

Spectral and temporal measures of coarticulation in child speech

Margaret Cychosz¹
Department of Linguistics
University of California, Berkeley, Berkeley, CA 94720, USA
mcychosz@berkeley.edu

Jan R. Edwards
Department of Hearing and Speech Sciences
University of Maryland-College Park, College Park, MD 20724, USA
edwards@umd.edu

Benjamin Munson
Department of Speech-Language-Hearing Sciences
University of Minnesota, Twin Cities, Minneapolis, MN 55455, USA
munso005@umn.edu

Keith Johnson
Department of Linguistics
University of California, Berkeley, Berkeley, CA 94720, USA
keithjohnson@berkeley.edu

¹ Corresponding author

MEASURING COARTICULATION IN CHILD SPEECH

Abstract: Speech produced by children is characterized by a high fundamental frequency which complicates measurement of vocal tract resonances, and hence coarticulation. Here two whole-spectrum measures of coarticulation are validated, one temporal and one spectral, that are less sensitive to these challenges. Using these measures, consonant-vowel coarticulation is calculated in the speech of a large sample of four-year-old children. The measurements replicate known lingual coarticulatory findings from the literature demonstrating the utility of these acoustic measures of coarticulation in speakers of all ages.

© 2019 Acoustical Society of America

Keywords: coarticulation, speech development, Mel-frequency, spectrum, validation

1. Background

Coarticulation, or the articulatory overlap of speech segments, reflects a crucial equilibrium between speaker efficiency and listener comprehension and has played numerous illustrative roles for phonetic theory (Bradlow, 2002; Matthies, et al., 2001; Whalen, 1981). It served as an empirical motivation for Articulatory Phonology, a theory that takes individual, overlapping gestures as the basic units of speech (Browman & Goldstein, 1989). Its ubiquity and relevance for accurate perception also challenge purely abstractionist models (Lahiri & Marslen-Wilson, 1991).

For child language development, appropriate coarticulatory overlap indicates mature, adult-like speech. Consequently, coarticulation is a metric for the development of speech production and planning capabilities (Gerosa, et al., 2006; Nittrouer, et al., 1989; Zharkova, et al., 2014). Because child speech is characterized by underdeveloped motor and articulatory schemata (Green, et al., 2000; Smith & Goffman, 1998), it may logically follow that children would also have immature coarticulatory patterns. However, despite this logic and despite the fact that children speak slower and with less coordinated movement, which would suggest less coarticulation, much research into coarticulatory development suggests that children coarticulate *more* than adults (Gerosa, et al., 2006; Nittrouer et al., 1989; Rubertus, et al., 2013; Zharkova et al., 2011). The question of whether children coarticulate more or less than adults remains unanswered with the literature suggesting mixed results (Barbier et al., 2013, 2015; Goffman et al., 2008; Noiray et al., 2013; Noiray et al., *to appear*; Rubertus & Noiray, 2018; Whiteside & Hodgson, 2000).

To determine whether or not children coarticulate like adults, coarticulation must be quantified using valid, replicable, and, ideally, automated acoustic measures. However, from infancy into puberty (when fundamental frequencies lower to adult levels), the child speech apparatus creates multiple issues for the study of acoustic phonetics in general and spectral analyses in particular (Hillenbrand et al., 1995; Vorperian & Kent, 2007). Small vocal folds and high fundamental frequency results in widely spaced harmonics in the spectral envelope (Ramsay, 2018). This can render an undersampled spectral shape obfuscating formant frequency peaks. Consequently, traditional formant-based measurements may be unreliable for young children's speech. This unreliability does not preclude the use of formant tracking in child speech. However, often the only remedy for formant tracking errors in young, high voices is to

make arbitrary data cleaning decisions such as excluding all measurements outside of a predetermined range (Lee et al., 1999), to painstakingly hand-check individual formant peaks in spectral slices (Nittrouer, 1993), or to rely on only those data points where formant measurements could reliably be found (approximately 50% in Nittrouer et al., 1989). Hand-checking may be unrealistic for studies with large sample sizes or if the formant peaks are not visible in the spectral structure.

Acknowledging these difficulties inherent to child acoustics, Gerosa et al. (2006) employed two novel acoustic measures of coarticulation to study consonant-vowel transitions in adult and child speech. The first, a spectral measure, calculates the distance between Mel-frequency cepstral coefficient (MFCC) vectors averaged over adjacent phones. The second measure, a temporal estimation, dynamically calculates the transition duration between phones in a given CV sequence as a function of their spectral overlap (explained in further detail in 3.1). This measurement reflects what proportion of the CV sequence is spent in *transition* where a greater proportion of transition time indicates more coarticulation (i.e. a larger chunk of the CV sequence is devoted to transitioning between segments).

