# Acoustic correlates and listener ratings of function word reduction in child versus adult speech

Melissa A. Redford

The University of Oregon

and

Phil J. Howson

Leibniz-Zentrum Allgemeine Sprachwissenschaft

**Running title**:

Acoustic correlates and listener ratings of reduction

**Please address correspondence to**:

Melissa A. Redford, Professor; Department of Linguistics | 1290 University of Oregon | Eugene, OR 97403; redford@uoregon.edu.

**Abstract**

The present study investigated "the" reduction in phrase-medial Verb-*the*-Noun sequences elicited from five-year-old children and college-aged adults. Several measures of reduction were calculated based on acoustic measurement of these sequences. Analyses on the measures indicated that the determiner vowel was reduced in both child and adult speech relative to content word vowels, but it was reduced less in children's speech compared to adults' speech. Listener ratings on the sequences indicated a preference for adult speech over children's speech. Acoustic measures of reduction also predicted goodness ratings. Listeners preferred sequences with shorter and lower amplitude determiner vowels relative to content word vowels. They also preferred a more neutral schwa over more coarticulated versions. In sequences where ratings differed by age group, the effect of coarticulation was limited to adult speech and the effect of relative schwa duration was limited to children's speech. The results are discussed with reference to communicative pressures on function word production in speech, including the rhythmic and semantic pressure towards reduction versus the pressure to convey adequate information in the acoustic signal. It is argued that competing pressures on production may delay the acquisition of adult-like function word reduction.

*Keywords*: speech production; speech acquisition; speech rhythm; speech motor skills

## I. INTRODUCTION

Function words, like determiners, refine the message and help to define the grammatical structure of a sentence. As high frequency items with minimal semantic weight, function words are typically unstressed and phonetically reduced relative to lower frequency content words with maximal semantic weight (Bell et al., 2009). The 10 most frequent function words in English are monosyllabic (Jurafsky et al., 1998). These alternate with lexically-stressed content words, many of which are also monosyllabic (Cutler and Carter, 1987). In spoken language, the recurrent alternation of unstressed function words with stressed content words, though not itself periodic, contributes substantially to the rhythm of English (Allen and Hawkins, 1978; Dauer, 1983; Deterding, 2001). Given the importance of lexical stress and rhythm to speech processing (e.g., Cutler and Butterfield, 1992; Mattys et al., 2005; Dilley and McAuley, 2008), it is reasonable to think that listeners come to expect function word reduction along the temporal and amplitude dimensions that define speech rhythm (Grabe and Low, 2002; He, 2012; Tilsen and Arvaniti, 2013). But reduction along these dimensions is also typically associated with greater coarticulation (i.e., greater gestural "overlap" or hypo-articulated speech; Agwuele et al., 2008; Moon and Lindblom, 1994). When coarticulation is extreme, vowel quality may be impacted to the point of distorting the phonetic shape of the determiner vowel, rendering it more difficult to process. And, of course, reduction-related decreases along the temporal and amplitude dimensions can also render function words inaudible to the listener (see, e.g., Dilley and Pitt, 2010). These observations suggest that the rhythmic requirements of speech production may compete with the functional pressure to produce intelligible speech. This competition may complicate the acquisition of function word reduction during spoken language development. It is an interest in children's acquisition of function word reduction that motivates the present study, which investigates the relationship between the acoustic correlates of "the" reduction and adult listener ratings of speech rhythmicity.

### A. The developmental context

In adult speech, reduced unstressed syllables are shorter, quieter, and more coarticulated with adjacent speech sounds than unreduced stressed syllables (Dauer, 1983; Fourakis, 1991; Fowler, 1981;

