

# Measuring coarticulation in spontaneous speech: a preliminary report

Melinda Fricke

Keith Johnson

University of California, Berkeley

## Introduction: why study spontaneous speech?

Constantly improving processing capabilities have recently opened the door to a type of language data that was previously impossible to study: the phonetic properties of spontaneous speech. Within the past two decades in particular, studies of the acoustics of conversational speech have become more common, and more sophisticated. The present study seeks to further expand our understanding of natural speech processes by examining coarticulatory patterns in the spontaneously produced speech of adults and children.

The study of spontaneous speech can shed considerable light on the workings of the language production system. In normal, everyday conversation, speakers are under quite different (and likely more stringent) pressures than in a laboratory setting. Laboratory speech is typically carefully controlled and often minimally creative; subjects often read words or sentences from a computer screen, repeating the same sentence or type of sentence many times, thus eliminating the usual need to go from concept to sentence construction to articulation. While laboratory studies have been invaluable for developing models of language production, the true testing ground for such models must be spontaneous speech, since it uniquely reveals how language production must proceed in the real world, in real time.

The present study is concerned with coarticulatory patterns in particular. 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). This study will be concerned with vowel-on-fricative coarticulation: the present research question is whether acoustic measurements can be identified that distinguish fricatives in a round vowel context from fricatives in a non-round vowel context in spontaneously produced speech. In addition to the main research question, the acoustic measurements under investigation will be ap-

<b>adult data</b>	non-round	round	TOTAL
anticipatory	1362	1535	2897
perseverative	618	279	897
TOTAL	1980	1814	3794

Table 1: Number of [s] tokens analyzed from the Buckeye Corpus.

<b>child data</b>	non-round	round	TOTAL
anticipatory	615	103	718
perseverative	1801	516	2317
TOTAL	2416	619	3035

Table 2: Number of [s] tokens analyzed from the Davis Corpus.

plied to corpora of adult and child speech, and will be used to compare anticipatory versus perseverative coarticulation.

## Experiment

### Method

#### Corpora

The adult corpus used in the present study is the Buckeye Corpus of Conversational Speech [Pitt et al., 2007]. Forty adults (twenty men, twenty women) were recorded in natural, sociolinguistic-style interviews, for a total of one hour each. The child data come from the Davis corpus of the CHILDES database [Davis et al., 2002, MacWhinney, 2000]. Twenty-one children were each recorded for one hour per week, for a period of several weeks to months. Both corpora were already 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 ε]) using an automated computer script. The child data were then hand segmented by the first author and a research assistant<sup>1</sup>.

Tables 1 and 2 show the number of [s] tokens analyzed, broken down by the direction of coarticulation (anticipatory vs. perseverative) and the adjacent vowel type (round vs. non-round). For the child data, only tokens of [s] that occurred in identifiable words were used. However, we did not exclude tokens that differed from the adult target pronunciation. For example, if the word *nose* was transcribed [nos] (and it was), it was counted as an instance of round, perseverative coarticulation. Future analyses may investigate whether pronunciations that matched the adult target differed in some way from “adapted” pronunciations.

Of the 21 children who were recorded for the Davis corpus, 11 (5 boys) produced instances of [s] in identifiable words. The ages of the 11 children ranged from 1;1 (years;months) to 3;1,

<sup>1</sup>Many thanks to Vanessa Chew for help with the segmentation.