The applicability of traditional acoustic measures of coarticulation, such as formant transitions (Lehiste & Shockey, 1972; Öhman, 1966) or the more dynamic Peak ERB_N (Reidy et al., 2017), may be limited to speakers with longer vocal tracts or to certain segments such as fricatives. However, the measures employed in Gerosa et al. (2006) rely on a cepstral representation of the audio signal, as the frequency scale of FFT spectra is transformed to the (log) Mel scale and a discrete cosine transformation is applied. This method is superior to formant tracking because it is a measure of distance between two overall *shapes* rather than a measure based on potentially unreliably tracked peaks in the spectral envelope. Consequently, these measures should be versatile and reliable for a broader range of speakers and consonant manners. The primary objective of this paper is to validate these two relatively novel acoustic measures of coarticulation to ensure their applicability for young children’s speech and a variety of consonants.

2. Current study

2.1 Calculations

Following Gerosa et al. (2006), we quantified coarticulation using two automatically-extracted acoustic measures, one spectral and one temporal. Both measures were made using

custom Python scripts running Librosa functions (McFee et al., 2015). And analyses were made over a single cohort of four-year-old children.

The spectral coarticulation measure is the difference between the averaged Mel-frequency log magnitude spectra from two adjacent phones. To compute this, the acoustic signal was first downsampled to 12 kHz. Then, each phone was segmented into 25.6ms frames, with a 10ms step. All the Mel-frequency spectral vectors from a given phone were then averaged. Finally, we measured the Euclidean distance between the averaged Mel spectral vector for both phones in the relevant biphone sequences for each word as displayed in Eq.(1):

$$d_{sa} = \sqrt{\sum (\bar{x}_s - \bar{x}_a)^2} \quad (1)$$

where d_{sa} is the Euclidean distance between segments /s/ and /a/ in the biphone sequence /sa/ (for example), and \bar{x}_s and \bar{x}_a are the averaged Mel spectral vectors of each segment. Unlike Gerosa et al. (2006), who computed the averaged MFCC vector from each adjacent phone, we did not apply a discrete cosine transformation to the Mel-frequency spectra to compute MFCCs because the compression of Mel spectra to MFCC can result in the loss of acoustic information.

We also implemented Gerosa et al.’s temporal coarticulation measure. This measure reflects the duration of the transition between adjacent phones. The acoustic signal was again downsampled to 12 kHz with each phone was segmented into 25.6ms frames with a 10ms step. The region of the transition duration was determined dynamically, based on acoustic difference between a given Mel-frequency spectral frame and the average spectrum of each phone. As in Gerosa et al. (2006), this first required that we compute a function for the distance between each sampled spectrum and the average Mel-frequency spectrum for each phone in the sequence as shown in Eq.(2):

$$f_{sa}(i) = d(\bar{x}_s, x_i) - d(\bar{x}_a, x_i) \quad (2)$$

where \bar{x}_s is the average Mel spectral vector for /s/, \bar{x}_a is the same for /a/. i is the spectral vector to be compared to the average spectrum (iteratively sampled over the phone), and d denotes the distance between the single spectral vector and the averaged spectral vector for that phone. The function $f(i)$ centers around zero and is negative over the first segment and positive over the second segment in the biphone sequence.

The number of frames where $f(i)$ is between an upper and lower bound is n and $n \cdot t$ is the duration of the transition in milliseconds, with step size $t=10\text{ms}$. The transition region, determined by the upper and lower bounds, was set to be the portion of $f(i)$ that spanned the middle 80% of the range $f(i)$. Transition duration was then scaled by the duration of the CV sequence dur_{sa} to compute the *relative* transition duration between phones in the CV sequence as shown in Eq.(3).

$$\frac{n \cdot t}{dur_{sa}} \quad (3)$$

2.2 Hypotheses

We make two important predictions regarding coarticulation in CV sequences:

1) Place of vowel articulation: In fricative-vowel sequences, fricative segments consistently show assimilatory effects to the following vowel. For example, in anticipation of the lip rounding required for [u], peak fricative frequencies are lower in the sequences [su] and [ʃu] than [si] and [ʃi] (Soli, 1981), reflecting anticipation of the upcoming round vowel.

Furthermore, larger distances traveled along the palate during the articulation of a CV sequence result in increased coarticulatory influence of one phone on another when compared to segments that are articulated in the same region. For two biphone sequences of equal duration, speakers may be more capable of differentiating the fricative and vowel in [sæ] than in [su] due to the time constraints of articulating both segments in a given window.

Anticipatory coarticulation in fricative-vowel sequences is one of the most well-documented cases of coarticulatory influence in the literature: fricatives articulated at or behind the alveolar ridge consistently demonstrate anticipatory coarticulation effects that vary by vocalic context, particularly before front and round vowels, in a variety of languages (Hoole et al., 1993; Mann & Repp, 1980; Soli, 1981; Whalen, 1981). Fricatives articulated at the alveolar ridge show more evidence of the upcoming vowel when that vowel is both front and round than when the vowel is not front and round.