Plag et al., 2011). This is especially true for monosyllabic function words, which are typically more reduced than unstressed syllables in content words (Fuchs, 2016; van Bergem, 1993). Although children acquire the overall temporal and amplitude patterns associated with lexical stress by age 2 or 3 years (Ballard et al., 2012; Kehoe et al.,1995; Schwartz et al., 1996), they elide unstressed function words in extra-metrical positions in early speech (see, e.g., Gerken, 1996) and do not reduce these in running speech to the same extent as adults until sometime in middle childhood (Allen and Hawkins, 1978; Nittrouer, 1993; Goffman, 2004; Redford, 2018). For example, Nittrouer (1993) found that the vowel of the indefinite determiner "a" was longer in speech produced by 3-, 5-, and 7-year-old children than in adult speech, but that their stressed vowel durations were similar to adults' stressed vowel durations. Redford (2018) found that, in comparison to adults, 5- and 8-year-old children produced longer vowels in the definite determiner "the" relative to adjacent content word vowels. In addition, she found that the determiner vowel was louder relative to the content word vowel in children's speech compared to adults' speech, but that determiner vowel formant frequencies varied as a function of the following content word onset in both child and adult speech. One aim of the present study is to confirm these limited findings by investigating "the" reduction in speech elicited from different English-speaking 5-year-old children and college-aged adults.

Function word reduction, or lack thereof, likely impacts speech rhythm. Interval-based studies of rhythm indicate that English-speaking children's speech is both more vocalic overall than adults' speech until late middle childhood and also that the vocalic intervals in children's speech vary less in duration than in adults' speech (Payne et al., 2011; Polyanskaya and Ordin, 2015). Whereas the greater vocalicness of children's speech might be explained by age-related differences in articulation rate (see below), reduced variability in vowel durations suggests a pattern that speech-language pathologists might characterize as "excessive equal stress." Such a pattern impedes speech intelligibility (e.g., Shriberg et al., 2003). Since the relative duration and amplitude patterns typical of lexical stress is mastered early (Ballard et al., 2012), the reduced variability of older children's speech requires another explanation. Hawkins and Allen (1978) long ago suggested that the explanation might be incomplete function word

reduction. Sirsa and Redford (2011) provided some support for this explanation when they found that interval-based measures of school-aged children's speech rhythm were better predicted by a ratio of determiner vowel duration to a subsequent noun vowel duration (an acoustic measure of function word reduction) than by a ratio of unstressed vowel duration to stressed vowel duration within a disyllabic word (an acoustic measure of lexical stress patterning) or by a ratio of final vowel duration to the mean of non-final vowel durations (an acoustic measure of final lengthening).

Of course, school-aged children's speech differs from adults' speech along a variety of dimensions that intersect with function word reduction and, by extension, with speech rhythm. The most obvious of these is articulation rate, which is slower in children's speech than in adults' speech until at least age 12 years (Lee et al., 1999). Whereas rate changes in adults' speech are associated with targeted changes in stressed vowel production (Gay, 1981), children's distinctively slower articulation rate is due to slower articulatory movements into and out of all segmental targets, including unstressed vowels (Redford, 2014). These slower movements are part of an overall pattern of articulation that is associated with immature motor skills, including larger amplitude articulatory movements relative to oral-facial size (Riely and Smith, 2003) and greater spatial-temporal variability (Smith and Zelaznik, 2004). Since these patterns would seem to conspire against the production of very short, quiet, coarticulated unstressed vowels, it is likely that age-related differences in function word reduction are due to immature speech motor skills, not to immature prosodic representations. This conclusion is consistent with results from kinematic and acoustic studies on the production of different metrical structures in child versus adult speech (e.g., Goffman, 2004; Redford, 2018): both children and adults produce trochaic and iambic patterns; the patterns produced by children are simply less contrastive than those produced by adults. It is also consistent with the evidence from cross-linguistic studies on the acquisition of prosody, which suggests the acquisition of language-specific rhythmic structures and intonational patterns by age 3 years (see Fikkert, Liu, and Ota, 2021) – long before the adult-like production of these patterns.

