

Development of Initial Clusters in American English by Fraternal Twins: An Acoustic Study

Abigail C. Cohn[†] and Emily K. Kishel[‡]

[†] Cornell University, Ithaca, NY, USA

and Laboratoire de Phonétique et Phonologie (UMR 7018), CNRS/Sorbonne Nouvelle, Paris, France

[‡] Franziska Racker Centers, Ithaca, NY, USA

E-mail: acc4@cornell.edu, ekk5@cornell.edu

ABSTRACT

We investigate the phonological development of initial consonants and consonant clusters in a pair of fraternal female twins acquiring American English. At age 4 years, 1 month, twin A had achieved a nearly adult phonology; while twin B evidenced a reduced inventory of surface contrasts, due to the multiple effects of substitution, deletion, and coalescence. The relative intelligibility of twin B's speech leads one to wonder whether covert contrasts exist, that is, low-level phonetic differences which provide cues to the intended target forms. This question was investigated through an acoustic study with multiple repetitions of target forms, where measurements were made for duration, intensity, and spectral balance. Differences were found for both duration and intensity, providing further evidence of the role of covert contrast in phonological acquisition.

1. INTRODUCTION

As reported by Cohn and Kishel [1, 2], the phonological development of initial consonants and consonant clusters in a pair of fraternal female twins acquiring American English showed marked differences. Studies of fraternal twins are particularly useful, as they allow for systematic comparisons that are otherwise difficult to achieve in language acquisition studies.

In the present study, at age 4 years, 1 month, twin A had achieved a nearly adult phonology; she was not yet able to pronounce /θ/ and /ð/, and produced a slightly odd pronunciation of /r/, although clearly distinct from /w/. Twin B's system differed much more markedly from an adult phonology. Substitutions occurred for /r/, /θ/, and /ð/. There was devoicing of some tokens of target voiced fricatives. Most tokens of both /s/ and /ʃ/ were produced as a somewhat palatal [s]. Affricates were reduced to the corresponding stop. Twin B was just starting to produce clusters. Consistent with observations in the literature ([13] and references cited therein), the first clusters were those exhibiting the greatest sonority distance: obstruent+glide. At this stage, twin B also produced voiceless stop+/l/, /s+/l/, and voiced stop+/l/ some of the time. The substitution of [w] for /r/ persisted throughout CC and CCC clusters and there was some substitution of [w] for /l/. The

/s/+obstruent and /s/+nasal sequences were reduced to a single consonant. Deletion occurred with coronals and velars and in all cases, it was C1 that was retained. For the labials (/sp/ and /sm/), coalescence was observed, with the manner of the fricative and the place of the labial maintained. All CCC clusters were realized as the combined patterns predicted from reduction of /s/+C2 and substitutions for the liquid as C3. Both twins perceived a full range of contrasts, even those that they could not yet produce. (See [1, 2] for a fuller report of the observed phonological patterns and discussion of recent phonological approaches to describing such patterns of acquisition.)

The multiple effects of substitution, deletion, and coalescence resulted in massive neutralization in twin B's speech. Most striking are the range of targets intended by surface [f], [s], [fw], and [sw], as summarized in Table 1:

surface	Intended target			
	direct mapping	substitu-tion	deletion	coales-cence
[f]	f	θ		sp, sm
[s]	s	ʃ	st, sn, sk	
[fw]		fl, fr, θr		spl, spr
[sw]	sw	sl, ʃr	skw, str, skr	

Table 1: Surface realization of intended targets, twin B.

The relative intelligibility of twin B's speech, despite these massive neutralizations, leads one to wonder whether complete neutralization is observed or whether a covert contrast exists. "Covert contrast" refers to cases where phonetic cues produced by young children differentiate target forms, but are not sufficient for the intended contrast to be identified by adult listeners based on impressionistic listening. The presence of such acoustic cues in the case of VOT is well documented [10, 14]. The evidence for covert contrast in reduced clusters is less clear. Early acoustic studies [6, 7, 11, 12] provide examples, but do not quantify the results in such a way that it can be determined how systematic these effects are. Two more comprehensive studies show little systematic difference. In the acoustic analysis of a longitudinal study of one child, Ringo (1985) [13] made a wide range of measurements for the target forms /s, st, sk, sn/; no significant spectral or durational differences were found, except that the reduced form of /sn/

was longer than /s/ at certain stages. In a study of eight children with delayed speech and reduced clusters, Louko (1998) [9] measured duration and found no significant differences.

