/sP/ consonant clusters in Swedish: Acoustic measurements of phonological development *

Fredrik Karlsson & Kirk P. H. Sullivan

Abstract

This paper presents a cross-sectional and longitudinal investigation into the development of a contrast in voicing production in simple and complex consonant clusters for 21 children aged 1;6 to 4;2. Voice onset time and length of aperiodic noise after plosive release were measured and the group trend as well as the developmental paths of the individual children were extracted. The results showed that, although the group results indicated a progression that may have been caused solely by development in articulatory proficiency, individual children develop in a manner that is not consistent with a developmental model postulating only articulatory development.

1 Introduction

The Swedish phonotactic system allows complex syllable onsets and codas. In initial position the syllable onset may consist of a consonant sequence of up to three elements (Sigurd, 1965). The most frequent initial two-element cluster type in spontaneous Swedish, accounting for 53.3% of these consonant clusters is the s+plosive cluster (Bannert and Czigler, 1999). As these clusters are frequent in the ambient language of children learning Swedish, investigations of the child's early attempted productions of word-initial s+plosive clusters may serve as a tool for investigating the transition from one-element syllable onsets to complex syllable onsets.

In Swedish, the distinction between the voiced and voiceless plosive cognates (e.g. /p/ and /b/) has traditionally been described through combinations of the two articulatory features [±voice] and [±asp]. Out of the four possible combinations of these features per place of articulation, three plosives ([b], [p] and [p^h]) are part of the Swedish consonant system; [b] and [p^h] are found in word-initial position and [p] is found in word-medial position.

The child's task of producing an adult word initial consonant cluster involving medial plosives, therefore, includes noticing the allophonic variation in the acoustic properties of plosives in different positions within a word. Furthermore, it involves making appropriate, position-dependent, modifications in the motor commands so that the acoustic output form of the plosive is not perceived, by the adult listener, as significantly different from the intended allophone.

Before accurate, adult-like productions have been achieved by the child, reduced forms of consonant clusters have been reported to be produced (e.g. Catts and Kamhi, 1984). Researchers, primarily through auditory investigations, have identified two types of processes, cluster reduction and cluster simplification. Cluster reduction is when, for example, the cluster /sP/ is reduced to the single plosive or, more rarely, to a single [s]. Cluster reduction has been reported in studies investigating the earlier stages of cluster production. Cluster simplification is when children substitute the voiced cognate of the plosive for the cluster (e.g. [sp] → [[b]][1]). Compared to cluster reduction, cluster simplification has been reported to occur during the later stages of development

*In Marina Tzakosta, Claartje Levelt & Jeroen van de Weijer (eds.), Developmental Paths in Phonological Acquisition. Special issue of Leiden Papers in Linguistics 2.1 (2005), 1–12. `http://www.ulcl.leidenuniv.nl`, ISSN 1574-4728.

[1] In this paper, the [[*utterance*]] notation will be used in descriptions of child output forms, in order to distinguish them from the [*utterance*] adult target form.

(e.g. Bond and Wilson, 1980). In their summary of trends found in previous research, McLeod et al. (2001) noted a tendency towards a trading relation between cluster reduction and cluster simplification throughout speech development.

Both cluster reduction and cluster simplification may cause productions that, to the adult observer, are perceptually inseparable from adult forms of words with initial singleton consonants. For instance, in the case of the target word [spɑːk] (lever) that when reduced to [[sɑːk]], by application of the process of cluster reduction, may be perceptually identical to another potential target word [sɑːk] (thing). Similarly, the same target word [spɑːk] may be produced as [[bɑːk]] due to application of the phonological process of consonant cluster simplification and, thus, possibly be perceived as the potential target word [bɑːk] (bottom). As Bond and Wilson (1980) pointed out, the child's productions will change from the reduced or simplified form towards the adult target production as the child develops.

1.1 Models of phonological acquisition

A number of models of the acquisition of phonological contrasts have been proposed. These models may broadly be divided into two major groups based upon the model's explanation of the relationship between the child's internal representation and the adult's model form (Maxwell, 1984).

The first group of models assumes that the child's internal representation is identical to the adult model. Consequently, any discrepancy between the model and the produced form is caused only by realisation processes (Smith, 1973).

The second group of models allows for the possibility that perceptual processes influence the child's internal representation of a word and permit the representation to be different in form from the output form of the adult model (Macken, 1980) From this perspective, the child's internal representation of the produced speech tokens may change throughout development. Therefore, this group of models implies the existence of two separate rule systems; the first generating the internal representation from the input form of an external model utterance and the second generating the output forms from the internal representation produced by the first system.

These two groups of models generate distinct predictions of the acoustic manifestations of homophones produced by children. In the first group of models, where the child's underlying representation is identical to the adult model, development of two distinct productions is solely influenced by the maturity level of the realisation processes. Once a proficiency to produce a specified feature has been established, the acoustic manifestation of the feature should be present in the output forms of all words where it is specified in the adult representation in a context of a similar articulatory complexity. Furthermore, there should be a divergence in the child's output forms of phonologically different adult target words at the time when the feature is first physically manifested.