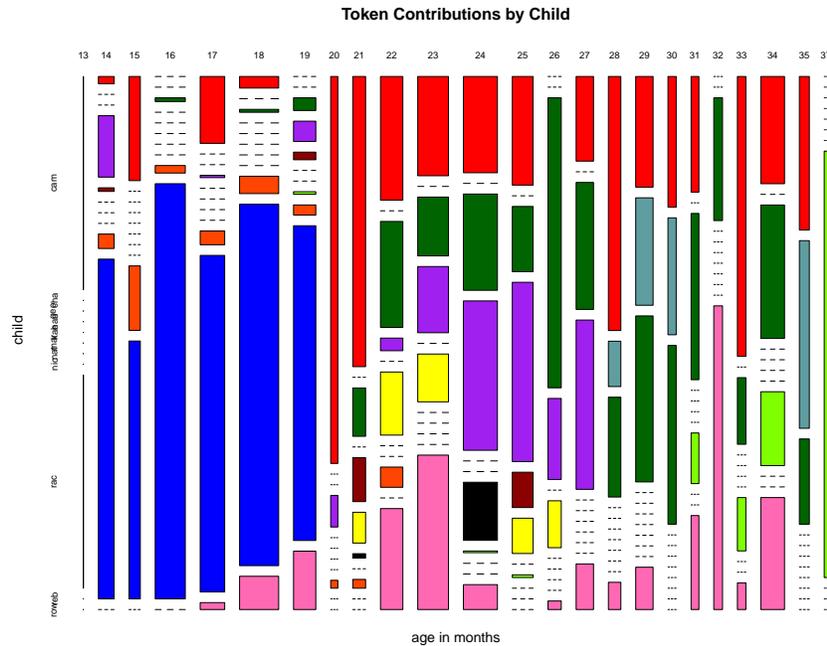


Figure 1: Proportion of data contributed by each child from the Davis Corpus, with each child represented by a unique color. See text for details.

with some children contributing far more data than others. Figure 1 is a graphical depiction of the number of [s] tokens contributed by each child, over time. Each column represents one month's worth of data, ranging from 13 months of age to 37 months of age. Each child is represented by a different color, and the height of each colored bar is proportional to the number of tokens contributed by that child for that month<sup>2</sup>.

### Procedure and analysis

Three acoustic measures will be reported in this paper: high frequency centroid, amplitude ratio, and kurtosis. A description of each measurement is provided in turn, below. In order to derive the acoustic measures, spectra were generated at four time points for each fricative token. A 40 millisecond Hamming window was centered at three points during the fricative: 20%, 50%, and 80% of the fricative's duration. Fricatives lasting less than 100 ms were excluded, such that the 40 ms window was always entirely contained within the fricative noise itself. A fourth window, centered at 20 ms into the adjacent vowel, was used to generate vowel spectra. The spectra were then averaged across all tokens and used to produce the plots in Figures 2 and 3. Figure 2 shows the averaged fricative spectra for anticipatory

<sup>2</sup>At present, we have yet to systematically investigate whether any of the children's pronunciations changed over time. While it is likely that time plays a significant role, we leave that analysis for a future report.