In order to investigate whether low-level phonetic details provide cues to the intended target forms, an acoustic study of a controlled set of data was carried out. This study was designed to address the following: 1) Are these cases of true neutralization or are there systematic low-level phonetic differences? 2) If differences occur, in what phonetic dimensions are they found? And 3) Are there phonetic differences between the realization of substitution, coalescence, and deletion?

2. METHODOLOGY

To address the need for systematic data across a full range of consonants and consonant cluster, in the present study two forms of each possible initial consonant or consonant cluster were included. Most of the forms (of the shape (C)(C)(C) aj (C) and (C)(C)(C) I (C)) were real words of English that were part of the active vocabulary of both twins. (The few less common items and nonsense forms that were included to complete the full set of forms were produced with no apparent difficulty.) The subset of the list that formed the basis of this acoustic study is presented in Table 2.

[f]	[aj]	[ɪ]	[fw]	[aj]	[ɪ]
f	fight	fit	fl	fly	flip
θ	thigh	thick	fr	fry	frizz
sp	spy	spit	θr	thrice	thrift
sm	smile	Smith	spr	spry	spring
			spl	splice	split
[s]	[aj]	[ɪ]	[sw]	[aj]	[ɪ]
s	sigh	sit	sw	swipe	swim
ʃ	shy	ship	sl	slide	slip
st	sty	stick	ʃr	shrine	shrimp
sn	snipe	snip	str	stripe	strip
sk	sky	skip	skr	scribe	script
			skw	squite	squish
[l]	[aj]	[ɪ]	[sl]	[aj]	[ɪ]
l	light	lip	sl	slide	slip
[w]	[aj]	[ɪ]			
w	white	which			
r	ride	rip			

Table 2: Word list used for acoustic study.

The data, which were recorded on analog tape in a soundproof booth, were elicited through a language game where one of the experimenters said "I can say X, can you say X?" and both twins repeated each form at least three times. This allowed collection of multiple repetitions in a systematic fashion. The production was quite consistent throughout the task and was also consistent with spontaneous production at the time.

The data were digitized using Creative Recorder and the digital files were extracted using CoolEditPro. All usable tokens were included and counting both /aj/ and /ɪ/ forms, this resulted in 3-8 tokens per target consonant type. Acoustic analysis, including measurements for duration, intensity, and spectral balance, was done using Praat. The specific measurements made are reported below.

3. RESULTS

Careful listening by both authors (of whom one is the mother of the twins) is the basis of the impressionistic transcription presented here and coincided with the phonological patterns described above and summarized in Table 1. Individual inspection of all the tokens was undertaken, and some individual tokens were very suggestive of the expected differences. Results are pooled across the two vowel contexts. Due to space limitations, we focus our discussion on surface [f] and [s], as the results for surface [fw] and [sw] show similar patterns. Due to the relatively small number of tokens and the variability in the results, statistical analysis was not done. We report here on results for duration, intensity, and spectral balance in turn.

3.1. DURATION

Durations were measured for the initial consonant(s), following vowel, and postvocalic consonant(s): C(C)V(C). The prediction is that surface [f] or [s] from target clusters should be longer than from target single consonants, e.g. [s] from target /st/ or /sk/ should be longer than [s] from target /s/ or /ʃ/. Durations for surface [f] and [s] are shown in Figure 1.

