Development of Coarticulatory Patterns in Spontaneous Speech

Melinda Fricke and Keith Johnson
University of California, Berkeley

This study seeks to better understand natural speech processes by examining coarticulatory patterns in the spontaneously produced speech of adults and children. Coarticulation, the process by which any articulatory gesture affects adjacent articulatory gestures, can be *anticipatory* (as when knowledge of an upcoming gesture affects the realization of the gesture currently being executed) or *perseverative* (when an already initiated gesture carries over onto the articulatory realization of a following gesture). We examine three acoustic measurements in an attempt to distinguish productions of [s] in round vs. non-round vowel contexts, and we compare the results for adults vs. children.

Experiment

Corpora. The adult data in this study come from the Buckeye Corpus of Conversational Speech (Pitt et al., 2007), and the child data come from the Davis corpus of the CHILDES database (Davis et al., 2002; MacWhinney, 2000). Both corpora were phonetically transcribed by their respective developers, making it possible to identify all instances of [s] in the context of either a high rounded vowel ([u ʊ o]) or an analogous unrounded vowel ([i ɪ e ɛ]) using an automated computer script. The child data were then hand segmented by the first author and a research assistant[1]. Table 1 shows the number of [s] tokens analyzed, broken down by corpus, direction of coarticulation (anticipatory vs. perseverative) and adjacent vowel type (round vs. non-round). For the child data, only tokens of [s] that occurred in identifiable words were used.

adult data	non-round	round	TOTAL
anticipatory	1362	1535	2897
perseverative	618	279	897
TOTAL	1980	1814	3794
child data	non-round	round	TOTAL
anticipatory	615	103	718
perseverative	1801	516	2317
TOTAL	2416	619	3035

Table 1 Number of [s] tokens included in the present analysis.

Procedure and analysis. We report three acoustic measures: high frequency centroid, amplitude ratio, and kurtosis. A description of each measurement is provided below. To calculate the acoustic measures, a 40 millisecond Hamming window was centered at three time points: 20%, 50%, and 80% of the fricative's duration. (Fricatives shorter than 100 ms were excluded from the analysis.) A fourth window, centered at 20 ms into the adjacent vowel, was used to generate vowel spectra for comparison. The spectra were then averaged across all tokens and used to produce the plots in Figure 1. The left panel of Figure 1 shows the averaged fricative spectra for the anticipatory direction, based on the 80% duration time slice, juxtaposed with the adjacent vowel spectra. The right panel shows the perseverative spectra for the 20% duration time slice, also juxtaposed with the adjacent vowel spectra. We now describe the measurements used to characterize the differences in spectral shape.

High frequency centroid. Following previous work, we calculated the high frequency centroid – the weighted mean frequency above the F2 region – since excluding the noise within

[1] Many thanks to Vanessa Chew for help segmenting.

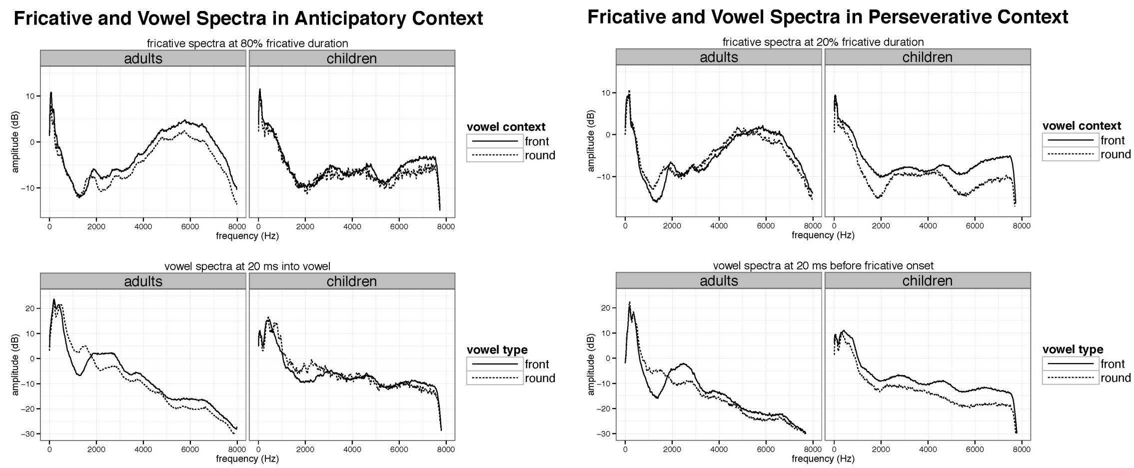

Figure 1 Fricative and vowel spectra for adults and children, in anticipatory and perseverative contexts.

the F2 region provides a better estimate of the size of the front cavity alone (Li et al., 2007). The F2 region was defined as 2500 – 3500 Hz for all children, 1500 – 2500 Hz for women, and 1125 – 2125 Hz for men (spanning the neutral tube F2 for vocal tract lengths of 8.75, 13, and 16 centimeters, respectively). These frequency ranges were chosen by visually examining plots of the averaged spectra for each group, identifying the approximate F2 peak, and defining a 1000 Hz band around that average peak.