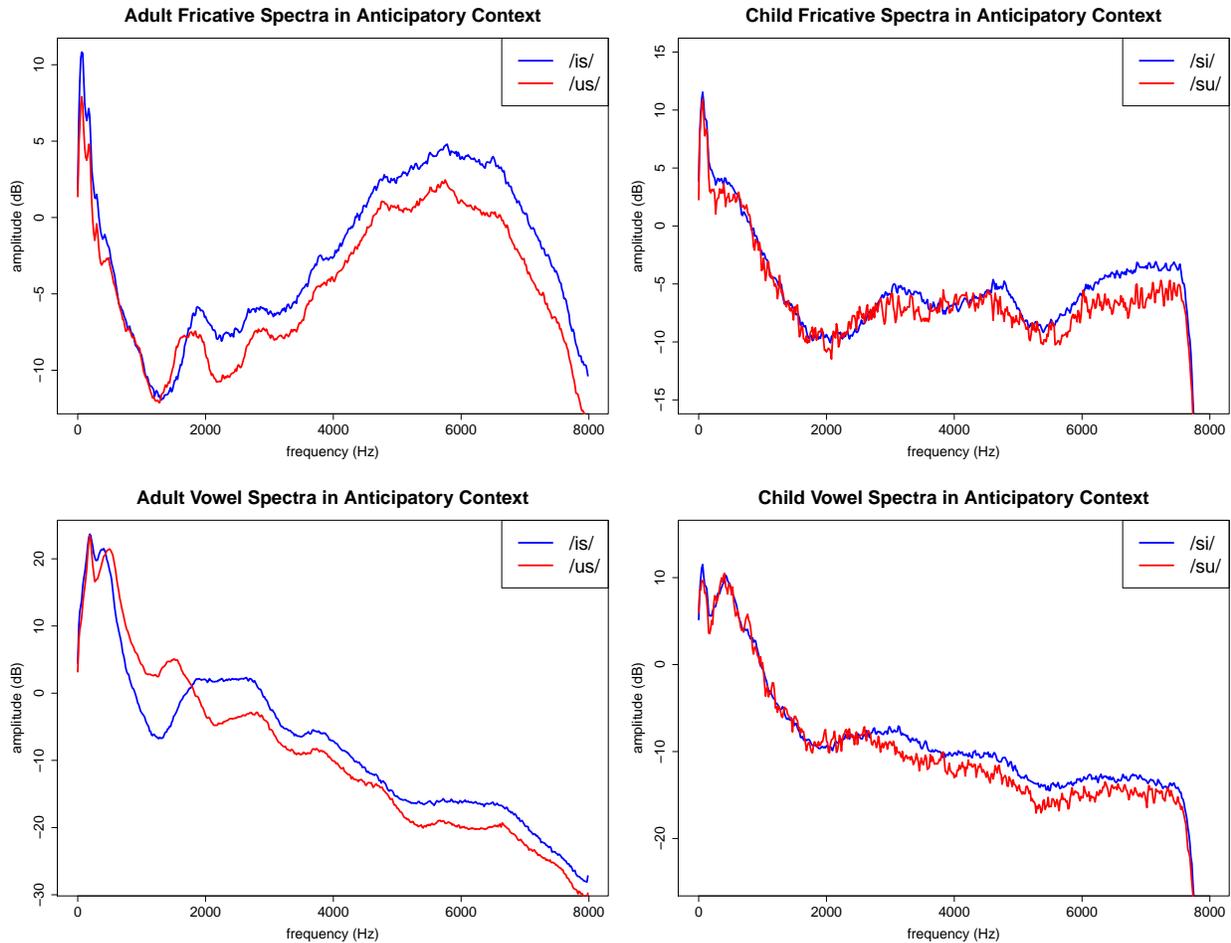


Figure 2: Averaged spectra generated 80% of the way into the fricative (top) and 20 ms into the adjacent vowel (bottom), preceding either a round or non-round vowel, for adult (left) and child (right) data.

coarticulation, based on the 80% duration time slice, juxtaposed with the adjacent vowel spectra. Figure 3 shows the perseverative coarticulation spectra based on the 20% duration time slice, juxtaposed with the vowel spectra for the adjacent vowel.

The most obvious difference between the adult and child spectra is their overall tilt and shape; the adult fricatives have clearly defined peaks (one around the region of the second formant, and a second, higher amplitude peak in the high frequency range), while the child spectra are much flatter, with less prominent high frequency noise<sup>3</sup>. These shapes reflect important differences in how the adults and children in this study produced their [s]’s, and we will return to the interpretation of the spectral shapes in the general discussion. We now turn to a description of the measurements implemented to quantify the differences in

<sup>3</sup>Note that the overall flatter spectral shape does reduce the reliability of ‘moments’-based measures such as the centroid and kurtosis.