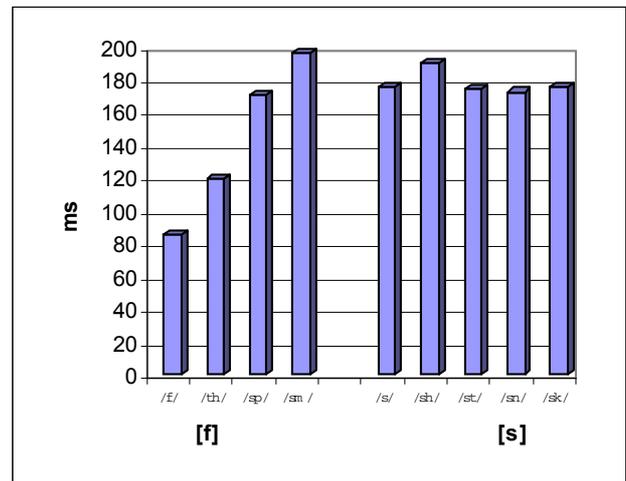


Figure 1: Average durations (in ms.) of surface [f] and surface [s] according to target sound.

On the average, surface [s] is longer than surface [f] (mean duration for surface [s]: 177ms; mean duration for surface [f]: 140ms). This is consistent with results reported in the literature for adult fricative production [5], although the differences found here are greater than those usually observed in adults. For surface [f], both target /sm/ and /sp/ are noticeably longer than for target /f/ or /θ/. They are in the range of surface [s] durations, but not as long as surface

clusters (mean duration for surface [sw]: 279ms; mean duration for surface [fw]: 205ms). This suggests that the realization of coalesced target forms is distinct from target single segments. On the other hand, we see no such difference between target clusters where one member is deleted and target single segments. Target /st, sn, sk/ have durations comparable to target /s/ and /ʃ/. This suggests a difference in the realization of forms involving coalescence (/sp, sm/) and deletion (/st, sn, sk/). Interestingly, realization of /sp/ was quite variable, with some targets realized as [f] and others as [fw]. Twin B may have been on the verge of realizing this target sequence as a cluster.

3.2. INTENSITY

Root mean square (RMS) calculations were made at approximately one fourth and three fourths of the way into the consonant and a ratio was calculated with the RMS of the midpoint of the vowel: RMS-C.25/RMS-V.5 (RMS A) and RMS-C.75/RMS-V.5 (RMS B). This gave a relative measure of intensity. The prediction was that greater intensity would be found for sibilant targets over non-sibilant targets, e.g. /sp/ vs. /f/ or /θ/. Also for target clusters, the relative intensity at point A would be great than at point B, e.g. in surface [f] from target /sp/, there would be greater intensity at point A, corresponding to the target /s/ compared to point B corresponding to the target /p/. Results for surface [f] and [s] are shown in Figure 2.

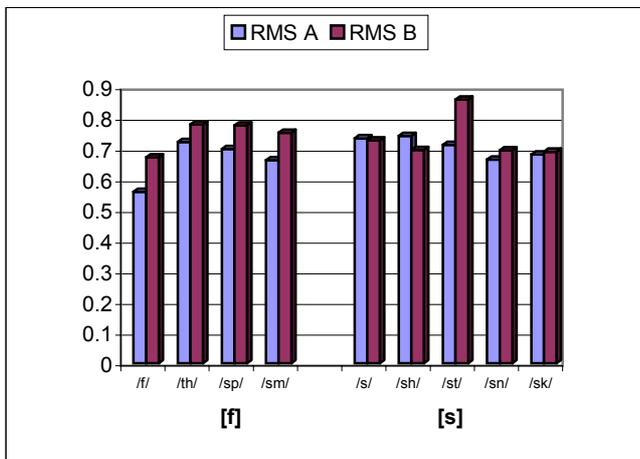


Figure 2: Average RMS Values (as a ratio to the following vowel) at Points A and B for surface [f] and surface [s].

Consider first the surface [f] cases: Noteworthy is the fact that there is a difference in relative intensity between the coalesced forms of the targets /sp, sm/ and the target /f/. However target /θ/ also shows greater intensity than target /f/. This might be interpreted as a negative result, but in fact some of the /θ/ tokens have a somewhat sibilant-like quality to them which might account for their higher relative intensity. For the second prediction, the opposite of what is predicted is observed, that is, relative intensity at point B is greater than at point A in all the cases of target CC. This is part of a broader pattern for surface [f] and a tendency for surface [s] and may have to do with the aerodynamics of fricative production.

