# Spectral and temporal measures of coarticulation in child speech

Margaret Cychosz, Jan R. Edwards, Benjamin Munson, and Keith Johnson

Citation: The Journal of the Acoustical Society of America **146**, EL516 (2019); doi: 10.1121/1.5139201
View online: https://doi.org/10.1121/1.5139201
View Table of Contents: https://asa.scitation.org/toc/jas/146/6
Published by the Acoustical Society of America



# Spectral and temporal measures of coarticulation in child speech

**Margaret Cychosz**[a)]
*Department of Linguistics, University of California, Berkeley, Berkeley, California 94720, USA*
*mcychosz@berkeley.edu*

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

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

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

**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 4-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.


[MG]
**Date Received:** August 20, 2019 **Date Accepted:** November 11, 2019

## 1. Introduction

Coarticulation reflects a crucial equilibrium between speaker efficiency and listener comprehension. For child language development, appropriate coarticulatory overlap indicates mature, adult-like speech. Consequently, coarticulation is a metric for development of speech production and planning (Gerosa *et al.*, 2006; Nittrouer *et al.*, 1989). Because child speech is characterized by underdeveloped motor schemata (Green *et al.*, 2000), it may follow that children would also have immature coarticulatory patterns. However, 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). Still, the question of whether children coarticulate more or less than adults remains unanswered (Barbier *et al.*, 2013; Noiray *et al.*, 2013).

To measure child coarticulation, coarticulation must be quantified using valid, replicable, and, ideally, automated acoustic measures. However, from infancy into puberty, the child speech apparatus creates multiple issues for the study of acoustic phonetics and spectral analyses (Vorperian and Kent, 2007). Small vocal folds result in widely spaced harmonics in the spectral envelope. This can render an undersampled spectral shape obfuscating 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 is to make arbitrary data cleaning decisions, painstakingly hand-check individual peaks in spectral slices, or rely on data points where formant measurements could reliably be found (Nittrouer *et al.*, 1989). Hand-checking may be unrealistic for studies with large sample sizes or if the formant peaks are not visible.

[a)]Author to whom correspondence should be addressed.

 

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

The applicability of traditional acoustic measures of coarticulation, such as formant transitions (Lehiste and Shockey, 1972; Öhman, 1966) or Peak equivalent rectangular bandwidth ($ERB_N$) (Reidy *et al.*, 2017), may be limited to speakers with longer vocal tracts or to certain segments such as fricatives. However, the measures in Gerosa *et al.* (2006) rely on a cepstral representation of the audio signal, as the frequency scale of fast Fourier transformed (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. These measures should be 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).

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

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

where $d_{sa}$ is the Euclidean distance between segments /s/ and /a/ in the biphone sequence /sa/, 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 the temporal coarticulation measure of Gerosa *et al.* (2006). This measure reflects the duration of the transition between adjacent phones. 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 as shown in Eq. (2),

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

where $\bar{x}_s$ is the average Mel spectral vector for /s/, and $\bar{x}_a$ 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 $t = 10$ 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 $\text{dur}_{sa}$ to compute the *relative* transition duration between phones in the CV sequence as shown in Eq. (3),

$$\frac{n \cdot t}{\text{dur}_{sa}}. \tag{3}$$

### 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: fricatives articulated at or behind the alveolar ridge consistently demonstrate anticipatory coarticulation effects that vary by vowel (Mann and Repp, 1980; Soli, 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.

We 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 and Espinosa, 2009). 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 a 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 (Recasens and 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 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 4-year-old children's speech recordings.

### 2.3 The corpus

Data come from 103 four-year-old children (56 girls, 47 boys; range = 3;3–4;4 [years;-months], mean = 3;5, standard deviation (SD) = 0;3). All children were monolingual speakers of English. Children were participating in a longitudinal study of lexical and phonological development. We report on data collected at the second of three time points. Each child passed a hearing screening in at least one ear at 25 dB for 1000, 2000, and 4000 Hz. 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 did not differ significantly from the remaining children. Consequently, data from all children were used.

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 CV sequence in word-initial position and were bisyllabic with penultimate stress. Words were chosen from the MacArthur Bates Communicative Development Inventory (Fenson *et al.*, 2007), the Peabody Picture Vocabulary Test-4 (Dunn and Dunn, 2007), and other sources (e.g., Morrison *et al.*, 1997).

Here we analyze a subset of five of the original test items (Table 1). *Sandwich* and *suitcase* evaluate the place of articulation hypothesis by measuring the anticipatory coarticulation of [s] before [ae] versus [u]. *Sister*, *chicken*, and *window* test manner of articulation by measuring the coarticulation between CV segments where the manner of consonant articulation varies. 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 (Marantz, Kanagawa, Japan). 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 15 min.

### *2.4 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 ICC between raters was 0.881, which was significantly greater than chance [$F(374,375) = 15.9$, $p < 0.001$, 95% confidence intervals (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 and 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. Coarticulation measures are highly dependent upon segmentation decisions. We took a number of steps to standardize alignment. 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. 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. 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.