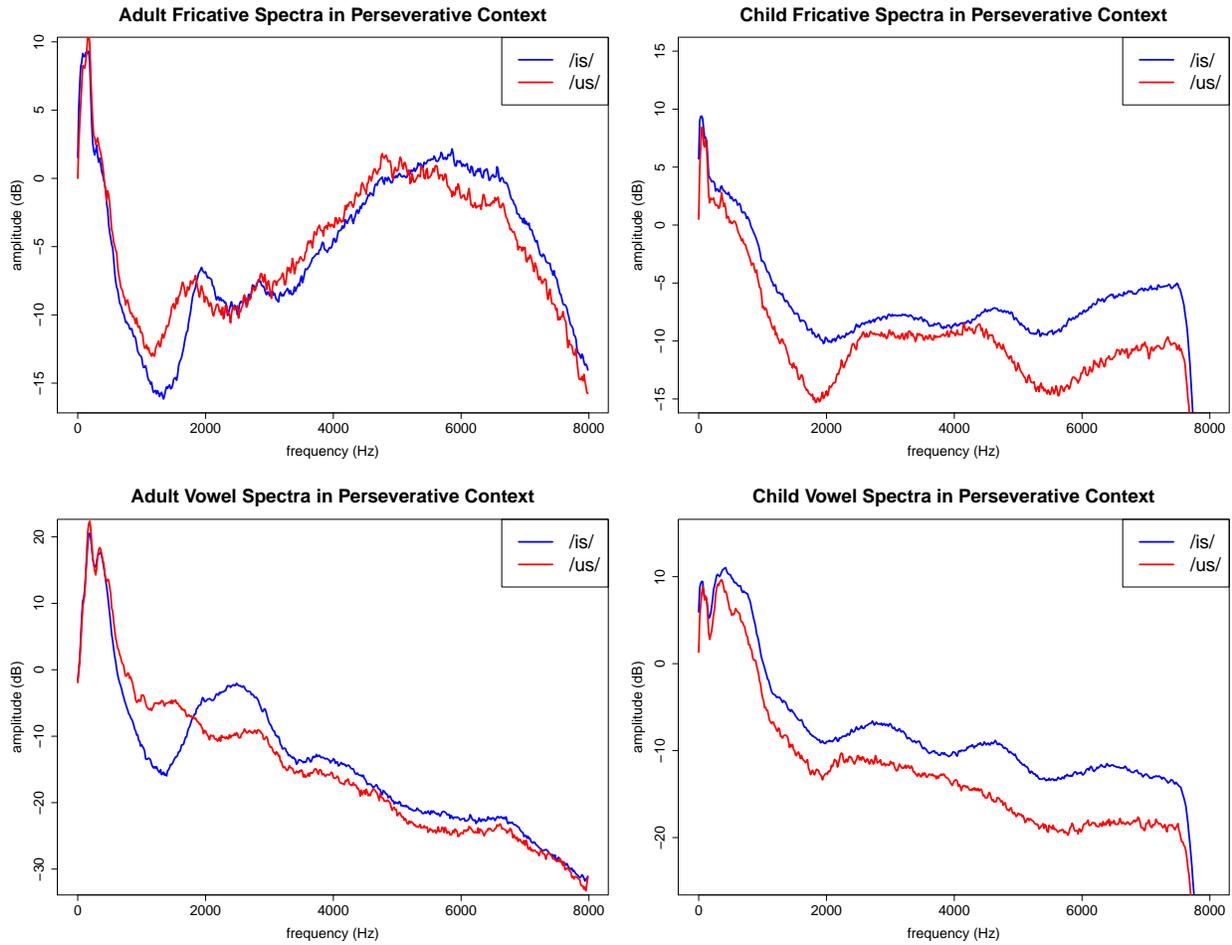


Figure 3: Averaged spectra generated 20% of the way into the fricative (top) and 20 ms into the adjacent vowel (bottom), following either a round or non-round vowel, for adult (left) and child (right) data.

spectral shape seen in Figures 2 and 3.

### **High frequency centroid**

The child spectra in Figures 2 and 3 reveal relatively prominent peaks in the region of the second formant (F2; around 3000 Hz for these children). McGowan and Nittrouer [1988] have argued that children's relatively high F2 amplitudes (as compared to adults) are due to coupling between atmosphere and the back cavities. Li et al. [2007] also point out that the length of the fricative constriction itself has an effect on the amount of coupling between the front and back cavities, with a shorter constriction length leading to more coupling. Li et al. therefore calculated the weighted mean of the frequency distribution *above* the F2 region, since excluding the noise within the F2 region provides a better estimate of the size of the front cavity during fricative production.

Therefore, in order to reduce the influence of the lower frequency noise present due to coupling with the back cavities, we also calculated the high frequency centroid, defined as the weighted mean frequency above the F2 region. The F2 region was set at 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**

To quantify the relative prominence of the F2 region, we also measured the amplitude of the F2 region relative to that of the highest peak above the F2 region. This measure was also used by Li et al. [2007], and a slightly different version was implemented by McGowan and Nittrouer [1988]. 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 a low value indicates that the amplitude of the F2 region is relatively high.

As discussed in the description of the high frequency centroid measurement, a relatively high F2 amplitude is correlated with greater coupling between the front and back cavities during fricative production, and greater coupling is related to a shorter constriction length. For this reason, we interpret a high value for amplitude ratio to be indicative of a flatter, more palatal tongue posture (high amplitude ratio < relatively low F2 amplitude < less coupling < longer constriction).

### **Kurtosis**

Kurtosis, the fourth spectral moment, has been used in many studies to quantify differences in fricative noise [Forrest et al., 1988, Jongman et al., 2000, Li et al., 2009]. It is generally

described as the “peakedness” of a distribution, with greater values corresponding to more defined peaks, which are typically a reliable indicator of greater lip rounding [Shadle and 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. The child models all included a random effect for speaker and for word, while the adult models included a random effect for speaker only<sup>4</sup>. Predictors were added in a step-wise fashion as follows. A fixed effect for measurement location (20% fricative duration vs. 80% fricative duration) was first added to the model, 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.

The decision was made to only test for coarticulation effects at the beginning and end of the fricatives in order to facilitate interpretation of the models. In all cases, measurements taken at the fricative midpoint were significantly more extreme than either the 20% or 80% duration time points, due to the trajectory of the articulatory gesture (the tongue reaches its most extreme position at or near the midpoint of the fricative). Since this is true across the data set, and since any coarticulatory effects present at the fricative midpoint were also present either at the beginning or end of the fricative, the statistical models were fit to only the 20% and 80% duration time points for simplicity’s sake.