For the current measurements, we first predict a *smaller* Euclidean distance between adjacent phones in [su] than [sæ], reflecting the greater influence of [u] on [s] than [æ] on [s]. In addition, we predict that sequences requiring a lingual transition from the palatal ridge to the

velar region, such as [su], will have a *longer* transition duration than segments such as [sæ], reflecting the increased movement required to articulate [s] and [u].

2) Manner of articulation: Consonant manner is a predictor of coarticulatory patterning with some manners demonstrating more COARTICULATORY RESISTANCE, or restraint from the coarticulatory influence of an adjacent segment, than others (Recasens & Espinosa, 2009; Recasens et al., 1997). Coarticulatory resistance decreases with lingual contact. Supra-glottal fricatives, for example, have a smaller surface contact area at the palate than glides which explains why anterior fricatives resist the influence of adjacent segments better than labiovelars or vowel-like rhotics (Recasens, 1985).

The relationship between coarticulatory resistance and lingual contact also interacts by speech articulator with segments realized with more sluggish articulators, such as the tongue dorsum, unable to resist coarticulatory influence as well as consonants articulated with the tongue blade (Mooshammer et al., 2006; Recasens & Espinosa, 2009).

We attempt to replicate these patterns of coarticulatory resistance in a hierarchy of sounds with different amounts of lingual contact and tongue dorsum involvement: alveolar fricatives > alveopalatal affricates > labiovelar glides. In this hierarchy, alveolar fricatives should show maximal coarticulatory resistance because articulation 1) involves the tongue tip (minimal palatal contact and tongue dorsum uninvolved) and 2) is highly constrained (to generate noisy turbulence). Alveopalatal affricates should exhibit relatively less resistance because tongue position is more flexible and lingual contact more fleeting (i.e. could be articulated at several points along the horizontal dimension to similar acoustic effect). Finally, labiovelar glides should show the least resistance because of a large area of lingual contact and articulation with a sluggish articulator (dorsum). This order by manner of articulation should translate to a *smaller* Euclidean distance between glide-vowel sequences than affricate-vowel sequences and smaller distance between affricative-vowels than fricative-vowels. For the temporal measure, we anticipate that glide-vowel sequences will show a longer transition duration than affricate-vowel and fricative-vowel, in that order.

To validate the novel temporal and spectral coarticulatory measures, we replicated these well-known coarticulatory patterns in a corpus of four-year-old children's speech recordings.

2.2 The corpus

MEASURING COARTICULATION IN CHILD SPEECH

Data for this study come from 103 four-year-old children (56 girls, 47 boys; range=3;3-4;4 [years;months], mean=3;5, SD= 0;3). All children were monolingual speakers of English. Children were participating in a longitudinal study of lexical and phonological development (see Mahr & Edwards [2018] for further detail). We report on data collected at the second of three timepoints. Each child passed a hearing screening in at least one ear at 25dB for 1000, 2000, and 4000Hz. Ninety (87.4%) of the children had normal speech and hearing development, per parental self-report. The 13 remaining children were identified as late talkers by their caregivers. However, the late talkers' scores on the series of language assessment tasks that we administered did not differ significantly from the remaining children. Consequently, data from all children were used in the current analysis.

For the data collection phase, each child completed a word repetition task where the participant repeated words after a model speaker. Children repeated a total of 94 words (including 4 training/practice items). All words contained a consonant-vowel (CV) sequence in word-initial position and were bisyllabic with penultimate stress. To ensure that children of this age would recognize words used in the task, we chose words from the MacArthur Bates Communicative Development Inventory (Fenson et al., 2007), the Peabody Picture Vocabulary Test-4 (Dunn & Dunn, 2007), and other sources (e.g. Morrison et al., 1997).

Here we analyze a subset of 5 of the original test items (Table 1). The first group of words, *sandwich* and *suitcase*, evaluates the place of articulation hypothesis by measuring the anticipatory coarticulation of [s] before [ae] versus [u]. The second group of words, *sister*, *chicken*, and *window*, tests manner of articulation by measuring the coarticulation between CV segments where the manner of consonant articulation varies but the vowel is constant.

TABLE 1 HERE

A young female speaker of Mainstream American English provided the recordings for the word stimuli. Recording prompts were digitized at a frequency of 44,100 Hz using a Marantz PMD671 solid-state recorder. Amplitude was normalized between words.

Each participant was guided through the repetition task by at least two experimenters. First, the child was seated in front of a computer screen and presented with a photo while the accompanying word played over external speakers. The child was then instructed to repeat the word. After each trial, the experimenter manually advanced to the subsequent trial. Stimuli were

presented randomly with E-prime software (Schneider et al., 2012). The task lasted approximately fifteen minutes.

