect of age (F 2,54)= 1.8802, p= .16241) was found: 2;0 year olds (mean: 40.63, SD: 12.59) did not differ significantly from either the 3;0 (mean: 35.16, SD: 9.50) or the 4;0 (mean: 37.19, SD: 8.59) year olds. Thus, VOT values of older children were not shorter in word-medial stops. Similarly, no interaction effect was found (F(4,54)=2.515, p= .05202).

Post-hoc Tukey HSD tests indicated that VOT values of word-medial voiceless bilabial plosives preceded and followed by the vowels /i/ vs. /ø/ did not differ from each other at p< .05. However, the VOTs of voiceless bilabial plosives preceded and followed by the phonologically long back rounded vowel /u:/ (mean: 42.44, SD: 10.05) differed from those preceded and followed by the phonologically short unrounded /i/ (mean: 32.74, SD: 9.53) but not the phonologically short rounded /ø/ vowel (mean: 37.80, SD: 9.87). These results are shown in Figure 2.

Figure 2: Mean plot of VOT values for word-medial /p/ grouped by preceding and following vowel quality categorized by age groups (in ms)



Overall, the VOT values of the word-medial /p/ indicate that children in the two older age groups do not produce stops with shorter VOT durations than 2;0 year olds.

Additionally, it is noteworthy that word-medial voiceless bilabial plosives embedded between front unrounded illabial vowels are charaterized by considerably shorter VOT durations than those embedded between phonologically long high back rounded vowels but not front rounded ones. Thus, once again data reflects considerably longer VOT durations for voiceless bilabial plosives embedded between rounded vowels with prominent (vs. much less prominent) lip rounding.

4.3. Comparing the results of word-initial vs. word-medial voiceless bilabial plosives

Results of a t-test for dependent samples analysis revealed that the VOT values (in ms) of word-initial voiceless bilabial plosives are significantly shorter on average (mean: 30.09, SD: 10.23) than word-medial values (mean: 37.66, SD: 10.45) at the $p < .00001$ level. [1]

[1] Previously reported results [27] on young children's avoidance to produce word-initial /p/ sounds while acquiring English were not detected in children here.

5. CONCLUSIONS

Mastery of gestural coordination in speech production takes many years in children to develop [27, 28, 29, 30] before reaching adult levels. This long period of speech production mastery can be explained by considering the important cognitive and physiological loads on the child's developing system represented by speech and language learning [31]. It has been shown that the effect of speaking rate on speech segment durations is talker-specific in adults [32]. However, developmental changes in children's speaking rate need further exploration to map out tendencies in acquisition for voice onset time durations.

The goal of this study was to examine whether children between 2;0 and 4;0 years acquiring Hungarian produce voiceless bilabial plosives with longer VOT durations than adults. Results suggest that children produce longer average voice onset time durations in all positions examined, in comparison with the adult data reported in the literature.

A second question examined the effects of lip rounding on VOT production in children. It was found that voiceless bilabial plosives followed by or embedded between rounded vowels with very prominent lip protrusion (such as the one required during the production of the Hungarian /u:/) are charaterized by longer VOT. In contrast, vowels with less prominent lip rounding (such as the one exhibited during the production of the Hungarian /ø/) do not co-occur with longer VOT. Thus, more prominent lip rounding has an effect on the duration of VOT in Hungarian-speaking children's speech.

A third question examined whether children's production data reflects increasingly adult-like values in speech timing patterns, as shown by VOT. Comparing the production of word-initial and word-medial voiceless bilabial stops, 3;0 and 4;0 years old boys produced stops with considerably shorter (and thus more adult-like) VOT than children at 2;0 years, indicating a more mature speech production system.

Maturational changes in children's speech output signal the emergence of an increasingly adult like speech production mechanism. Mature speech motor control coupled with more advanced cognitive skills pave the way for speech sequencing capacities that are executed at faster rates. Further research is required to continue to explore acquisition patterns for mature voice onset time production in children across languages.

6. REFERENCES

[1] UPSID database. http://web.phonetik.uni-frankfurt.de/upsid.html. Also available via the UCLA Linguistics Department website.

[2] Maddieson, I. 1984. *Patterns of sounds.* (Cambridge Studies in Speech Science and Communication.) Cambridge: Cambridge University Press.

[3] Lisker, L., Abramson, A.S. 1964. A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20, 384-422.

[4] Chao, K-Y., Chen, L-m. 2008. A cross-linguistic study of voice onset time in stop consonant productions. *Comput. Ling. and Chinese Lang. Proc.* 13, 215-232.

[5] Gósy, M. 2001. The VOT of the Hungarian voiceless plosives in words and in spontaneous speech. *Int. J. Speech Technology*, 4, 75-85.

[6] Zajdó, K. 2012. *Mothers` modeling of phonological vowel length and voice onset time in Hungarian: Similar or individual processes?* Poster presented *at the* 14th Meeting of the Int. Clinical Phonetics and Linguistics Association, Cork, University College, Cork, Ireland.

[7] Westbury, J.R., Keating, P.A. 1986. On the naturalness of stop consonant voicing. *J. Linguist.* 22, 145-166.