In the second group of models, however, the child's internal representation of the two target words may be under-specified with regard to the properties that are necessary in the production for the adult observer to perceive them as different. In this view, it may not be assumed, *a priori*, that the acoustic manifestation of a feature will be present in contexts of similar complexity. In addition, the occurrence of an acoustic manifestation of a phonological feature in the child's output form of one target word does not necessarily imply the occurrence of that feature in all contexts of similar articulatory complexity.

1.2 Teasing the models apart through acoustic measurements

The output form of the phonological processes cluster reduction and cluster simplification, when applied to the target cluster, /sp/, is either the voiced, /b/, or voiceless, /p/, variant of the plosive. In the acoustic domain, measurement of voice onset time (VOT), i.e. the time between plosive release and onset of voicing, has been shown to be a strong acoustic cue to voicing quality in plosives across languages (Lisker and Abramson, 1964).

It has furthermore been argued that the production of voiceless plosives with a long lagged VOT involves a more complex articulatory movement than voiced plosives with a short lagged VOT (Kewley-Port and Preston, 1974). In the domain of contrast acquisition, this would suggest that voiced plosives should be predominant in the productions of a child until sufficient articulatory control has developed. Additionally, since there is a phonetic difference in aspiration in Swedish between the voiceless plosive in cluster context compared to initial position, the distinction between reduced and unreduced productions, such as [[pɑːk]] and [pɑːk], can be examined through acoustic measurement of aspiration duration.

According the group of phonological acquisition models that include only output realisation processes (H1), the initial voiced and voiceless plosives should be separable through VOT and aspiration measurements from the onset of the occurrence of long lagged VOTs in the child's productions. Furthermore, there should be a linear progression towards the adult VOT model range. In contrast, for the group of models that allows for input processes to influence the child's internal phonological representation of an utterance (H2), there is no *a priori* reason for the progression from the earliest production towards the adult model to be linear. Indeed, as the child's internal representation of the intended utterance may vary at different stages of development, a substantial degree of experimentation with different lengths of voice onset and aspiration is to be expected.

In order to evaluate the validity of the two hypotheses presented above, VOT and aspiration duration were investigated in a case-study of the development of one male Swedish child using monthly recordings between 18–31 months (Karlsson et al., 2002). The results showed an onset of productions with a longer aspiration and a longer VOT for all the possible Swedish plosive types (voiced, voiceless unaspirated and voiceless aspirated) at 27 months. The results were interpreted as indicating a change in production strategy at 27 months in terms of the voicing distinction. The co-occurrence of longer VOTs for all plosive categories created by the features voicing and aspiration was tentatively interpreted as an indication that the child had noticed the difference timing of voice onset and aspiration between his own output form and that of the ambient language. The application of longer VOT across all voicing categories was interpreted as a sign of uncertainty in the application of the discovered component of speech, giving support to H2.

The plausibility of H2 was further increased by a follow-up study investigating the nature of the frictious portions of the syllable onsets (Karlsson et al., 2003). The results presented in Karlsson et al. (2003) showed that some of the plosives produced by the child, that were marked as unaspirated by an adult listener and acoustic measurement using traditional markup criteria for aspiration, were in fact produced in breathy voice. Measuring the total length of frictious portion after plosive release, the results showed a significantly longer post-release friction length in plosives produced with a long lagged VOT (VOT $>$ 40ms) compared to plosives produced with a short lagged VOT (VOT $<$ 40ms). A more detailed examination of the tokens produced showed that in instances, where a long portion of breathy voice was present, the target word always involved an aspirated plosive in initial position. In this respect, the data presented by Karlsson et al. (2003) can be intepreted as indicating that the child had acquired a general understanding of the distribution of aspiration, but that an understanding of adult-like acoustic manifestation of aspiration had not yet been acquired. Together, the initial tendency towards VOT in the short-lagged range, the increase in VOT across plosive voicing category and the substitution of length for power in friction after plosive release are interpreted as signs of development of overall understanding of voicing and aspiration contrasts through advances in both production ability and awareness of the relevant acoustic cues. Since there is advancement in both perception and production, the hypothesis H1 is not given support by the collected data.