2.3 Segmentation

We first scored the production accuracy of each CV sequence. Accuracy scoring was conducted offline in a feature-based system by a trained phonetician who is a native speaker of American English. Child participants had to produce the correct consonant voicing, manner of phone articulation, and place of articulation. Children additionally had to produce the correct height, length, and backness for the vowel and repeat the word's prosodic structure correctly (number of syllables, consonant in correct position, and vowel in correct position). Scoring was conducted auditorily and by reviewing the acoustic waveform. To ensure scoring accuracy, a second rater, also a trained phonetician and native speaker of American English, scored a 10% subset of the original words. An intraclass correlation (ICC) statistic assessed inter-rater agreement. The intraclass correlation between raters was 0.881, which was significantly greater than chance ($F(374,375)=15.9$, $p<.001$, 95% CI=0.86, 0.90). Only CV sequences that were produced correctly underwent acoustic analysis. Acoustic analysis and accuracy scoring were conducted on separate occasions for different research programs. The number of tokens for each word used in the current study is listed in Table 1.

The words that were repeated correctly then underwent acoustic analysis. Each correct CV sequence was manually transcribed in a Praat TextGrid (Boersma & Weenik, 2018) by a native speaker of American English who is a trained phonetician. The audio files were aligned using the visual representation from the waveform and spectrogram in addition to auditory analysis. As coarticulation measures are highly dependent upon segmentation decisions, we took a number of steps to standardize alignment. We developed alignment conventions for each CV sequence. The start of affricate/fricative-vowels corresponded to the onset of high-frequency energy in the spectrogram. For affricate/fricative-vowel sequences, the start of the vowel corresponded to the onset of periodicity in the waveform and formant structure in the spectrogram. These criteria were sufficient to demarcate all affricate/fricatives from vowels. Delimiting glide-vowel sequences was more gradient: a steady state formant delimited glide offset and vowel onset. Transcribers were encouraged not to rely on auditory analysis for glide-vowel segmentation decisions, which is notoriously problematic (McAllister Byun et al., 2016).

In the rare event that a steady-state formant could not be identified, 50% of the sequence was assigned to the consonant and 50% to the vowel.

A second transcriber, blind to the validation experiment objectives, independently aligned a 10% subset of the words. The difference between the coders' average consonant duration was 2 ms and the average difference in vowel duration was 10 ms. Pearson correlations between the coders were significant for consonants: $r=0.96$ $p<.001$, 95% CI=[0.95, 0.96] and vowels: $r=0.87$ $p<.001$, 95% CI=[0.85, 0.89], suggesting high fidelity to the alignment procedure. Despite these efforts, it is important to note that hand-segmentation is often highly subjective.

4. Results

We first evaluate the hypothesis that these acoustic measures of coarticulation should predict differences in anticipatory coarticulation in fricatives depending on the place of vowel articulation. Two mixed effects linear regression models were fit using the lme4 package in the R computing environment (Bates et al., 2015). Each model included Speaker as a random effect. One model predicted the temporal coarticulatory measure and the other the spectral. In both cases, the effect of Context significantly improved baseline model fit. Specifically, for the spectral model, there is a smaller distance between phones in the sequence [su] than [sæ] ($\beta=-1.56$, $t=-3.31$, $p=.002$), indicating greater coarticulation between [s] and [u] than [s] and [æ] as children anticipate the roundedness of [u] (Figure 1). In the temporal model, the transition duration between [s] and [u] is longer than [s] and [æ] ($\beta=1.08$, $t=1.99$, $p=.05$), again indicating greater coarticulation between the segments in [su]. Thus, both the temporal and spectral measures capture coarticulatory differences by place of articulation in fricative-vowel sequences in the vertical dimension (i.e. backness) and by vowel quality (roundedness), but the spectral model may be a more reliable indicator of anticipatory coarticulation for these segments.

FIGURE 1 HERE

Next, we evaluate the hypothesis that the coarticulatory measures should predict coarticulatory differences by consonant manner in CV sequences, after controlling for vowel identity. Two mixed effects linear regression models were again fit as before with Speaker as a random effect. The fixed effect Consonant Manner improved both model fits. Specifically, in the spectral model, [sɪ] reliably differed from [tʃɪ] ($\beta=-2.67$, $t=-4.74$, $p<.001$) and [wɪ] ($\beta=-3.19$, $t=-5.93$, $p<.001$) – [s] and [ɪ] were less acoustically overlapped than the segments in [tʃɪ] or [wɪ], suggesting less coarticulation. However, a post-hoc test with [tʃɪ] as the reference level

demonstrated that [tʃɪ] did not differ significantly from [wɪ] ($p=.78$). Still, the trend by consonant manner follows the anticipated direction: there was a larger acoustic distance between segments in [tʃɪ] (median=7.39, SD=4.19) than [wɪ] (median=6.72, SD=2.00) suggesting less coarticulation in [tʃɪ] than [wɪ] (Figure 2). For the temporal model, [sɪ] reliably differed from [tʃɪ] ($\beta=1.98$, $t=3.42$, $p<.001$) and [wɪ] ($\beta=7.71$, $t=14.04$, $p<.001$). Another post-hoc test also demonstrated that along the temporal dimension, [tʃɪ] differed significantly from [wɪ] ($\beta=5.56$, $t=5.71$, $p<.001$). The transition between segments in [wɪ] was longer than the transition between segments in [tʃɪ]. These results suggest that both the temporal and spectral coarticulation measures reliably capture known coarticulatory differences by consonant manner, after controlling for vocalic environment.