But even if the acquisition of adult-like function word reduction is "merely" a product of speech motor development, it must still be learned. This learning begins with patterns extracted from the ambient

language. And findings from studies of early child language strongly suggest that this happens early. In particular, developmental studies of speech processing indicate that English-speaking toddlers make use of function words to identify noun-picture correspondences (Kedar et al., 2006); soon thereafter, they regularly produce these words in syntactically correct sentences (~ age 3 years; Abu-Akel et al., 2004). The question is: Once children use function words correctly, how do they know that their speech is still not phonetically accurate? Relatedly, what motivates children to adjust their production of function words across developmental time until they achieve adult-like patterns of reduction? If the answer to the second question is that children strive to communicate with others (adults included), then the answer to the first is that they are not always successful in doing so. Specifically, both the elision of function words and their too-fulsome production may impede communication because it disrupts speech rhythm. Communication failure motivates the child to try again, leading to the adjustments that characterize speech motor learning and the acquisition of adult-like speech patterns.

**B. The present study**

The hypothesis that competing pressures on function word production delays the acquisition of adult-like reduction predicts that children do not reduce function words to the same extent as adults. The hypothesis that communicative pressures help shape children's speech motor learning, including learning that underpins the acquisition of function word reduction, predicts that adults prefer adult-like speech rhythm patterns over children's speech rhythm patterns. This prediction is in line with the evidence that adult listeners prefer more intelligible children's speech than less intelligible children's speech (e.g., Redford et al., 2018). In the context of the current study, the more specific prediction is that adult listeners will prefer adult-like reduction of function words over incomplete reduction of these words. These predictions motivate the current study. Here, we sought to identify the acoustic correlates that best distinguished child from adult productions of a determiner, and then asked whether these same correlates account for adult listener ratings of speech rhythmicity on sequences that contained the determiner.

## II. EXPERIMENT 1

The aims of Experiment 1 were to confirm previous study findings on age-dependent differences in function word reduction and to identify those acoustic correlates of reduction that are most salient to adult listeners and so are most likely to influence the predicted preference for adult speech rhythm over children's speech rhythm. Simple Subject-Verb-*the*-Noun sentences were elicited from a group of 5-year-old children and college-aged adults. Both the verb and the object noun were monosyllabic. The sentence frame included a final "today" in order to avoid phrase-final lengthening effects on the object noun. The consonantal context on either side of "the" was controlled. The stressed vowels in the verb and noun were manipulated to investigate vowel-to-vowel coarticulatory effects on the unstressed determiner vowel. Several measures of reduction were calculated from the acoustic data and analyzed for an effect of age. These were the duration of schwa divided by the duration of adjacent content word vowels (i.e., relative duration), the amplitude of schwa divided by the amplitude of adjacent content word vowels (i.e., relative amplitude), and the effect of vowel context on schwa formant frequencies (i.e., coarticulation). Based on previous findings, the predictions were that children would produce relatively longer and higher amplitude determiner vowels compared to adults, but that the unstressed vowels would be similarly coarticulated with adjacent vowels in children's and adults' speech.

### A. Methods

#### *1. Participants*

Participants were 12 school-aged children (7 female) and 12 college-aged adults (6 female) drawn from a larger project on speech rhythm acquisition. Children ranged in age from 5;3 to 6;2 with a mean age of 5;8 (SD = 3 months). They were recruited via a database built and maintained by a group of developmental labs at the University of Oregon and from summer camps run by the YMCA in Eugene, Oregon. Typical development in children was determined based on parental report and on an in-laboratory assessment of speech-language skills using the Diagnostic Evaluation of Articulation and Phonology (DEAP; Dodd et al., 2002) and the Clinical Evaluation of Language Fundamentals (CELF-5; Wiig et al., 2013). Inclusion criteria were standardized scores within 1 standard deviation of the mean on

both the DEAP and the CELF-5. Further selection criteria were based on age (the larger sample included 8-year-olds), the order in which children participated in the study (recruitment was on-going), and the quality of the video recording (not relevant to the present study). The college-aged adults were recruited by word-of-mouth from the University of Oregon student body. None of the adult participants reported a history of speech-language therapy. All participants, including adults, completed and passed a pure-tone hearing screen (1000, 2000, and 4000 Hz at 25 dB). All participants were financially compensated for their time. Children also earned a small prize at the end of the study session.