However, one should keep in mind that the studies presented in Karlsson et al. (2002, 2003) concerned speech development of one child. Considering the large degree of variance found in investigations of acoustic measurements of grouped child speech (e.g. Eguchi and Hirsh, 1969) and the tendency towards diverging developmental paths in word learning across subjects, it is likelly that multiple routes in s+plosive acquisition exist. Although informative as to the developmental paths possible for the child acquiring the ability for adult-like production of word-initial consonant clusters, the results from Karlsson et al. (2002, 2003) do not provide any indication as to the gener-

C^{h}V(C)		C̬V(C)		CCV(C)	
kal	[k^{h}ɑːl]	*gal*	[gɑːl]	*skal*	[skɑːl]
pår	[p^{h}oːr]	*bår*	[boːr]	*spår*	[spoːr]
tå	[t^{h}oː]	*då*	[doː]	*stå*	[stoː]
pak	[p^{h}ɑːk]	*bak*	[bɑːk]	*spak*	[spɑːk]

Table 1: The corpus of target word triplets used in the study.

ality of the observed path. The development in perception may be dependent on the general stage of speech development of a specific child. Therefore, this study investigates intra-group variability and variance in VOT and aspiration length, and identifies rivalling developmental paths for these acoustic features. Results showing developmental adjustments in these acoustic parameters that are manifested only in the correct adult voicing category will be interpreted as supporting the H1 hypothesis. Changes occuring in incorrect segments according to adult speech will be interpreted as an indication of a developing underlying phonological representation and as providing support for hypothesis H2.

2 Method

The method selected for the investigation was cross-sectional and longitudinal.

2.1 Subjects

Twenty-one monolingual children, 12 male and 9 female, were recruited for the study. Information regarding the children's hearing status and dialect of Swedish spoken in the home environment was gathered from the parents/care-takers before the child began the recording phase of the investigation. All the children were growing up in a monolingual home environment and there were no known hearing problems.

Each participant was recorded at monthly intervals for a period of twelve months. The recording phase of the investigation resulted in 229 acoustically satisfactory recordings. The children were between 77 and 218 weeks old at the time of the recording sessions.

2.2 Speech material

The target words elicited were four triplets of monosyllabic words. Each triplet contained target words with an identical syllable nucleus and coda. The onsets in each triplet were a voiced plosive, a voiceless aspirated plosive and an s+plosive consonant cluster in which the plosive is voiceless unaspirated in adult speech. The target words are presented in table 1.

2.3 Procedure

The recordings were made in a sound treated room at the Department of Philosophy and Linguistics, Umeå University. The target words were elicited by the parent or care-taker accompanying the child to the recording session. The target words were elicited using black-and-white picture prompts. Each parent or caretaker was informed that they should aim for spontaneous productions. When it proved impossible to elicit a spontaneous production of a target word, they were instructed to elicit the target word through a delayed repetition task by providing the target word in non-final position in a carrier sentence (1). In the cases where a spontaneous production was not obtained, the parents and caretakers were asked to elicit the target word through direct repetition of an adult model production (2).

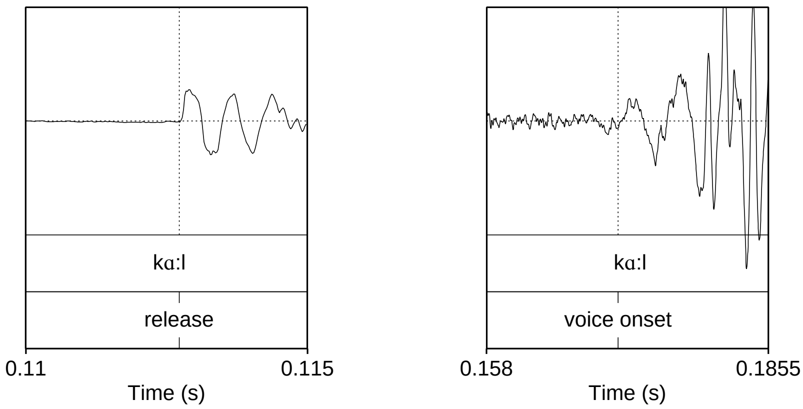

Figure 1: Example markup of time of plosive release (left) and time of voicing onset (right) for the target word [kɑːl]. Source: "Development of a Gender Difference in Voice Onset Time" by Karlsson, F., Zetterholm, E., & Sullivan, K. P. H. 2004. Proceedings of the Tenth Australian International Conference on Speech Science & Technology. pp 316–321. Reprinted with permission.

(1) *Är det en tå eller en fot?*
Is it a *toe* or a *foot*?
'Is it a toe or a foot?'

(2) *Titta, en tå! Kan du säga tå?*
Look, a *toe*! Can you say *toe*?
'Look, a toe! Can you say toe?'

2.4 Phonetic markup and analysis

Initially, the children's productions of the target words were identified and extracted from the recorded sessions. A child's production was extracted from the session recording and included into the set of analysed productions iff the context in which the production was made and the parent's reaction towards the production indicated that the child had attempted to produce the target word in question. Productions that co-occurred with an external sound, such as banging on the table or an utterance made by someone other than the child, were not included in the set of analysed productions.

The extracted productions were marked for presence of frictious noise after plosive release and voicing:

- The onset and offset of friction after plosive release was marked at the boundaries of the portion of the speech signal with an audible frication noise after the release of the plosive.
- The release of the plosive was marked at the last occurrence of a transient in the waveform that corresponded in time with an audible release of airflow in the plosive.
- The onset of voicing was marked at the first positive zero crossing before the onset of periodicity in the waveform.