Table 1. Stimuli used in validation experiments.

| Word | Transcription | CV sequence | Hypothesis | # of children who correctly produced |
|---|---|---|---|---|
| sandwich | [sændwɪtʃ] | [sæ] | Place of articulation | $N = 73$ (70.87%) |
| suitcase | [sutkes] | [su] | Place of articulation | 74 (71.84) |
| sister | [sɪstɚ] | [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) |

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: $r = 0.96$ $p < 0.001$, 95% CI = [0.95, 0.96] and vowels: $r = 0.87$ $p < 0.001$, 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.

## 3. 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 spectral. 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] ($\beta = -1.56$, $t = -3.31$, $p = 0.002$), indicating greater coarticulation between [s] and [u] than [s] and [æ] (Fig. 1). In the temporal model, the transition duration between [s] and [u] is longer than [s] and [æ] ($\beta = 1.08$, $t = 1.99$, $p = 0.05$), 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.

Next, we evaluate the hypothesis that the coarticulatory measures should predict coarticulatory differences by consonant manner in CV sequences. 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] ($\beta = -2.67$, $t = -4.74$, $p < 0.001$) and [wɪ] ($\beta = -3.19$, $t = -5.93$, $p < 0.001$)—[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] ($p = 0.78$). 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] (Fig. 2). For the temporal model, [sɪ] reliably differed from [tʃɪ] ($\beta = 1.98$, $t = 3.42$, $p < 0.001$) and [wɪ] ($\beta = 7.71$, $t = 14.04$, $p < 0.001$). Another *post hoc* test also demonstrated that along the temporal dimension, [tʃɪ] differed significantly from [wɪ] ($\beta = 5.56$, $t = 5.71$, $p < 0.001$). 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.

## 4. Discussion and conclusion

In this study, we used two relatively novel acoustic measures of coarticulation to replicate previous acoustic correlates of coarticulation. We demonstrated that both of the 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 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 toward 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 longer transition duration between phones, in the sequence [su] than [sæ], replicating known coarticulatory patterns by place of vowel articulation and quality (Mann and Repp, 1980; Soli, 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 hitting a vowel target at the beginning of a vowel-consonant sequence, then their transition to the following consonant could be faster than a speaker who did not start at the same halfway

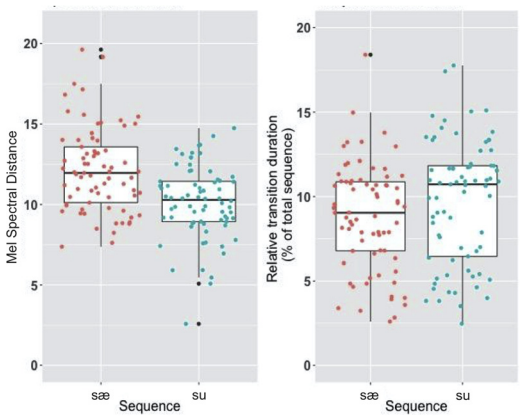


Fig. 1. (Color online) Fricative-vowel coarticulation by place of vowel articulation. Computed temporally (R) and spectrally (L).

point. Yet acoustic measures might say that these speakers "coarticulated" in different amounts, without acknowledging the underlying reasons.

Next, we attempted to capture differences in coarticulation by manner articulation. Consonants whose manner requires less lingual contact, particularly when realized with the tongue blade, are able to resist coarticulation with adjacent segments more than consonants whose manner requires more lingual contact with the sluggish dorsum (Recasens and Espinosa, 2009). We replicated these patterns 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. Furthermore, the measures are relatively immune to the many challenges that children's voices, breathy with high fundamental frequencies, bring to traditional acoustic analyses. Finally, these measurements can be made automatically, over small samples of speech, without specialized equipment. As a result, these measures may have broad applications for clinical populations or understudied groups. The measures can be used as an index of speech maturity or a fine-grained way to measure speech disfluencies in clinical populations on the basis of small samples collected in the home or clinic. 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. The speed of the measures also evades some of the challenges inherent to articulatory data collection outside of the lab or with children (children are reticent to wear ultrasound stabilization helmets or paste pellets on the tongue for electromagnetic articulatography).