***2. Speech Elicitation***

Speech materials were designed to investigate carryover and anticipatory V-to-V effects on the production of "the" in simple Subject-Verb-Object sentences, where the verb and object noun were varied to create different stressed vowel environments for the determiner. A total of 16 sentences were created. Here, we focus on a subset of 4 sentences that were designed to investigate the effect of vowel height and backness on schwa production using the vowels [ɑ] and [i]. The verbs were "shock" and "tweak" and the nouns were "god" and "geek." The target sentences, which had a first person singular subject and the object noun in penultimate position, were as follows: *I shock the god today; I shock the geek today; I tweak the god today; I tweak the geek today*. An adult female speaker of west coast American English recorded the target sentences with a second person singular subject (e.g., *You shock the geek today*), followed by the question: *What do you do today?* Care was taken to produce the model sentence under a single intonational contour, followed by a clear prosodic break before the question.

Participants were introduced to the object nouns with different cartoon pictures; namely, a studious looking boy for *geek* and an unfamiliar mythological deity for *god*. Verbs were associated with different hand gestures (a flicking gesture for *tweak* and a whole-hand expansion gesture for *shock*), which the experimenter deployed next to the cartoon picture during elicitation when each stimulus sentence was played. The participant's task was to repeat back the model sentence in first person after the question prompt. The experimenter controlled the pace of sentence elicitation and provided feedback on production during practice. If a repetition was deemed errorful or disfluent, the experimenter elicited a

new repetition by replaying the stimuli at the end of a block. The sentences were elicited in random order four times, once per repetition block. Speech was audio-visually (AV) recorded with a Panasonic AJ-PX270. Audio for the video was recorded with Shure SM81 Condenser at 44100 Hz.

### *3. Acoustic Segmentation and Measurement*

The first three good repetitions of each sentence from every speaker were selected for measurement, resulting in a total of 288 sentences for analysis. A good repetition was defined as a fluent utterance that the speaker produced while looking face-on at the video camera. The AV recording of each sentence was isolated and saved. Audio was stripped from the files and displayed for vowel segmentation as an oscillogram and a spectrogram in Praat (Boersma and Weenink, 2019). Utterance boundaries were identified by the onset and offset of acoustic energy in the sentence and its duration measured. The verb, determiner, and noun vowel were then segmented and durations measured based on repeated listening, visible periodicity and abrupt changes in the oscillogram and/or the presence of formant structure.