FIGURE 2 HERE

5. Discussion and Conclusion

In this validation study, we used two relatively novel acoustic measures of coarticulation, one temporal and one spectral, to replicate previous acoustic correlates of segmental coarticulation. Through a series of comparisons, we demonstrated that both of the tested acoustic measurements were generally robust enough to capture known patterns of coarticulation.

We first tested the hypothesis that the coarticulation measures would capture differences in fricative-vowel coarticulation by place of vowel articulation and vowel quality. Specifically, speakers are known to anticipate vowel quality, especially roundedness, in fricative-vowel sequences, and should exhibit increased coarticulation in sequences such as [su]. Furthermore, speakers should show anticipate the upcoming vowel in sequences with segments that differ in place of articulation, such as [su], than with segments that do not, such as [sæ], because the articulation of the former requires a transition from a lingual articulation at the alveolar ridge to an articulation towards the velum.

Our measures captured both of these coarticulatory patterns, though the spectral measure was more reliable. We found that speakers showed more acoustic overlap of phones, and a longer transition duration between phones, in the sequence [su] than the sequence [sæ], replicating known coarticulatory patterns by place of vowel articulation and vowel quality (Hoole et al., 1993; Mann & Repp, 1980; Soli, 1981; Whalen, 1981). However, acoustic measures of coarticulation are imperfect and acoustic similarity/transition duration does not necessarily indicate greater coarticulation. For example, if a speaker were already halfway to

MEASURING COARTICULATION IN CHILD SPEECH

hitting a vowel target at the beginning of a VC sequence, then their transition to the following consonant could be faster than a speaker who did not start at the same halfway point. Yet acoustic measures might say that these speakers “coarticulated” in different amounts, without acknowledging the underlying reasons for their dissimilarity.

Next, we attempted to capture differences in coarticulation by manner of consonant articulation. Consonants whose manner requires less lingual contact, particularly when realized with the tongue blade such as alveopalatal and alveolar stops, are able to resist coarticulation with adjacent segments more than consonants whose manner requires more lingual contact with the sluggish dorsum, such as rhotics and labiovelar glides (Recasens & Espinosa, 2009; Recasens et al., 1997). We replicated these patterns by manner using both coarticulation measures. As predicted, speakers coarticulated less in sequences with more resistant consonants in the following hierarchy: [sɪ] < [tʃɪ] < [wɪ].

These coarticulatory measures are important tools for speech research, particularly developmental. Both measures have broad applicability for a variety of consonant types – unlike center of gravity they are not limited to fricatives or sonorant sounds. Furthermore, the measures are relatively immune to the many challenges that children’s voices, generally breathy with high fundamental frequencies, bring to traditional acoustic analyses. Finally, these measurements can be made automatically, over small samples of speech, without equipment more specialized than a portable audio recorder. As a result, these measures may have broad applications for clinical populations or understudied groups. Researchers and clinicians can use the measures as an index of speech maturity in children or as a more fine-grained way to measure speech disfluencies in clinical populations on the basis of small samples collected in the home or clinic (however we stress that results are heavily dependent upon text to audio alignment choices which can still be time-consuming). Field linguists and clinicians working in under-served communities can use these measures to document speech patterns in populations who cannot feasibly be reached with articulatory apparatuses such as ultrasounds or electromagnetic articulography. The speed of the measures also evades some of the challenges inherent to articulatory data collection outside of the lab environment or with young children (children are reticent to wear ultrasound stabilization helmets or paste pellets on the tongue for electromagnetic articulography). Thus, while both articulatory and acoustic measures are valid methods to study child speech development, acoustic tools can broaden the communities that are represented in this research.

Future work in this realm could continue to test these coarticulation measures on additional segments to ensure that they capture other coarticulatory patterns such as nasality which were not tested here (Beddor et al., 2018). We also did not compare coarticulatory patterns across adults and children of different ages, which may be an important step towards assuring that the measures capture coarticulation equally in the two populations. However, we stress that comparison of adults and children would likely be inconclusive as the directionality of coarticulatory development is unclear (Barbier et al., 2013; Gerosa et al., 2006; Nitttrouer et al., 1989; Zharkova et al., 2014). It is also important to note that the word repetition employed here could have resulted in phonetic convergence between the children and the adult model speaker, though hopefully the presentation of test items in a random order mitigated any effect. Furthermore, future work explicitly contrasting formant-based measurements with those outlined here is warranted. Finally, though anticipatory coarticulation is generally analyzed in adjacent segments and considered a short-distance phenomenon, long-distance coarticulation may exhibit different developmental patterns (Barbier et al., 2013; Goffman et al., 2008). The current measurements could also efficiently shed light on this outstanding question.