[8] Macken, M. A., Barton, D. 1980. 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.

[9] Allen, G. D. 1985. How the young French child avoids the pre-voicing problem for word-initial voiced stops. *J. Child Lang.* 12, 37–46.

[10] Clumeck, H., Barton, D., Macken, M. A., Hunting-ton, D. A. 1981. The aspiration contrast in Cantonese word-initial stops — Data from children and adults. *J. Chin. Linguist.* 9, 210–225.

[11] Davis, K. 1995. Phonetic and phonological contrasts in the acquisition of voicing — Voice onset time production in Hindi and English. *J. Child Lang.* 22, 275–305.

[12] Gandour, J., Petty, S. H., Dardarananda, R., Dechongkit, S., Mukngoen, S. 1986. The acquisition of the voicing contrast in Thai — A study of voice onset time in word-initial stop consonants. *J. Child Lang.* 13, 561–572.

[13] Eilers, R. E., Oller, D. K., Benitogarcia, C. R. 1984. The acquisition of voicing contrasts in Spanish and English learning infants and children — A longitudinal study. *J. Child Lang.* 11, 313–336.

[14] Kim, M., Stoel-Gammon, C. 2009. The acquisition of Korean word-initial stops. *J. Acoust. Soc. Am.* 125, 3950-3961.

[15] Kim, M., Stoel-Gammon, C. 2010. Segmental timing of young children and adults. *Int. J. of Speech-Lang. Path.* 12, 221-229.

[16] Kim, M., Stoel-Gammon, C. 2011. Phonological development of word-initial Korean obstruents in young Korean children. *J. Child Lang.* 38, 316-340.

[17] Zajdó, K. 2007. Hungarian speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition*. Clifton Park, NY: Thomson Delmar, 412-436.

[18] Zajdó, K. 2002. *The acquisition of vowels by Hungarian-speaking children aged two to four years: A cross-sectional study.* Unpublished doctoral dissertation, University of Washington (Seattle), US.

[19] Zajdó, K., Stoel-Gammon, C. 2003. The acquisition of vowels in Hungarian: Developmental data. *Proc. 15th ICPhS* Barcelona, 2229-2232.

[20] Zajdó, K. 2013. Cross-linguistic trends in the acquisition of speech sounds. In: B. Peter and A. MacLeod (Eds.), *Comprehensive perspectives on speech sound development and disorders: Pathways from linguistic theory to clinical practice.* New York: Nova Science Publishers, 249-274.

[21] Zajdó, K., van der Stelt, J.M., Wempe, T., Pols, L.C.W. 2005. Cross-linguistic comparison of two-year-old children's acoustic vowel spaces: Contrasting Hungarian with Dutch. *Proc. InterSpeech 2005* Lisbon, 1173-1176.

[22] Zajdó, K., Van der Stelt, J.M., Wempe, T., Pols, L.C.W. 2005. *Proc. of InterSpeech* Lisbon, 1173-1176.

[23] Zajdó, K., Wempe, T., van der Stelt, J., Pols, L.C.W. 2011. The acquisition of Hungarian high front unrounded short. vs. long vowels. *Proc. ICPhS XVII* Hong Kong, 2252-2255.

[24] Smith, A. 2013. Development of neural control of orofacial movements for speech. In Hardcastle, W. J., Laver, J., Gibbon, F. E. (eds), *The Handbook of Phonetic Science* (2nd ed.). Oxford, Blackwell, 251-296.

[25] Riely, R., Smith, A. 2003. Speech movements do not scale by orofacial structure size. *J. of Applied Physiology*, 94, 2119-2126.

[26] Boersma, P. Weenink, D. 2014. Praat: Doing phonetics by computer (Version 5.4.04, 32 bit, for Windows) . http://www.fon.hum.uva.nl/praat/

[27] Lowenstein, J.H., Nittrouer, S. 2008. Patterns of acquisition of native voice onset time in English-learning children. *J. Acoust. Soc. Am.* 124, 1180-1191.

[28] Lee, S., Potamianos, A., Narayanan, S. 1999. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *J. Acoust. Soc. Am.* 105(3), 1455-1468.

[29] Zajdó, K., Powell, S. (2008). The acquisition of phonological vowel length in children acquiring Hungarian. In *Proc. 8th Int. Seminar on Speech Production*, Strasbourg, France: 173-176.

[30] Buder, E.H., Stoel-Gammon, C. 2002. American and Swedish children's acquisition of vowel duration: Effects of vowel identity and final stop voicing. *J. Acoust. Soc. Am.* 111, 1854-1864.

[31] Davis, B.L. 2013. Speech acquisition. In Hardcastle, W. J., Laver, J., Gibbon, F. E. (eds), *The Handbook of Phonetic Science* (2nd ed.). Oxford, Blackwell, 299-315.

[32] Theodore, R.M., Miller, J.L., DeSteno, D. 2007. The effect of speaking rate on voice-onset-time is talker specific. *Proc. of 16th ICPhS* Saarbrücken, 473-476.