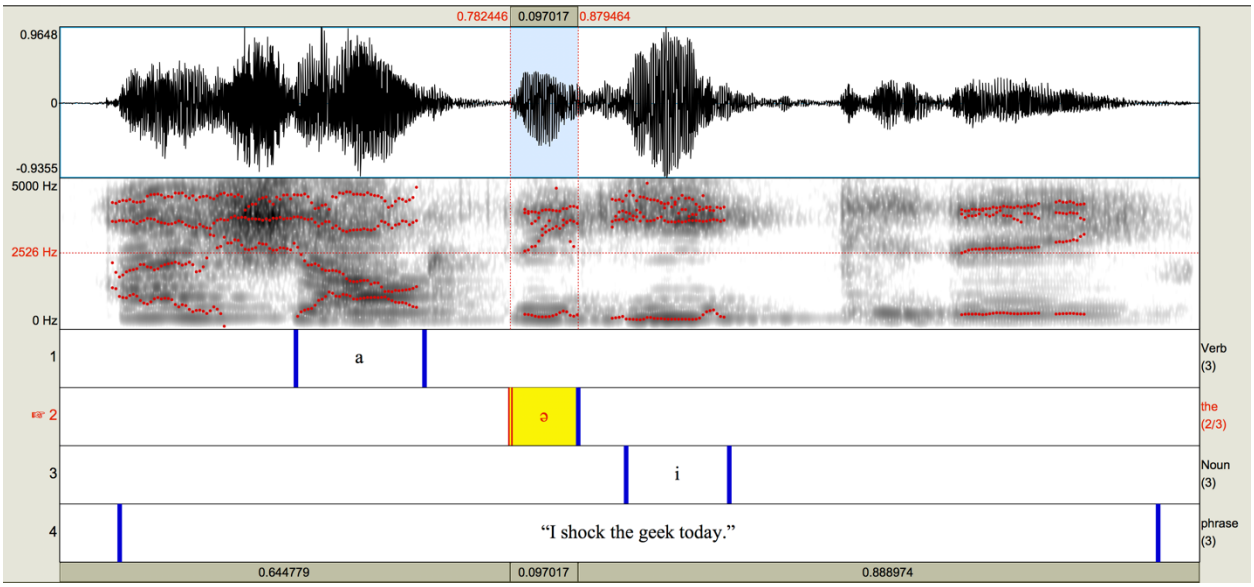


*Figure 1. Example of vowel segmentations for a Verb-the-Noun sequence embedded in a simple S-V-O sentence elicited from a 5-year-old child.*

Figure 1 illustrates the segmentation criteria for the sentence, *I shock the geek today*, produced by a 5-year-old boy. As in Figure 1, stressed vowels in the monosyllabic verb and noun were typically

produced in such a way that the expected visible cues were robustly present. Determiner vowels were also easily identified based on some subset of the cues. The reliability of the authors' segmentation was assessed on 25% of the data. A research assistant, blind to the purpose of the experiment, was asked to segment 3 sentences chosen at random from each speaker according to the above criteria. Interval durations from the new segmentations were correlated with those based on the original segmentations on the same data. The bivariate correlation indicated very high inter-rater reliability, $r(216) = 0.922$.

Amplitude (dB) and formant frequency were measured at 10 evenly-spaced intervals across the entire schwa duration, and at 5 evenly-spaced intervals in the latter half and first half of the verb and noun vowel, respectively. The spectrogram settings were as follows: the view range was set from 0 to 7000 Hz; the window length was 0.005 s; the dynamic range was 50 dB. The standard pre-emphasis view setting of 6 dB per octave was used. Amplitude used the standard Praat settings including a minimum pitch setting of 75 Hz, a time step that was one quarter the spectrogram window length divided by the minimum pitch, and a cubic interpolation method. If the automatic track was visibly inaccurate relative to the spectrogram, tracking was adjusted by changing the number of formants tracked or the range in Hz used for picking peaks. The solution for improving tracking typically differed by vowel type: for example, the number of formants increased for low vowels if F1 and F2 were not separately tracked; the range in Hz increased for high vowels if F2 was poorly tracked.

***4. Analyses***

Articulation rate was calculated for each target sentence (syll/s). Sentences with pauses were excluded from this calculation ($N$=36). Relative schwa duration and amplitude was calculated by dividing the determiner vowel duration/amplitude by the duration/amplitude of the verb (Det:V) or by the duration/amplitude of the noun (Det:N). Lower ratios signaled greater reduction than higher ratio. Although all participants generally produced fluent sentences, a few sentences were produced with a pause between the verb and object noun phrase ($N$=24) or between the object noun phrase and the adverbial ($N$=2). Since a prosodic break is associated with pre-final lengthening, Det:V and Det:N vowel durations were calculated only when the content word was not also at a prosodic boundary. All vowels

from one hyperarticulated sentence, where all elements were uttered under narrow prosodic focus, was also excluded from the analyses.

A linear mixed effects model tested for the fixed effects of age group (Group, 2 levels: child, adult) and vowel context (Context, 4 levels: [ɑ]_[ɑ], [ɑ]_[i], [i]_[ɑ], [i]_[i]) on the temporal and amplitude measures of reduction. The model was built using the lme4 package (Bates, Mächler, Bolker, and Walker, 2015). Speaker and repetition were included as random effects. Repetition was removed when shown to have no significant effect on the results. The lmerTest package (Kuznetsova et al., 2017) was used to estimate the degrees of freedom with Satterthwaite's method (Satterthwaite, 1946). Box and whisker plots present all data used in the analyses. The whiskers represent 1.5 times the interquartile range. The potential outliers (circles) and extreme values (stars) that are shown in the plots were not excluded from the analyses.