## **Results**

### **High frequency centroid**

The raw means for the high frequency centroid data are shown in Figure 4, and a summary of the statistically significant predictors of centroid frequency are provided in Table 3 for the anticipatory direction, and in Table 4 for the perseverative direction. For both adults and children, the only significant predictor in the anticipatory direction was a main effect of round vowel. In the perseverative direction, the models for adults versus children were quite different. The adult model yielded a main effect of round vowel, and a highly significant interaction between vowel and measurement location, indicating that the effect of the round vowel was completely gone by the end of the fricative. For the child data, there was a main effect of measurement location (the centroid frequency was slightly lower at the end of the fricative), and a main effect of round vowel (the centroid frequency was nearly 20 Hz lower

---

<sup>4</sup>The individual word data were unavailable for the adult analysis at the time of writing. However, the addition of the random effect for word to the child models resulted in no qualitative changes, suggesting that the adult results will remain qualitatively the same once the random effect for word is added.

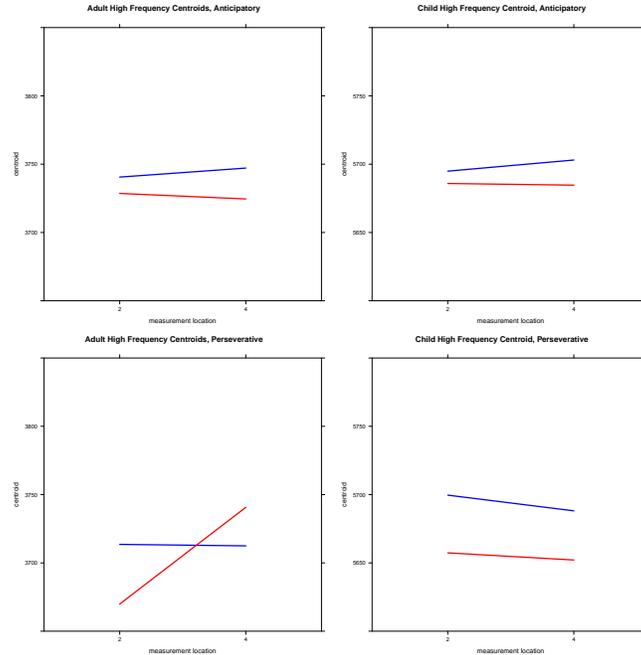


Figure 4: High frequency centroid measurements taken at the 20% and 80% duration time points in the fricative, averaged across all tokens. Lower values (red lines: round vowel context) indicate a longer front cavity, caused by lip rounding and/or retracted place of articulation. Left: adult data. Right: child data. Top: anticipatory coarticulation. Bottom: perseverative coarticulation.

following a round vowel, with the lack of interaction indicating that this was true for the duration of the fricative).

### Amplitude ratio

The results for the amplitude ratio data are shown in Figure 5, and the statistical models are given in Tables 5 and 6. For the anticipatory data, the adult model yielded significant effects of both measurement location and round vowel, with both predictors resulting in slightly lower amplitude ratios (recall that a lower ratio indicates shorter constriction length, or less palatalization). The child anticipatory amplitude ratios had no significant predictors.

In the models for the perseverative direction, the adult data yielded no significant main effects, but a significant interaction between measurement location and round vowel; higher amplitude ratios were found at the end of fricatives following round vowels. The child data showed a main effect of measurement location only; amplitude ratios were significantly lower at the end of fricatives. (The interaction term approached significance, at  $p = 0.10$ , but it was dropped from the model.)

adult anticipatory			child anticipatory		
predictor	coefficient	significance	predictor	coefficient	significance
intercept	3737	***	intercept	5696.9	***
location = 80%	–	n.s.	location = 80%	–	n.s.
round vowel	- 31	***	round vowel	- 19.6	*
interaction	–	n.s.	interaction	–	n.s.

Table 3: Statistical models for high frequency centroid data, anticipatory direction.

adult perseverative			child perseverative		
predictor	coefficient	significance	predictor	coefficient	significance
intercept	3703	***	intercept	5692.8	***
location = 80%	9.4	n.s.	location = 80%	- 9.8	***
round vowel	- 45	***	round vowel	- 19.6	*
interaction	72	***	interaction	–	n.s.

Table 4: Statistical models for high frequency centroid data, perseverative direction.

adult anticipatory			child anticipatory		
predictor	coefficient	significance	predictor	coefficient	significance
intercept	15.1	***	intercept	3.63	***
location = 80%	- 1.1	***	location = 80%	–	n.s.
round vowel	- 1.0	***	round vowel	–	n.s.