Future work could continue to test these coarticulation measures on additional segments to ensure that they capture other coarticulatory patterns such as nasality. We also did not compare coarticulatory patterns across adults and children of different ages, which may be an important step toward assuring that the measures capture coarticulation

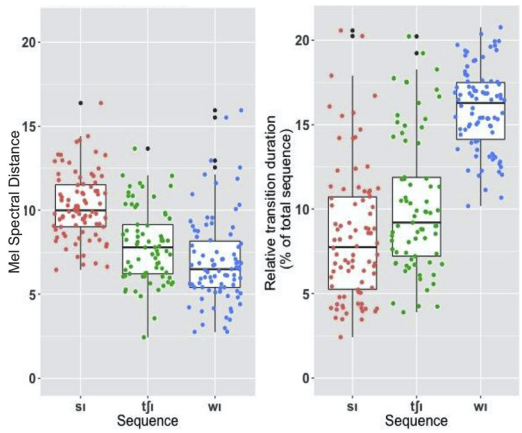


Fig. 2. (Color online) CV coarticulation by consonant manner. Computed temporally (R) and spectrally (L).

equally in the two populations. However, we stress that a 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; Nittrouer *et al.*, 1989). It is also important to note that the word repetition employed here could have resulted in phonetic convergence between the children and the model speaker, though hopefully the presentation of test items in a random order mitigated any effect. Future work explicitly contrasting formant-based measurements with those outlined here is warranted.

## Acknowledgments

The authors thank the participating families and Learning to Talk lab members, especially Rebecca Higgins and Michele Liquori. Research was supported by NIDCD Grant No. R01 02932 to J.R.E., B.M., and Mary E. Beckman and a U.C. Berkeley Dissertation Completion Fellowship to M.C.

## References and links

Barbier, G., Perrier, P., Ménard, L., Payan, Y., Tiede, M. K., and Perkell, J. S. (**2013**). "Speech planning as an index of speech motor control maturity," in *Proceedings of Interspeech 2013*, Lyon, France.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Software **67**(1), 1–48.

Boersma, P., and Weenik, D. (**2018**). Praat: Doing phonetics by computer (Version 6.0.42). Retrieved from www.praat.org (Last viewed March 15, 2018).

Dunn, L. M., and 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., and Bates, E. (**2007**). *MacArthur-Bates Communicative Development Inventories User's Guide and Technical Manual*, 2nd ed. (Singular, San Diego, CA).

Gerosa, M., Lee, S., Giuliani, D., and Narayanan, S. (**2006**). "Analyzing children's speech: An acoustic study of consonants and consonant-vowel transition," in *2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings*, Vol. 1, pp. 393–396.

Green, J. R., Moore, C. A., Higashikawa, M., and Steeve, R. W. (**2000**). "The physiologic development of speech motor control: Lip and jaw coordination," J. Speech Lang. Hear. Res. **43**(1), 239–255.

Lehiste, I., and Shockey, L. (**1972**). "On the perception of coarticulation effects in English VCV syllables," J. Speech Hear. Res. **15**(3), 500–506.

Mann, V., and Repp, B. (**1980**). "Influence of vocalic context on perception of the [sh]-[s] distinction," Percept. Psychophys. **28**, 213–228.

McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., and Nieto, O. (**2015**). "librosa: Audio and music signal analysis in python," in *Proceedings of the 14th Python in Science Conference*, pp. 18–24.

Morrison, C. M., Chappell, T. D., and Ellis, A. W. (**1997**). "Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables," Qtrly. J. Exp. Psychol. A **50**(3), 528–559.

Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (**1989**). "The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults," J. Speech Lang. Hear. Res. **32**, 120–132.

Noiray, A., Ménard, L., and Iskarous, K. (**2013**). "The development of motor synergies in children: Ultrasound and acoustic measurements," J. Acoust. Soc. Am. **133**(1), 444–452.

Öhman, S. E. G. (**1966**). "Coarticulation in VCV utterances: Spectrographic measurements," J. Acoust. Soc. Am. **39**(1), 151–168.

Recasens, D. (**1985**). "Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences," Lang. Speech **28**(2), 97–114.

Recasens, D., and Espinosa, A. (**2009**). "An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan," J. Acoust. Soc. Am. **125**(4), 2288–2298.

Reidy, P. F., Kristensen, K., Winn, M. B., Litovsky, R. Y., and Edwards, J. R. (**2017**). "The acoustics of word-initial fricatives and their effect on word-level intelligibility in children with bilateral cochlear implants," Ear Hear. **38**(1), 42–56.

Rubertus, E., Abakarova, D., Ries, J., and Noiray, A. (**2013**). "Anticipatory V-to-V coarticulation in German preschoolers," in *Phonetik Und Phonologie Im Deutschsprachigen Raum* [*Phonetics and Phonology in German-speaking Countries*], Munich, Germany, Vol. 12, p. 5.

Schneider, W., Eschman, A., and Zuccolotto, A. (**2012**). *E-Prime* (Psychology Software Tools, Inc., Pittsburgh, PA).

Soli, S. D. (**1981**). "Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation," J. Acoust. Soc. Am. **70**(4), 976–984.

Vorperian, H. K., and Kent, R. D. (**2007**). "Vowel acoustic space development in children: A synthesis of acoustic and anatomic data," J. Speech Lang. Hear. Res. **50**(6), 1510–1545.

Zharkova, N., Hewlett, N., and Hardcastle, W. J. (**2011**). "Coarticulation as an indicator of speech motor control development in children: An ultrasound study," Motor Control **15**(1), 118–140.