Schwa coarticulation was assessed based on formant frequencies in two vowel contexts – the [ɑ]_[i] and [i]_[ɑ], contexts. These contexts were chosen to test both the effect the stressed vowels on determiner production and the direction of this effect. Formant frequencies were analysed using a smoothing spline (SS)ANOVA (Gu, 2014) and an R script (Mielke, 2015; R Core Team, 2019). For each of the first three formants, best-fit frequency trajectories and 95% confidence intervals were calculated. Though not presented here, the first three formants of the stressed vowels were analysed in the same way and found to conform to expectations consistent with the low vowel versus high vowel target.

**B. Results**

As expected, children spoke more slowly than adults: mean articulation rate was 3.22 syllables per second in child speech (*SD* = 0.49 syll) and 4.37 syllables per second in adult speech (*SD* = 0.63 syll). The absolute duration of the determiner vowel was therefore longer in child speech [*M*(144) = 115 ms, *SD* = 26 ms] compared to adult speech (*M*(144) = 74 ms, *SD* = 18 ms). More importantly, the mixed effects models indicated that both children and adults produced a longer, higher amplitude schwa relative to adjacent stressed vowels, though the effect of age group on temporal reduction interacted with the particular Verb-*the*-Noun sequence produced. Specifically, the data in Figure 2 show that schwa duration

was generally longer in child speech after [ɑ] but not after [i]. This was true whether relative duration was measured in relation to the verb or noun. Accordingly, there was a significant interaction between age group and vowel context on both Det:V duration [$F(3,257) = 2.73$, $p = 0.044$] and on Det:N duration [$F(3,257) = 2.88$, $p = 0.037$]. The simple effect of context was also significant on both Det:V duration [$F(3,257) = 12.64$, $p < 0.001$] and Det:N duration [$F(3,257) = 72.26$, $p < .001$]. The simple effect of age group was not significant on either measure.

The main effect of vowel context was also significant on relative schwa amplitude [Det:V amplitude: $F(3,257) = 4.51$, $p = 0.004$; Det:N amplitude: $F(3,257) = 26.28$, $p < 0.001$. But the data in Figure 3 show that children produced a higher amplitude schwa compared to adults when this measure was calculated in relation to the verb [$F(1,22) = 12.28$, $p = 0.002$]. The interaction between age group and vowel context was not significant no matter how relative amplitude was calculated. The effect of age group on Det:V and not on Det:N amplitude could reflect an effect of phrase position on child speech or the more consistent modulation of amplitude across the phrase in adult speech.