Table 5: Statistical models for amplitude ratio data, anticipatory direction.

adult perseverative			child perseverative		
predictor	coefficient	significance	predictor	coefficient	significance
intercept	13.4	***	intercept	3.31	***
location = 80%	- 0.08	n.s.	location = 80%	- 0.98	***
round vowel	- 0.20	n.s.	round vowel	–	n.s.
interaction	2.4	***	interaction	–	n.s.

Table 6: Statistical models for amplitude ratio data, perseverative direction.

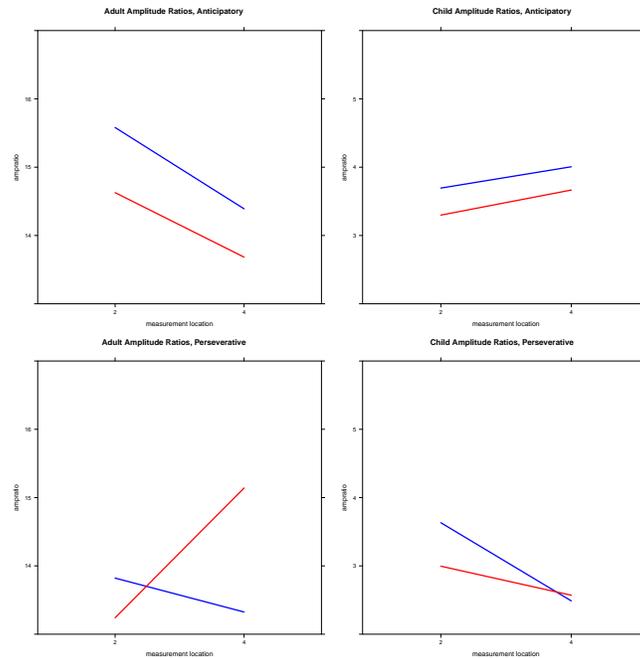


Figure 5: Amplitude ratio measurements taken at the 20% and 80% duration time points in the fricative, averaged across all tokens. Higher values (blue lines: non-round vowel context) indicate a more palatal tongue posture. Left: adult data. Right: child data. Top: anticipatory coarticulation. Bottom: perseverative coarticulation.

## Kurtosis

The kurtosis data are shown in Figure 6, and the statistical models are given in Tables 7 and 8. No significant predictors were found for any of the child data, suggesting that children's fricatives in the context of a round vowel were not reliably more rounded. For the adult anticipatory data, kurtosis was significantly higher preceding a round vowel. In the perseverative data, kurtosis was also significantly higher following a round vowel, but the effect was qualified by an interaction term; by the end of the fricative, the kurtosis value in a round vowel context was significantly higher than at the beginning. There was also a main effect of measurement location, such that with all other factors held constant, kurtosis was consistently slightly higher at the end of the fricative<sup>5</sup>.

<sup>5</sup>The statistical model seems to run contrary to Figure 6, which indicates that the random effect for speaker affected the pattern apparent from the raw means. Indeed, a model that allows the effect of measurement location to vary by speaker causes the main effect to disappear. Future analyses may further investigate these sorts of individual differences in coarticulatory patterns.

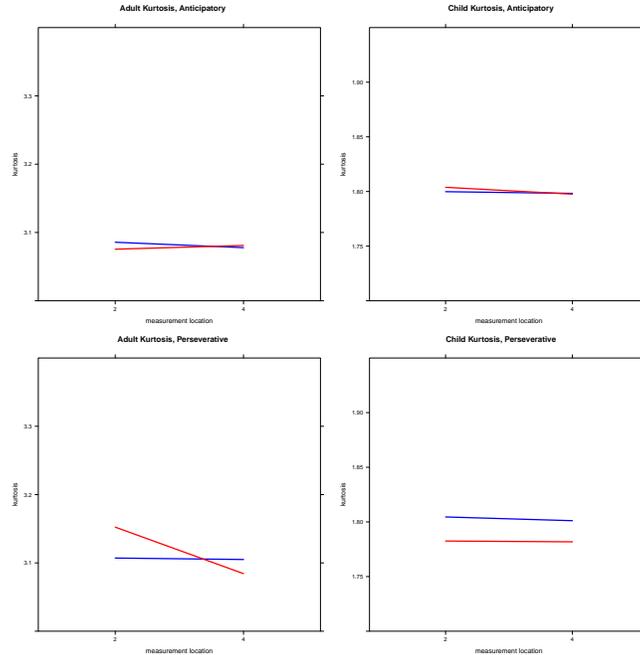


Figure 6: Kurtosis measurements taken at the 20% and 80% duration time points in the fricative, averaged across all tokens. Higher values (red lines: round vowel context) indicate a more peaked distribution, which is correlated with lip rounding. Left: adult data. Right: child data. Top: anticipatory coarticulation. Bottom: perseverative coarticulation.