An example of markup is presented in figure 1.

From these acoustic marks, the phonetic quantities of VOT and post release friction length were calculated. Voice onset time was calculated as the difference in time between the time of voicing onset and time of plosive release. Voice onset time was not calculated for productions with no release or no voicing onset. The analysed corpus included 3535 productions with measurable VOT.

These quantifications of the produced plosives' voicing type were then grouped according subject's age, the voicing type of the target word's plosive and place of articulation of the target word's plosive. The collected data were analysed statistically using the `nlme` (Pinheiro et al., 2004) package for mixed-effects models in the R statistical package (R Development Core Team, 2004).

3 Results

A LOESS line (Cleveland, 1979) of the group results obtained from the voice onset time measurements is presented in figure 2. The panels in figure 2 present the VOT measurements obtained divided into categories based on target plosive voicing type and place of articulation. Figure 2 shows that all productions made by the investigated group, towards a plosive in word initial position, were produced with a short lagged VOT in the early stages of development, regardless of target voicing and aspiration quality. Unaspirated targets are produced with a short lagged VOT throughout the investigated time period. For aspirated targets, however, a steady increase in VOT is observed, reaching a local maximum around 115 weeks for velar targets, 127 weeks for labial targets and 136 weeks for dental targets. At this point in development, the trend line keeps this steady state throughout the investigated period labial and velar target plosives. For dental targets, however, a decrease in predicted VOT is observed towards the investigated window.

A linear mixed-effects model, with Voicing category, Place of articulation and Subject's age as fixed effects and Subject as random effect, was fitted to the VOT measurements. The resulting model was tested against the residuals using an ANOVA and the results showed a significant effect of Voicing category ($F_{(2,2903)}$=146, p<0.01), Place of articulation ($F_{(2,2903)}$=15, p<0.01) and the interaction of Voicing category and Subject's age ($F_{(2,2903)}$=6.5, p<0.01).

Figure 3 presents a LOESS trend line for the measurements of the frictious portion after plosive release. A linear mixed-effects model, using the same independent factors as the VOT model, was fitted to the frication length measurements. As in the case of the linear model of VOT, an ANOVA investigating effects on frication length showed a significant effect of Voicing category ($F_{(2,1632)}$=13, p<0.01) and Place of articulation ($F_{(2,1632)}$=22, p<0.01). Regarding interaction effects, however, only the interaction between Voicing category and Place or articulation reached significance (F_2=3, p<0.05).

The similarities in the results from the statistical testing of the voice onset time and frication length models (figures 2 and 3) indicate a strong correlation between the two measurements. A Spearman rank correlation ρ test confirmed that that the acoustic measurements VOT and frication length were significantly correlated (ρ=0.297,p<0.01).

The results obtained from the grouped measurements of VOT and frication length deviate from the results obtained in the case study presented in Karlsson et al. (2002). In the case study, VOT increased for velar plosives across voicing categories. In contrast, the statistical testing of the group results in figures 2 and 3 showed that the increase in VOT to a region corresponding to an adult voiceless aspirated plosive only occured when the target word was, in fact, an aspirated plosive. Furthermore, inspection of the residuals of the two models shows that the degree of unexplained variance is substantial (figure 4). Together, these two observations suggests heterogeneous developmental paths for the investigated children. Therefore, more detailed investigations of individual developmental patterns were conducted.

3.1 Individual developmental paths

An examination of each child's data separately revealed a number of developmental types for voice onset time production. In figure 5, VOT data for labial target plosives produced by the male subject M2 is presented. This figure shows how VOT begins to increase at 123 weeks for