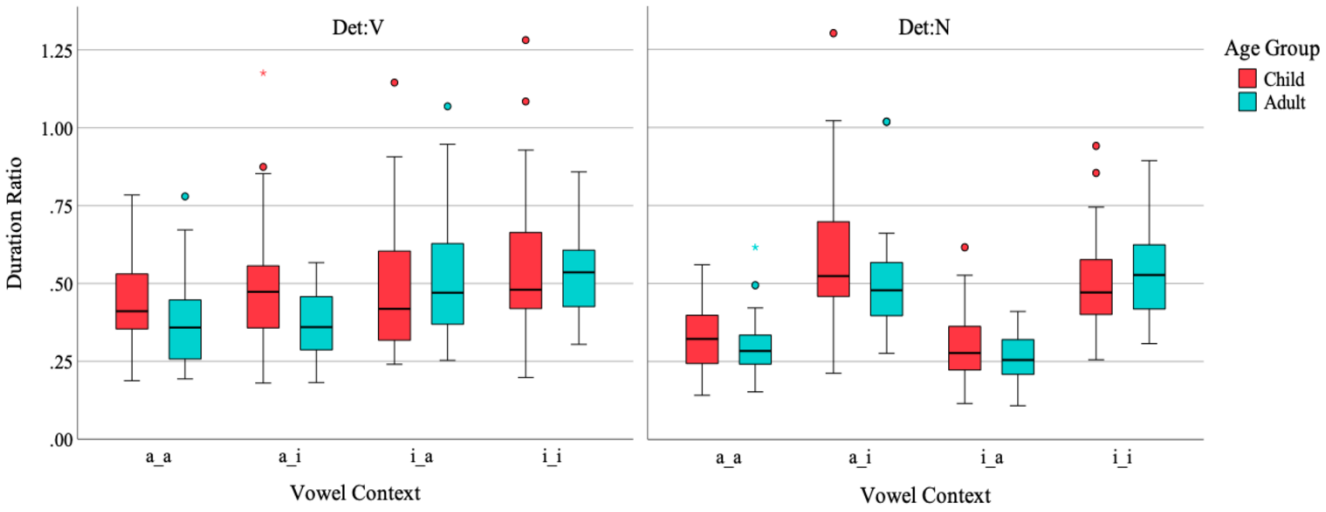


*Figure 2. Vowel duration ratios for Det:V (left) and Det:N (right) shown by age group and the vowel context within which the determiner appeared.*

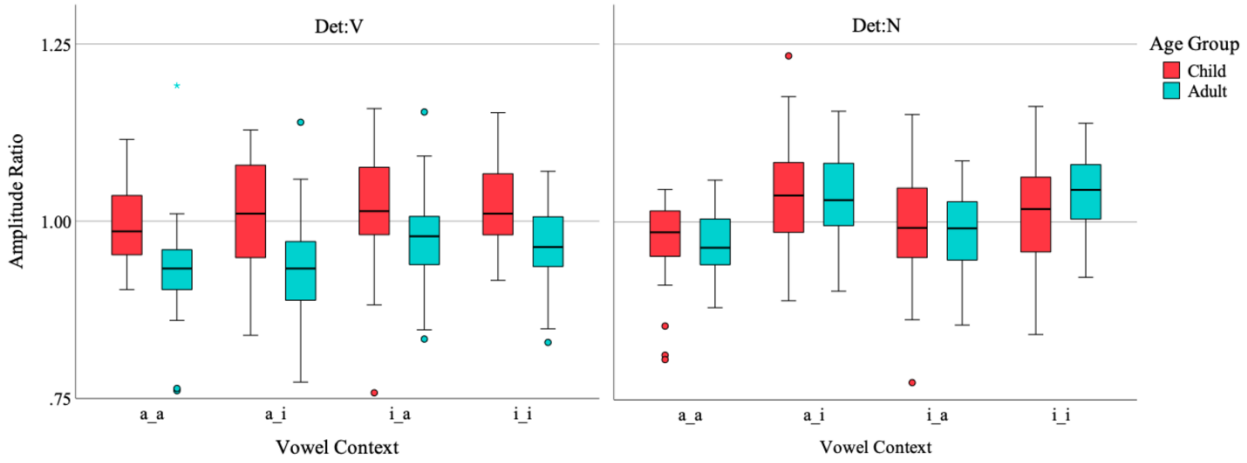


*Figure 3. Vowel amplitude ratios for Det:V (left) and Det:N (right) are shown by age group and the vowel context within which the determiner appeared.*

The effect of age group on function word reduction extended to coarticulation. Figure 4 shows that both children and adults produced schwa differently as a function of the upcoming stressed vowel, but not the preceding one. In child speech, the largest effect of vowel context is seen on F2, which was higher before [i] than before [ɑ]. In adult speech, the data indicate an additional effect of context on F3, which was also higher before [i] compared to [ɑ]. These results suggest that children move the

constriction location forward during schwa ahead of [i]; adults may also front the tongue body more before [i] than before [ɑ], but changes in F3 also suggest greater lip spreading and a higher overall constriction degree (= larger subapical aperture) in this context (Lindblom et al., 2011).