3.3. SPECTRAL BALANCE

To investigate spectral balance, for each token, spectra were produced at approximately one fourth and three fourths of the way into the first consonant and were visually examined. Then centroids (first spectral moment [4]) were calculated using a 20ms Hanning window, again at approximately one fourth and three fourths of the way into the first consonant, in order to quantify the average energy concentration across the spectrum. The predictions are that target /s/ should show a higher centroid than the target non-sibilants /f/ and /θ/ and also than /ʃ/ which has a lower locus of energy. Thus a difference would be predicted between [f] from target /sp/ and /sm/ vs. [f] from target /f/ or /θ/ and between [s] from target /s/ vs. /ʃ/. Careful inspection of the spectra led to no clearly identifiable differences in spectral peaks. This is consistent with the centroid results for surface [f] and [s] which also did not show evidence of the predicted differences as shown in Figure 3.

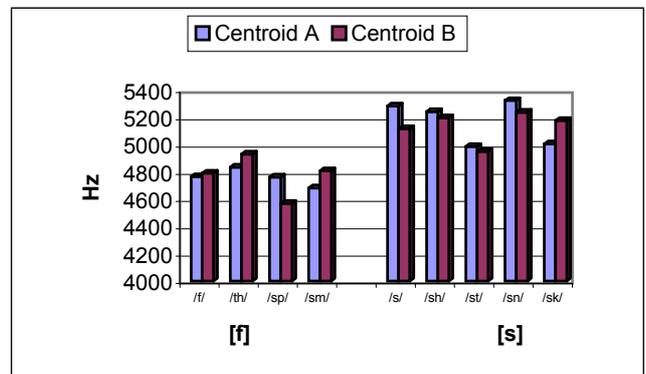


Figure 3: Average Centroid Values (in Hz.) at Points A and B for surface [f] and surface [s].

We do find a robust difference across the board in centroid values for surface [s] (mean values Point A: 5174 Hz.; Point B: 5149 Hz.) and surface [f] (mean values Point A: 4787 Hz.; Point B: 4818 Hz.), as predicted from adult speech [5], but there do not appear to be differences due to the target source of the form.

4. DISCUSSION AND CONCLUSIONS

In conclusion, we found that for some of the apparently neutralized cases in twin B's speech, low-level phonetic differences are observable. Differences were found in both duration and relative intensity. The lack of observable differences in spectral balance could be due to the fact that measures were only made at two points in the consonant. It is possible that additional measurements would reveal differences here too. Overall differences in the realization of surface [f] and surface [s] as a class were robust and showed patterns similar to those of adult speech.

For surface [f], the differences in duration and relative amplitude between target /f/ and target /sm, sp/ suggest that these cases of coalescence are not complete and that a covert contrast exists. This was not found for target cases involving deletion /st, sn, sk/, while in the case of

substitution of [s] for target /ʃ/ no differences were found, but for substitution of [f] for target /θ/ some differences were observed. The fact that duration differences were found is interesting as it is consistent with earlier less systematic studies that noted such differences. The fact that few systematic duration differences were found by [9, 13] may have to do with quite narrow stages of development. It appears that twin B was just in the process of breaking the coalesced forms into clusters. It is possible that just shortly before this recording, durational differences might not have been found. The pattern of coalescence observed for twin B may lend support to the view that in early stages of acquisition, all of the acoustic cues are piled up, in effect as a single unit and as acquisition progresses, these are pulled apart into clusters [6, 7, 11].

It is important to note that young children's speech shows greater variability than adult speech. If low-level phonetic differences exist, they may be difficult to quantify. Patterns of covert contrast may not be as systematic as overt contrasts and may not involve the most obvious phonetic cues; thus while playing a real role in differentiating target forms, they may be hard to document. Nevertheless, the existence of documented covert contrasts in at least some cases raises serious questions about analysis of child language production based only on impressionistic listening. Only through systematic phonetic study, with a full range of possible cues investigated, can it be determined whether perceived patterns are attributable to the phonology or the phonetics.