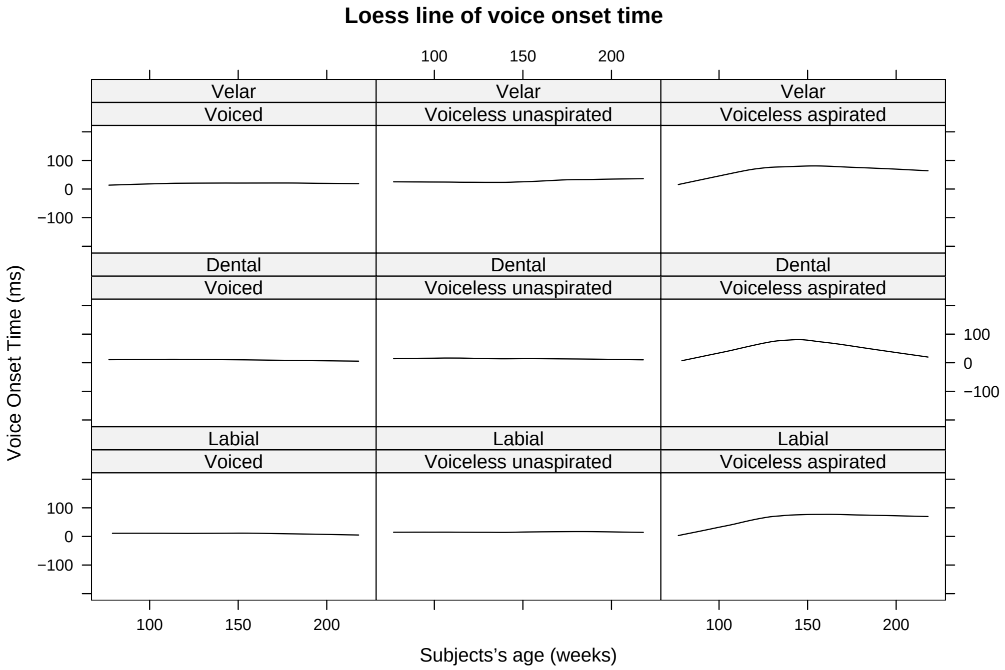

Figure 2: A LOESS line (Cleveland, 1979) indicating the group developmental trend in VOT at the investigated ages (in number of weeks). Panels present the VOT measurements obtained divided into categories based on target plosive voicing type (columns) and place of articulation (rows). The LOESS line was calculated with a span of 0.75.

for aspirated target plosives (figure 5: right panel) and that neither the same magnitude nor continuity of increase is found for the medial unaspirated nor the voice plosive targets. M2's pattern of development for labial plosives follows the group pattern presented in figure 2.

In contrast to M2, F10 follows a different developmental path. This path is illustrated in figure 6; this speaker produced unspirated target plosives at the onset of production with long portions of frictious noise after plosive release (figure 6: middle panel). At onset of production this speaker has similar degrees of frictious noise for productions towards both aspirated and unaspirated voiceless plosives. As this child developed, post-release friction length decreased for unaspirated plosives and increased for aspirated plosives. VOT for the voice plosives (figure 6: left panel) steadily decreased over the investigated period; at the end of the period, when the unaspirated and aspirated cognates had reached a more adultlike level, voiced labial plosives were produced with substantial prevoicing. The strategy of prevoicing initial voiced plosives as a way of contrasting voiced productions from unspirated plosives in consonant cluster environments is also used, although to a lesser extend by the female child F3 (figure 7). F3 does not, however, consistently use pre-voicing; the majority of her productions towards a voiced plosive are in the short lagged range. This speaker further illustrates the range of individuality in the developmental paths by producing no clear observable differences in their productions towards voiceless unaspirated and voiceless aspirated labial plosives; a developmental path feature that is not the groups path nor the same as M2 or F10's developmental paths.

The left panel of figure 6 shows that the VOT of voiced plosives is steadily decreased as development continues. At the end of the investigated period, when VOT for the unaspirated and aspirated cognates have reached a more adultlike level, voiced labial plosives targets are produced with substantial prevoicing.

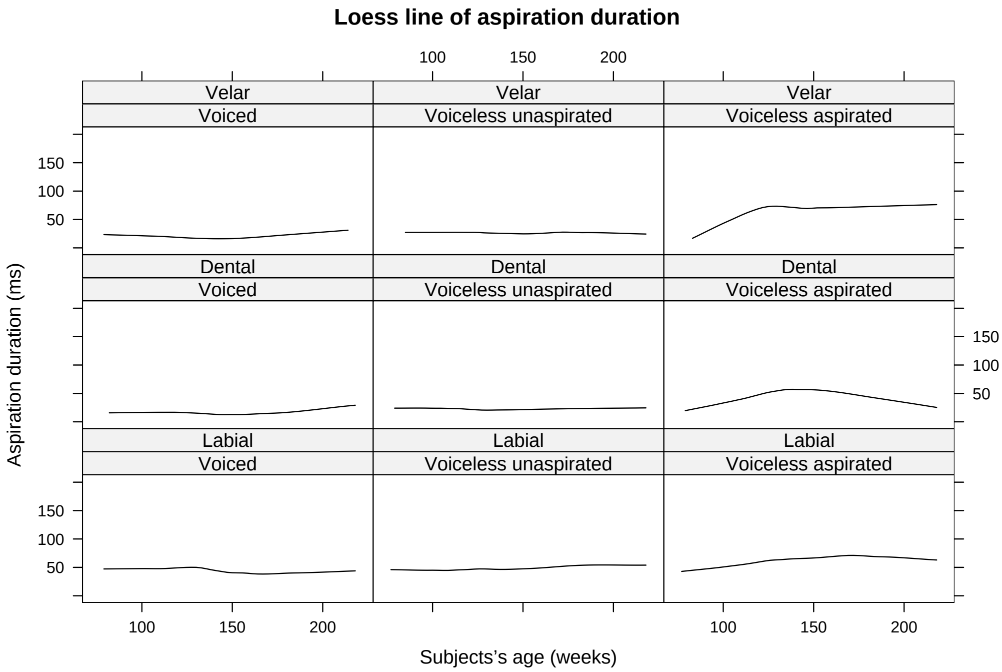

Figure 3: A LOESS line (Cleveland, 1979) indicating the group developmental trend in post-release frication length at the investigated ages (in number of weeks). Panels present the frication length measurements obtained divided into categories based on target plosive voicing type (columns) and place of articulation (rows). The LOESS line was calculated with a span of 0.75.