Amplitude ratio. We also measured the amplitude of the F2 region relative to that of the highest peak above the F2 region (Li et al., 2007). For each fricative, we found the highest amplitude peak above the F2 region, determined the average amplitude within a 1000 Hz band centered on that peak, then subtracted the average amplitude within the F2 region from the average amplitude of the high frequency peak. A high amplitude ratio therefore indicates that the amplitude of the F2 region is relatively low, and we interpret a low F2 amplitude as resulting from a flatter, more palatal tongue posture (high amplitude ratio < relatively low F2 amplitude < less coupling between the front and back cavities < longer constriction). Note, however, that the gross differences in spectral shape between adults and children mean that the amplitude ratio measurement is better suited for making within group comparisons.

Kurtosis. Kurtosis, the fourth spectral moment, is commonly used to quantify the "peakedness" of a distribution. Greater values correspond to more defined peaks, which are a reliable indicator of greater lip rounding (Shadle & Mair, 1996). We computed kurtosis based on the procedures described in Forrest et al. (1988), with the modification of calculating the spectrum over a 40 ms Hamming window rather than a 20 ms one.

Statistical modeling. Separate statistical models were fit for the child and adult data, and for the anticipatory and perseverative data, resulting in four models for each dependent measure. Predictors were added in a step-wise fashion as follows. A random effect for speaker was included in each model. A fixed effect for measurement location (20% fricative duration vs. 80% fricative duration) was then added, followed by a fixed effect for context (round vowel context vs. non-round vowel context), followed by an interaction term. Non-significant predictors were dropped before additional predictors were added.

Results and Discussion

The raw means for the high frequency centroid, amplitude ratio, and kurtosis data are shown in Figures 3, 4, and 5, respectively, and a summary of the statistically significant predictors is provided in Table 2.

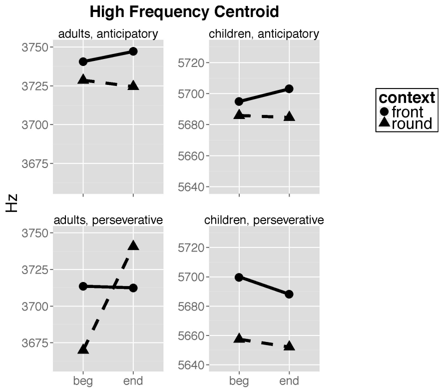

Figure 3

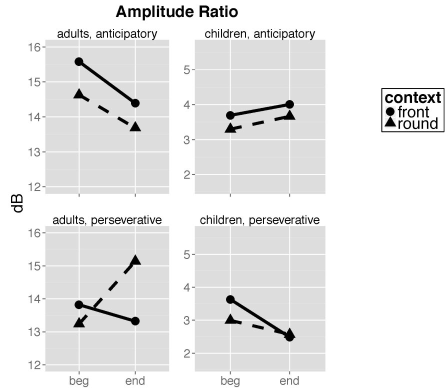

Figure 4

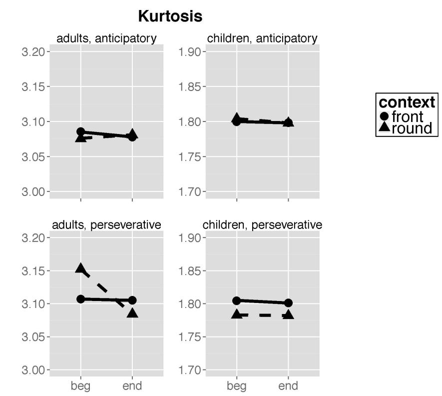

Figure 5

Centroid	intercept	time slice = 80 %	context = round	location x context
adult ant	3737 ***	--	-31 ***	--
adult pers	3703 ***	--	-45 ***	72 ***
child ant	5697 ***	--	-20 *	--
child pers	5693 ***	-10 ***	-20 *	--
Amp Ratio				
adult ant	15.1 ***	-1.1 ***	-1.0 ***	--
adult pers	13.4 ***	--	--	2.4 ***
child ant	3.63 ***	--	--	--
child pers	3.31 ***	-0.98 ***	--	--
Kurtosis				
adult ant	3.1 ***	--	0.01 ***	--
adult pers	3.1 ***	-0.01 *	0.05 ***	-0.07 ***
child ant	1.81 ***	--	--	--
child pers	1.81 ***	-0.003 *	--	--