Acknowledgements

The authors thank all of the participating families and Learning to Talk lab members, especially Rebecca Higgins and Michele Liquori, who made this research possible. Research supported by NIDCD grant R01 02932 to Edwards, Munson, and Mary E. Beckman and a UC Berkeley Dissertation Completion Fellowship to Cychosz. Publication made possible in part by support from the Berkeley Research Impact Initiative sponsored by the UC Berkeley Library.

References

- Barbier, G., Perrier, P., Ménard, L., Payan, Y., Tiede, M. K., & Perkell, J. S. (2015). Speech planning in 4-year-old children versus adults: Acoustic and articulatory analyses. *Proceedings of Interspeech 2015*. Presented at the Interspeech 2015, Dresden, Germany.
- Barbier, G., Perrier, P., Ménard, L., Payan, Y., Tiede, M. K., & Perkell, J. S. (2013). Speech planning as an index of speech motor control maturity. *Proceedings of Interspeech 2013*. Presented at the Interspeech, Lyon, France.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.

MEASURING COARTICULATION IN CHILD SPEECH

- Beddor, P. S., Coetzee, A. W., & Styler, W. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *94*(4), 931-968.
- Boersma, P., & Weenik, D. (2018). Praat: doing phonetics by computer (Version 6.0.42). Retrieved from www.praat.org
- Bradlow, A. R. (2002). Confluent talker- and listener-oriented forces in clear speech production. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology* (Vol. 7, pp. 241–273). Berlin and New York: Mouton de Gruyter.
- Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*(02), 201–251.
- Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories User's Guide and Technical Manual* (2nd Edition). San Diego, CA: Singular.
- Gerosa, M., Lee, S., Giuliani, D., & Narayanan, S. (2006). Analyzing Children's Speech: An Acoustic Study of Consonants and Consonant-Vowel Transition. *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, *1*, 393–396.
- Goffman, L., Smith, A., Heisler, L., & Ho, M. (2008). The Breadth of Coarticulatory Units in Children and Adults. *Journal of Speech Language and Hearing Research*, *51*(6), 1424.
- Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The Physiologic Development of Speech Motor Control: Lip and Jaw Coordination. *Journal of Speech Language and Hearing Research*, *43*(1), 239–255.
- Hillenbrand, J., Getty, L. A., Wheeler, K., & Clark, M. J. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111.
- Hoole, P., Nguyen-Trong, N., & Hardcastle, W. (1993). A Comparative Investigation of Coarticulation in Fricatives: Electropalatographic, Electromagnetic, and Acoustic Data. *Language and Speech*, *36*(2–3), 235–260.

MEASURING COARTICULATION IN CHILD SPEECH

- Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38, 245–294.
- Lehiste, I., & Shockey, L. (1972). On the perception of coarticulation effects in English VCV syllables. *Journal of Speech and Hearing Research*, 15(3), 500–506.
- Mahr, T., & Edwards, J. (2018). Using language input and lexical processing to predict vocabulary size. *Developmental Science*, 21(6), e12685.
- Mann, V., & Repp, B. (1980). Influence of vocalic context on perception of the [sh]-[s] distinction. *Perception & Psychophysics*, 28, 213–228.
- Matthies, M., Perrier, P., Perkell, J., & Zandipour, M. (2001). Variation in Anticipatory Coarticulation With Changes in Clarity and Rate. *Journal of Speech Language and Hearing Research*, 44(2), 340–353.
- McAllister Byun, T., Harel, D., Halpin, P. F., Szeredi, D. (2016). Deriving gradient measures of child speech from crowdsourced ratings. *Journal of Communication Disorders*, 64, 91–102.
- McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., & Nieto, O. (2015). librosa: Audio and Music Signal Analysis in Python. *Proceedings of the 14th Python in Science Conference*, 18–24.
- Mooshammer, C., Hoole, P., & Geumann, A. (2006). Interarticulator cohesion within coronal consonant production. *Journal of the Acoustical Society of America*, 120(2), 1028–1039.
- Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of Acquisition Norms for a Large Set of Object Names and Their Relation to Adult Estimates and Other Variables. *The Quarterly Journal of Experimental Psychology Section A*, 50(3), 528–559.
- Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech Language and Hearing Research*, 32, 120–132.
- Noiray, A., Ménard, L., & Iskarous, K. (2013). The development of motor synergies in children: Ultrasound and acoustic measurements. *Journal of the Acoustical Society of America*, 133(1), 444–452.
- Noiray, A., Wieling, M., Abakarova, D., Rubertus, E., & Tiede, M. (to appear). Back from the future: non-linear anticipation in adults and children's speech. *Journal of Speech, Language, and Hearing Research*.