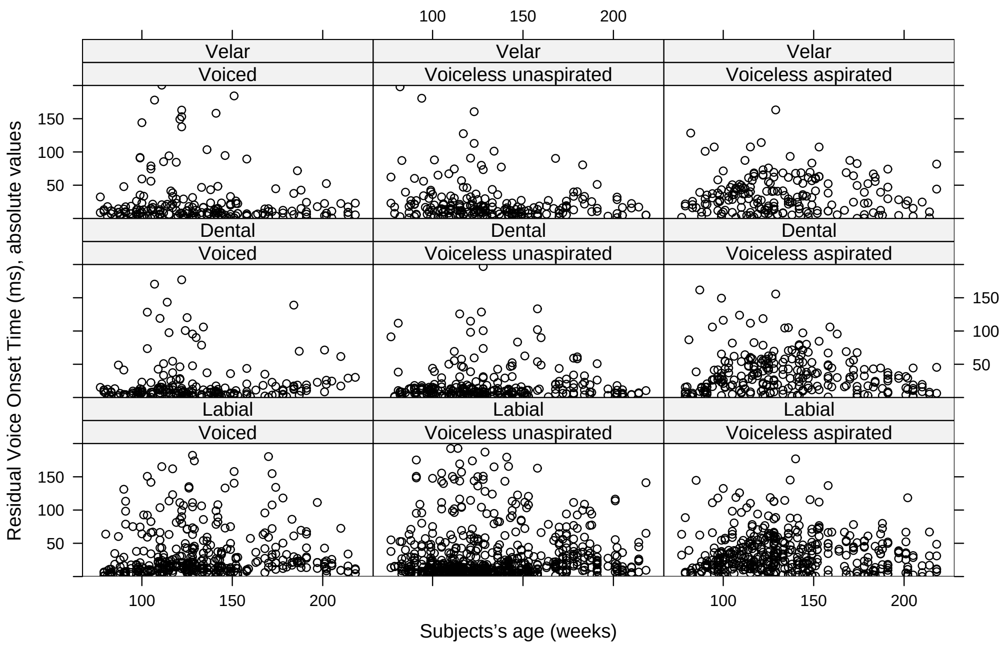

Figure 4: Absolute residuals for the loess model of VOT by subject's age in weeks. Panels present the VOT measurements obtained divided into categories based on target plosive voicing type (columns) and place of articulation (rows).

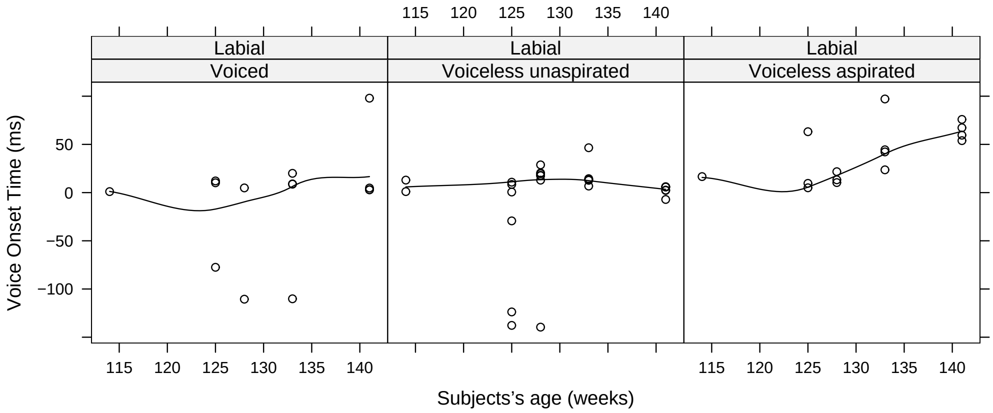

Figure 5: Voice onset time of attempted productions of labial plosives produced by the male child M2. Productions towards voiced targets are presented in the left panel, voiceless unaspirated plosives (that are constituents of a consonant cluster) in the middle panel and aspirated plosives in the right panel.