Table 2 Summary of statistical predictors.
'***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$

Overall adult vs. child differences. As seen in Figure 1, and as reflected in the large between group differences in amplitude ratios and kurtosis, the child fricative spectra were overall much flatter than the adult spectra, with less well defined peaks. In adult productions of [s], the apex of the tongue forms a groove, directing a stream of air against the back of the teeth and generating obstacle turbulences that result in high frequency, high intensity noise. This noise is further shaped by the anterior cavity, resulting in a particularly intense, peaked distribution of high frequency noise.

The relative lack of high intensity, high frequency noise in the child productions suggests that the children in the present study were unable to direct a stream of air to hit the teeth (at least not with adult-like precision). The disperse nature of the high frequency noise may also reflect a combination of a more laminal articulation (Li et al., 2007) and a lack of shaping by the front cavity. We conclude that the global differences between the adult and child spectra suggest that the children in this study were unable to produce the tongue shape necessary to direct a stream of air to hit the teeth and may have instead formed a flatter, more palatalized constriction.

Anticipatory vs. perseverative differences. For both adults and children, for every dependent measure, the anticipatory data showed a magnitude of coarticulatory influence that was constant throughout the fricative. This suggests that both adults and children have already planned and prepared the articulators for the fricative and its following vowel by the time they begin the fricative articulation. Otherwise, the influence of the round vowel on the fricative would increase during the fricative.

We note that the child data showed no effect of vowel context on kurtosis. One interpretation of this finding is that the children did not produce significant lip rounding in the context of the phonologically round vowels, suggesting that the difference in the centroid data could additionally be explained by a difference in place of articulation. (The high frequency centroid correlates inversely with the size of the front cavity, which can be lengthened by lip rounding *and/or* by producing a more posterior constriction location.) Since all of the round vowels in this study were also back vowels, it is possible that the coarticulation evidenced by the children's high frequency centroids was at least partially due to anticipatory retraction of the tongue. This would be consistent with our explanation for the overall differences in spectral shape, in that it suggests only gross motor control of the tongue (i.e., a lack of differentiated control for the front vs. the back of the tongue).

In the perseverative direction, for every acoustic measurement examined, the adult data revealed a significant interaction term of vowel context and measurement location, while the child data did not. Adults' [s] productions were apparently quite dynamic, since the influence of the round vowel on the following fricative was relatively short lived (differing significantly from the beginning to the end of the fricative). The child perseverative data, however, indicate a static production; while children may have anticipated their upcoming fricative–vowel gestures to the same extent as adults, they were not similarly able to correct for perseverative coarticulatory differences.

Conclusion

In this paper, we implemented three acoustic measurements that were successfully used to describe coarticulatory patterns in spontaneous speech. Both adults and children exhibited anticipatory and perseverative vowel-on-fricative coarticulation. For children, perseverative coarticulation persisted throughout the fricative, while for adults, it was short-lived. This result, combined with the overall differences in spectral shape, suggests that the development of adult-like coarticulatory patterns depends on the acquisition of fine-grained motor control of the tongue, and on the appropriate phasing of lingual gestures with other gestures. Our results also suggest that the development of adult-like gestural phasing may proceed along different timelines depending on the direction of influence (i.e. anticipatory vs. perseverative).

References

Davis, B.L., MacNeilage, P.F., and Matyear, C.L. (2002). Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production. Phonetica *59*: 75–107.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R.N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. The Journal of the Acoustical Society of America *84*: 115 – 123.

Li, F., Edwards, J., and Beckman, M. (2007). Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese. Proceedings of the XVIth International Congress of Phonetic Sciences *4*: 917–920.

MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk, Volume II: The Database. Lawrence Erlbaum.

Pitt, M.A., L. Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., and Fosler-Lussier, E. (2007). The Buckeye corpus of conversational speech (2nd release*)*. Department of Psychology, Ohio State University.

Shadle, C.H. and Mair, S.J. (1996). Quantifying spectral characteristics of fricatives. Proceedings of the Fourth International Conference on Spoken Language *3*: 1521–1524.