## Discussion

This study examined three acoustic measurements that could be used to detect vowel-on-fricative rounding coarticulation in the spontaneous speech of adults and children. Important differences were found between the adult and child data, and between the anticipatory and perseverative directions of influence.

### Overall adult vs. child differences

As seen in Figures 2 and 3, and as reflected in the large differences in amplitude ratios (approximately 15 vs. 3.5 dB for adults vs. children, respectively) and in kurtosis (about

adult anticipatory			child anticipatory		
predictor	coefficient	significance	predictor	coefficient	significance
intercept	3.1	***	intercept	1.81	***
location = 80%	–	n.s.	location = 80%	–	n.s.
round vowel	0.01	***	round vowel	–	n.s.

Table 7: Statistical models for kurtosis data, anticipatory direction.

adult perseverative			child perseverative		
predictor	coefficient	significance	predictor	coefficient	significance
intercept	3.1	***	intercept	1.81	***
location = 80%	- 0.01	*	location = 80%	- 0.003	*
round vowel	0.05	***	round vowel	–	n.s.
interaction	-0.07	***	interaction	–	n.s.

Table 8: Statistical models for kurtosis data, perseverative direction.

3 vs. 1, respectively), the child fricative spectra were overall much flatter than the adult spectra, with less well defined peaks. The articulation of [s] by adult English speakers results in a particularly prominent high frequency peak, as the fricative source and filter characteristics combine to produce an intense concentration of high frequency noise. For an adult [s], the apex of the tongue typically 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 high frequency, high intensity noise source is then shaped by the anterior cavity (the portion of the vocal tract in front of the fricative constriction), resulting in the intense, peaked distribution of high frequency noise that characterizes adult English [s].

The flatness of the child spectra are likely due to differences in both source and filter characteristics. The relative lack of high intensity, high frequency noise suggests that the child fricatives were not produced with appreciable obstacle turbulences; the children in the present study were likely unable to direct a stream of air to hit the teeth (at least not with adult-like precision), resulting in overall less intense high frequency noise. With respect to the filter characteristics, the disperse nature of the high frequency noise may reflect a combination of a more laminal articulation [Li et al., 2009] and a lack of shaping by the front cavity; labial fricatives, for example, are characterized by a nearly non-existent front cavity, disperse noise, and correspondingly low kurtosis values [Jongman et al., 2000].

Overall, then, the global differences between the adult and child spectra suggest that the children in this study possessed only gross motor control of the tongue. They appear to have been 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, resulting in less peaked, less intense high frequency noise.

### Anticipatory vs. perseverative differences

For both the adults and children, in every dependent measure examined, the anticipatory data showed a magnitude of coarticulatory influence that was constant throughout the fricative. For the adults, the high frequency centroid, amplitude ratio, and kurtosis all showed significant differences according to vowel context that neither increased nor decreased significantly during the fricative. For children, there were no significant differences in amplitude ratio or kurtosis according to vowel context, but the high frequency centroid data did show

a main effect of round vowel context. 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 amount of vowel-on-fricative coarticulation would increase during the fricative.

Of note, however, is the fact that the child anticipatory data showed a significant effect of vowel context on the high frequency centroid data only (no significant predictors obtained for either the amplitude ratio or kurtosis data). One interpretation of this finding is that the children did not produce significant lip rounding in the context of the phonologically round vowels. Indeed, the apparent difference in F2 and F3 between round and non-round vowels in the lower right plot of Figure 2 is quite small. If children did not in fact produce much lip rounding for round vowels, then the difference in the high frequency 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 “rounding” coarticulation measured by the difference in children’s high frequency centroids was at least partially due to retraction of the tongue in anticipation of the tongue dorsum constriction needed for the upcoming back vowel. This interpretation would be consistent with the “gross motor control” explanation of the overall differences in spectral shape that was proposed in the previous section.

Interestingly, then, both adults and children seem to have planned their fricative gestures in concert with the vowel gestures that follow them as early as the fricative onset, even if the gestures themselves are quite different.