MEASURING COARTICULATION IN CHILD SPEECH

- Öhman, S. E. G. (1966). Coarticulation in VCV Utterances: Spectrographic Measurements. *The Journal of the Acoustical Society of America*, 39(1), 151–168.
- Ramsay, G. (2018). Multitaper harmonic analysis of infant vocalizations. *Journal of the Acoustical Society of America*, 144(3), 1767–1768.
- Recasens, D. (1985). Coarticulatory Patterns and Degrees of Coarticulatory Resistance in Catalan CV Sequences. *Language and Speech*, 28(2), 97–114.
- Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *Journal of the Acoustical Society of America*, 125(4), 2288–2298.
- Recasens, D., Pallarès, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America*, 102(1), 544–561.
- Reidy, P. F., Kristensen, K., Winn, M. B., Litovsky, R. Y., & Edwards, J. R. (2017). The Acoustics of Word-Initial Fricatives and Their Effect on Word-Level Intelligibility in Children With Bilateral Cochlear Implants: *Ear and Hearing*, 38(1), 42–56.
- Rubertus, E., & Noiray, A. (2018). On the development of gestural organization: A cross-sectional study of vowel-to-vowel anticipatory coarticulation. (M. Perlman, Ed.) *PLOS ONE*, 13(9), e0203562.
- Rubertus, E., Abakarova, D., Ries, J., & Noiray, A. (2013). Anticipatory V-to-V Coarticulation in German Preschoolers. *Phonetik Und Phonologie Im Deutschsprachigen Raum*, 12, 5. Munch, Germany.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime*. Pittsburgh, PA: Psychology Software Tools, Inc.
- Smith, A., & Goffman, L. (1998). Stability and Patterning of Speech Movement Sequences in Children and Adults. *Journal of Speech Language and Hearing Research*, 41(1), 18.
- Soli, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *The Journal of the Acoustical Society of America*, 70(4), 976–984.
- Vorperian, H. K., & Kent, R. D. (2007). Vowel Acoustic Space Development in Children: A Synthesis of Acoustic and Anatomic Data. *Journal of Speech Language and Hearing Research*, 50(6), 1510–1545.

MEASURING COARTICULATION IN CHILD SPEECH

- Whalen, D. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[S] boundary. *The Journal of the Acoustical Society of America*, 69(1), 275–282.
- Whiteside, S. P., & Hodgson, C. (2000). Speech patterns of children and adults elicited via a picture-naming task: An acoustic study. *Speech Communication*, 32(4), 267–285.
- Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2011). Coarticulation as an Indicator of Speech Motor Control Development in Children: An Ultrasound Study. *Motor Control*, 15(1), 118–140.
- Zharkova, N., Hewlett, N., Hardcastle, W. J., & Lickley, R. J. (2014). Spatial and Temporal Lingual Coarticulation and Motor Control in Preadolescents. *Journal of Speech Language and Hearing Research*, 57(2), 374-388.

MEASURING COARTICULATION IN CHILD SPEECH

Table 1. Stimuli used in validation experiments.

Word	Transcription	CV sequence	Hypothesis	# of children who correctly produced
sandwich	[sændwɪʃ]	[sæ]	Place of articulation	N=73 (70.87%)
suitcase	[sutkes]	[su]	Place of articulation	74 (71.84)
sister	[sɪstə]	[sɪ]	Manner of articulation	86 (83.50)
chicken	[tʃɪkən]	[tʃɪ]	Manner of articulation	74 (71.84)
window	[wɪndo]	[wɪ]	Manner of articulation	89 (86.41)