A fuller understanding of covert contrast could come from perceptual studies of child production. These data should be subjected to perceptual tests to ascertain whether acoustic differences are identifiable. Further it would be interesting to conduct perceptual studies involving siblings and caretakers of children. In [8] it is reported that children reliably differentiate their own [w]'s from target /r/ vs. /w/. Research on twins shows that their increased ability to understand their siblings may account for many of the anecdotal descriptions in the literature of twin language [3]. In the present study, twin A quite consistently understood twin B. We might predict varying degrees of sensitivity to covert contrast based on degree of familiarity with the child producing the forms.

ACKNOWLEDGMENTS

We would like to thank Cécile Fougeron and Jacqueline Vaissière for their helpful comments on this paper and also the members of the Cornell Phonetics Lab and participants at LabPhon 8 who provided helpful comments on an earlier version. Special thanks to Amanda Miller-Ockhuizen for developing the centroid script in Praat. This research was conducted while the second author was a student at Cornell.

REFERENCES

[1] A. Cohn and E. Kishel, "Development of initial consonant clusters in English: A case study of fraternal

- twins," poster at *LabPhon 8*, New Haven, CT, 2002.
- [2] A. Cohn and E. Kishel, "Phonological neutralization of covert contrast: Evidence from fraternal twins acquiring initial clusters in English," poster to be presented at *2003 Child Phonology Conf.*, UBC, 2003.
- [3] B. Dodd and S. McEvoy, "Twin language or phonological disorder?" *Journal of Child Language*, vol. 21, pp. 273-289, 1994.
- [4] K. Forrest, G. Weismer, P. Milenhovic, and R. Dougall, "Statistical analysis of word-initial voiceless obstruents: Preliminary data," *JASA*, vol. 84, pp. 115-124, 1988.
- [5] A. Jongman, R. Wayland, and S. Wong, "Acoustic characteristics of English fricatives," *JASA*, vol. 108, pp. 1252-1263, 2000.
- [6] J. Kornfeld, "Theoretical issues in child phonology" *CLS*, vol. 7, pp. 454-69, 1971a.
- [7] J. Kornfeld, "What initial clusters tell us about a child's speech code," *Quarterly Progress Reports, MIT RLE*, vol. 101, pp. 218-221, 1971b.
- [8] J. Kornfeld and H. Goehl, "A new twist to an old observation: kids know more than they say," *CLS Parasession on Natural Phonology*, pp. 210-219, 1974.
- [9] L. Louko, *Acoustic and Perceptual Analysis of Word-Initial Consonant Clusters in Children With Normal and Disordered Phonology*, Syracuse University PhD dissertation, 1998.
- [10] M. Macken and D. Barton, "The acquisition of voicing contrasts in English: a study of voice onset time in word-initial stop consonants," *Journal of Child Language*, vol. 7, pp. 41-74, 1980.
- [11] P. Menyuk, "Clusters as single underlying consonants: Evidence from children's pronunciations," in *Proceedings of the Seventh International Congress of Phonetic Sciences*, A. Rigault and R. Charbonneau, Eds., pp. 1161-1165. The Hague, Netherlands: Mouton, 1972.
- [12] P. Menyuk, and D. Klatt "Child's production of initial consonant clusters," *Quarterly Progress Reports, MIT RLE*, vol. 91, pp. 205-213, 1968.
- [13] C. Ringo, *The Nature of Change in Phonological Development: Evidence from the Acquisition of /s/+stop and /s/+nasal Clusters*, Brown University PhD dissertation, 1985.
- [14] J. Scobbie, F. Gibbon, W. Hardcastle, and P. Fletcher, "Covert contrast as a stage in the acquisition of phonetics and phonology," in *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, M. Broe and J. Pierrehumbert, Eds., pp.194-207. Cambridge, England: Cambridge University Press, 2000.