While the findings for anticipatory coarticulation were thus quite similar regardless of speaker age, the findings for perseverative coarticulation differed in an important way for adults versus children. 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. For perseverative coarticulation, then, adults’ productions were quite dynamic; whatever the influence of the round vowel on the following fricative (lip rounding, constriction location, tongue posture), it seems to have been relatively short lived, since the centroid, amplitude ratio, and kurtosis values for fricatives that followed a round vowel all differed significantly from the beginning to the end of the fricative. There was no such interaction in the child perseverative data, however, suggesting a static production; while children may have planned their upcoming fricative–vowel gestures to the same degree as adults, they were not similarly able to correct for perseverative coarticulatory differences.

## **Relation to previous findings**

The picture that emerges is largely consistent with previous studies of child fricative coarticulation, even though the present study examined a younger group of children and used a somewhat different methodology. Zharkova et al. [2012], for example, used ultrasound imaging techniques to examine Scottish children’s production of [s] preceding /i u a/ and concluded that the lack of robust overall difference in tongue shape was due to a lack of fine

motor control; they concluded that the children in their study (aged 7;7) were unable to produce an initial [s] sound using the tongue tip while simultaneously adjusting the placement of the tongue body for the upcoming vowel.

The data in the present study are of course rather different, in that we were unable to directly observe the position of the articulators, but rather have attempted to infer their position based on acoustic data. The lack of a difference in kurtosis for [s] in a round vowel versus non-round vowel context suggests that the children in our study did not reliably produce lip rounding coarticulation. However, the significant difference in the high frequency centroid data depending on the vowel context indicates that children did produce *some* articulatory difference between [s] in the two contexts, and we suggest the difference may be one of tongue retraction in the context of a (back) round vowel. This hypothesis is consistent with the idea of gross motor control of the speech articulators, in that children may have anticipated the overall position of the tongue body, but were unable to differentially control the tongue tip versus the tongue dorsum. Further, it is not necessarily inconsistent with Zharkova et al. [2012]’s finding of no difference in overall tongue shape between vowel contexts; it is possible that the children in their study were unable to manipulate overall tongue shape, but *were* able to advance or retract the whole tongue, just as we have proposed for the children in the present study.

## Conclusion

In this paper, we implemented three acoustic measurements that were successfully used to describe coarticulatory patterns in spontaneously produced speech. While several important potential factors remain relatively unexplored (individual differences, word-specific differences, changes in children’s coarticulatory patterns over time, etc.), our findings generally square nicely with related patterns previously documented in the literature. Importantly, then, the methods and preliminary results reported here present several promising avenues for future research.

## References

- B.L. Davis, P.F. MacNeilage, and C.L. Matyear. Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production. *Phonetica*, 59(2-3):75–107, 2002.
- K. Forrest, G. Weismer, P. Milenkovic, and R.N. Dougall. Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America*, 84:115, 1988.
- A. Jongman, R. Wayland, and S. Wong. Acoustic characteristics of english fricatives. *The Journal of the Acoustical Society of America*, 108(3):1252–1263, 2000.

- F. Li, J. Edwards, and M. Beckman. Spectral measures for sibilant fricatives of english, japanese, and mandarin chinese. In *Proceedings of the XVIth International Congress of Phonetic Sciences*, volume 4, pages 917–920, 2007.
- F. Li, J. Edwards, and M.E. Beckman. Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in english and japanese toddlers. *Journal of Phonetics*, 37(1):111–124, 2009.
- B. MacWhinney. *The CHILDES Project: Tools for Analyzing Talk, Volume II: The Database*, volume 2. Lawrence Erlbaum, 2000.
- R.S. McGowan and S. Nittrouer. Differences in fricative production between children and adults: evidence from an acoustic analysis of / / and /s/. *The Journal of the Acoustical Society of America*, 83:229, 1988.
- M.A. Pitt, L. Dilley, K. Johnson, S. Kiesling, W. Raymond, E. Hume, and E. Fosler-Lussier. *The Buckeye corpus of conversational speech (2nd release)*. [www.buckeyecorpus.osu.edu](http://www.buckeyecorpus.osu.edu), Department of Psychology, Ohio State University (Distributor), 2007.
- C.H. Shadle and S.J. Mair. Quantifying spectral characteristics of fricatives. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, volume 3, pages 1521–1524. IEEE, 1996.
- N. Zharkova, N. Hewlett, and W. Hardcastle. An ultrasound study of lingual coarticulation in /sv/ syllables produced by adults and typically developing children. *Journal of the International Phonetic Association*, 42(2):193–208, 2012.