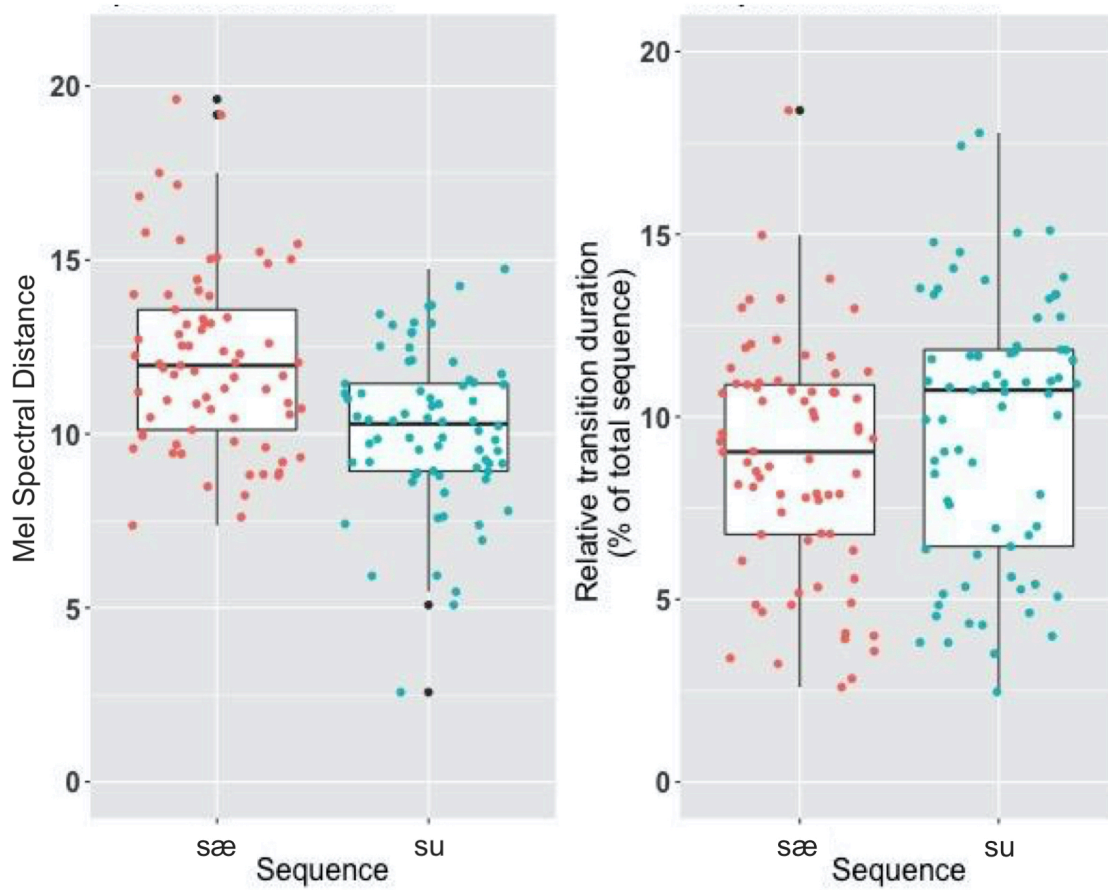


Figure 1. (color online) Fricative-vowel coarticulation by place of vowel articulation. Computed temporally (R) and spectrally (L).

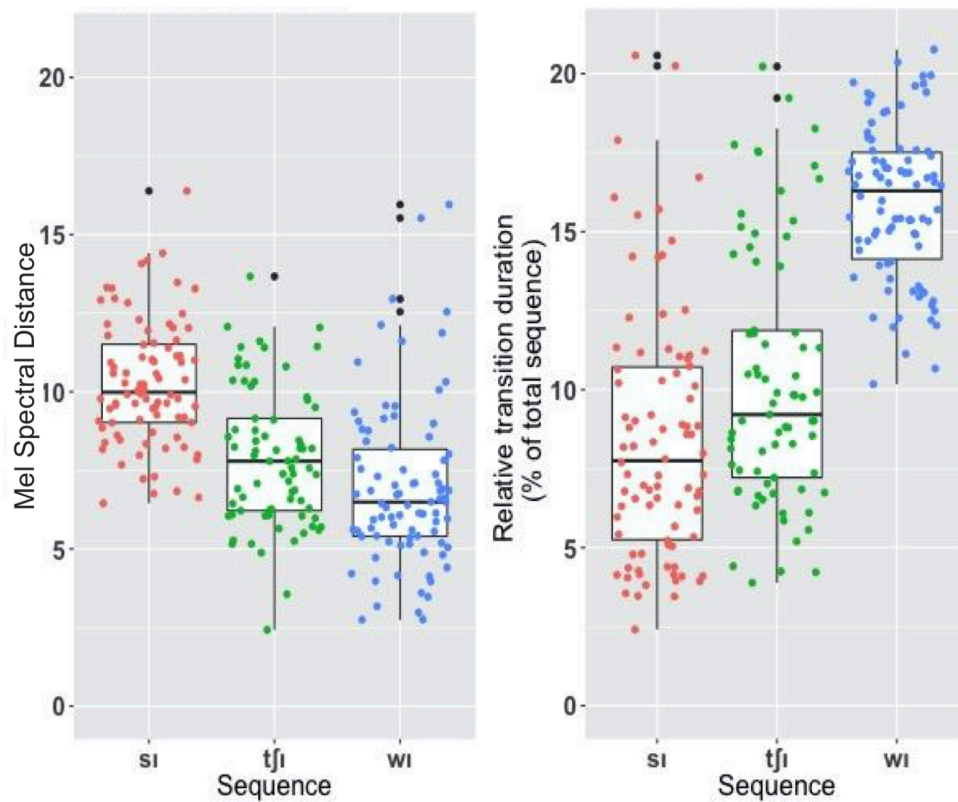


Figure 2. (color online) CV coarticulation by consonant manner. Computed temporally (R) and spectrally (L).