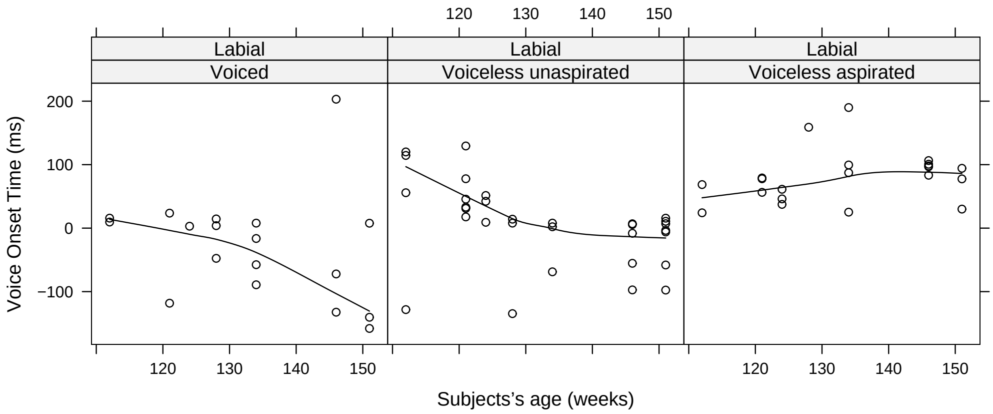

Figure 6: Voice onset time of attempted productions of labial plosives produced by the female child F10. Productions towards voiced targets are presented in the left panel, voiceless unaspirated plosives (that are constituents of a consonant cluster) in the middle panel and aspirated plosives in the right panel.

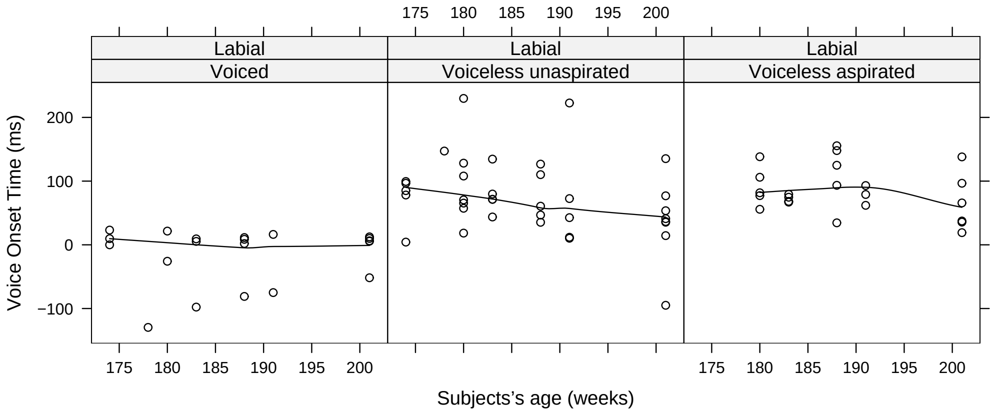

Figure 7: Voice onset time of attempted productions of labial plosives produced by the female child F3. Productions towards voiced targets are presented in the left panel, voiceless unaspirated plosives (that are constituents of a consonant cluster) in the middle panel and aspirated plosives in the right panel.

The strategy of prevoicing initial voiced plosives as a way of contrasting those productions from that of unaspirated plosives in a consonant cluster environment is also used, although to a lesser extent, by the female child F3. Although the bulk of her productions towards a voiced plosive are in the short lagged range, prevoicing is sometimes produced. Clustered and aspirated targets produced by speaker F3 are, however, never prevoiced.

4 Discussion

The case study presented in Karlsson et al. (2002) showed an increase in VOT at around 28 months across plosive voicing type. The increase in VOT was not correlated with adult use of aspiration and the data was interpreted as an indication of uncertainty in the acoustic manifestation and distribution of aspiration and voicing.

In this paper, the results from a larger sample of children were investigated using the same acoustic dimensions, VOT and aspiration. Theses results show a group trend for VOT for aspirated target words; depending on the place of articulation the length of VOT reached a maximum value between 115 and 136 weeks before, for velar and labial plosives, stabilizing around this value and, for dental plosives, reducing in length.

This group trend does not concur with the data presented in the Karlsson et al. (2002) case study and suggests the need for an in depth analysis of the developmental paths taken by individual children. An inspection of the progression of the individual child towards the adult target revealed individual developmental paths resembling the pattern observed for the group. For instance, the VOT measurements of the labial productions produced by the male speaker M2 between the ages 113–140 weeks (figure 5) show a strong increase for aspirated targets at 128 weeks. However, as shown in figures 6 and 7, the pattern displayed by the group, and some of the individuals (e.g. M2) does not extend to all the investigated children.

The development in VOT in labial plosives by speaker F3 (figure 7) is one example of the diverse developmental trend in the corpus. For voiced target plosives, values are kept at an adult-like level, with prevoicing and VOT in the short lagged range. However, a tendency towards a decrease in the level of prevoicing may be observed. This pattern is also present in the voiced labial and dental targets produced by other subjects. The voiceless plosives produced by F3 that are initially produced with a long VOT lag show a tendency towards a shortening of the onset throughout thought the investigated time window. Looking solely at the trend line for unaspirated and aspirated voiceless plosives, one could hypothesize that subject F3 is making micro adjustments in voicing delay towards a more adult like range. However, investigation of the individual data points reveals that there is a large degree of variation in this subject's productions; they range from prevoiced tokens to tokens with a very long lagged interval. The subtlety of this gradual change coupled with the large variation across productions makes it, therefore, difficult to determine whether the internal representation of the target word is being refined, or whether the change in the data is caused by a refinement in the execution of articulatory motor commands.

Another example in the analyzed data is the individual data collected from speaker F10 (figure 6. Here, however, the developmental trends are not as subtle as the those of F3. For this child, an interplay between productions of the contrasting target words may be observed. In the initial stages, voiceless unaspirated and aspirated targets are both produced with a substanital VOT lag, and voiced plosives are produced with lag within the unmarked range hypothesized by Kewley-Port and Preston (1974). As the child develops, the VOT of aspirated plosives are steadily increased and the corresponding measure for voiceless unaspirated plosives are decreased. This decrease is parallelled by a change in voiced targets towards increasingly early onsets of voicing. At the end of the investigated time window voiced productions are mainly (with two substantial outliers) produced with heavy prevoicing and voiceless unaspirated plosives are produced with a VOT in the short lagged range. At this point in her development, F10 produces all aspirated targets with a long lagged voice onset time.

The results from the labial productions, as summarized in the group data (figure 2) and from individual children such as M2, reveal no sign of a contrast between the candidate target plosives in the initial stages of development. At the age of 127 weeks, however, aspirated target plosives was started to be produced with a long VOT. Since the increase in VOT is not extended to the other plosive categories, the cause of the change is unclear. This could be caused both by a sudden increase in the ability to produce aspiration or a noticing of the linguistic importance of that acoustic property of plosives. The group summary and M2's productions, unfortunately, do not assist in teasing apart of the two groups of theories, H1 and H2, on the driving forms of phonological development.

The productions made by F10, and to some extent also F3 (figures 6 and 7), are more revealing. At the onset of the investigated window of development, voiced plosives are produced with short VOT lag, voiceless unaspirated plosives are produced with a substantial lag and aspirated plosives are produced with a lag slightly smaller than the unaspirated plosives. Consequently, an argument based on articulatory development is less plausible since the child is able to physically produce both long lagged, intermediate lagged and short lagged VOT at the earlier stages of development. Although towards the end of the period of study, the balance in the application of these acoustic correlates is adjusted towards adult target, the correlates cannot be described as being applied in an adult-like manner. It is, therefore, concluded that the data from F10 and F3 lend support to a model of phonological acquisition that presupposes an increase in articulatory proficiency together with a gradually emerging structure. The development of only one of these components is not sufficient to explain the collected speech data.

References

Bannert, R. and Czigler, P. E. (1999). *Variations in consonant clusters in Standard Swedish.* Number 7 in PHONUM. Umeå University.

Bond, Z. and Wilson, H. (1980). /s/ plus stop clusters in children's speech. *Phonetica*, 37:149–158.

Catts, H. W. and Kamhi, A. G. (1984). Simplification of /s/ + stop consonant clusters: A developmental perspective. *Journal of Speech and Hearing Research*, (27):556–561.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74:829–836.

Eguchi, S. and Hirsh, I. J. (1969). Development of speech sounds in children. *Acta Orolaryngologica*, (Supplement No. 257).

Karlsson, F., Sullivan, K. P., Czigler, P. E., and van Doorn, J. (2002). Acoustic correlates of voicing in a child's production of plosives. In *Proceedings of the 9th Australian International Conference on Speech Science & Technology*, pages 385–390, Melbourne, Australia.

Karlsson, F., Sullivan, K. P. H., van Doorn, J., and Czigler, P. E. (2003). Då or tå, pår or bår — seeing is believing! In *Proceeding to the 15th International Congress of Phonetic Sciences*, pages 1967–1970, Barcelona, Spain.

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

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

Macken, M. A. (1980). A child's lexical representation: The '*puzzle-puddle-pickle*' evidence. *Journal of Linguistics*, (16):1–17.

Maxwell, E. M. (1984). On determining underlying phonological representations of children: A critique of the current theories. *Phonological theory and the misarticulating child*, ASHA Monograph(22):18–29.

McLeod, S., van Doorn, J., and Reed, V. A. (2001). Consonant cluster development in two-yeas-olds: General trends and individual difference. *Journal of Speech, Language, and Hearing Research*, 44:1144–1171.

Pinheiro, J., Bates, D., DebRoy, S., and Sarkar, D. (2004). *nlme: Linear and nonlinear mixed effects models.*

R Development Core Team (2004). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Sigurd, B. (1965). *Phonotactic structures in Swedish.* Scandinavian University Books, Lund.

Smith, N. V. (1973). *The Acquisition of Phonoloy; A Case Study.* Cambridge University Press.