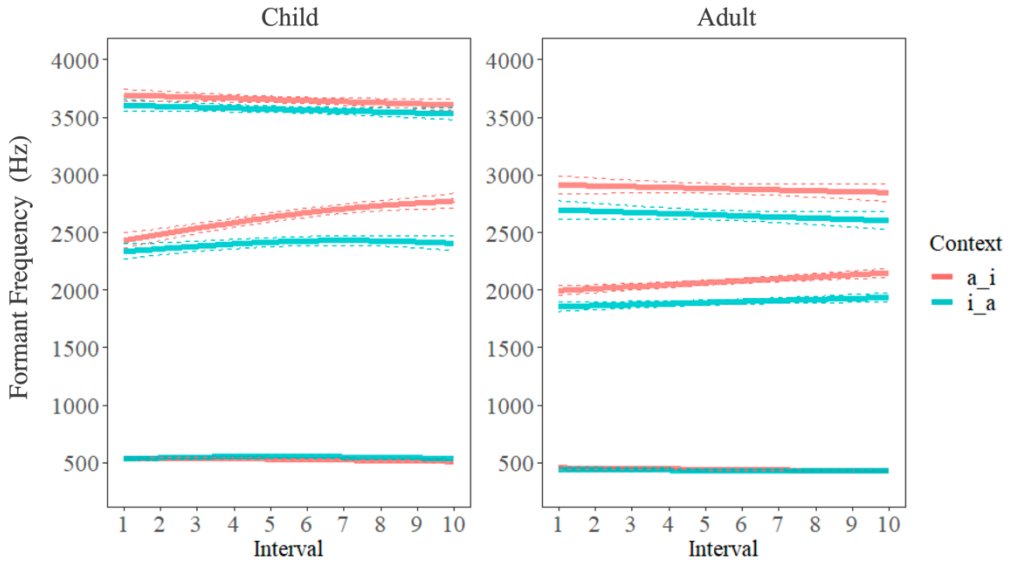


*Figure 4. Best-fit curves for schwa F1, F2, and F3 are shown as a function of the stressed vowel context within which the determiner was produced in child (left) and adult (right) speech. Dotted lines represent the 95% confidence interval.*

**C. Discussion**

The results from Experiment 1 indicate that when group differences are observed, relative schwa duration and amplitude is higher in child speech compared to adult speech. This result replicates previous findings, including from studies where the elicited speech was less well controlled and so also less contrived (e.g., Sirsa and Redford, 2011). The results also indicate that both children and adults coarticulated the determiner vowel with the following noun vowel, but how this was done differed by age. Whereas the results from child speech suggest an adjustment along the front-back dimension (i.e., F2) during schwa articulation when it was produced in advance of [i], the results from adult speech suggest that a single posture was adopted across the entire schwa duration that adjusted for both constriction location as well as the shape of the front cavity (i.e., F3) in advance of [i]. These age-related differences

in vowel-to-vowel coarticulation may be related to children's overall slower articulation rate and longer determiner vowels (see, e.g., Agwuele et al., 2008; Moon and Lindblom, 1994) or to differences in the speech plan representation (see General Discussion). Either way, slower movements into and out of a vowel target cannot explain why the *relative* duration and amplitude of schwa varied with age. Instead, the Det:V and Det:N results are consistent with the interpretation that 5-year-old children do not reduce function words to the same extent as adults, at least along the temporal and amplitude domains most closely associated with speech rhythm.

## III. EXPERIMENT 2

In Experiment 2, we investigate whether age-related differences in function word reduction influence adult listeners' ratings of speech rhythmicity. Verb-*the*-Noun sequences were excised from the subset of sentences with different vowel contexts. The sequences were blocked by speaker and age group and presented to listeners, who were instructed to rate the rhythmic quality of the sequence on a goodness scale. The blocked design was used to encourage listeners to attend to within speaker variability in production rather than to the many acoustic characteristics that distinguish individual speakers from one another and children from adults. Despite this encouragement, we expected listeners to rate children's speech as less good overall than adults' speech – either because children do not reduce function words to the same extent as adults or because listeners make global intelligibility judgments even when instructed to attend to speech rhythm. We also expected that the acoustic correlates of function word reduction would account for listener ratings of rhythmicity independently from a preference for adult speech, assuming that listeners are indeed as sensitive to speech rhythm as the psycholinguistic literature would suggest. We were particularly interested in the character of this sensitivity: Do certain correlates of function word reduction matter more to listeners than others? Do the same correlates that influence ratings of child speech also account for ratings of adult speech? Under the hypothesis that communicative pressures help shape children's speech motor learning, including the learning that underpins the acquisition of function word reduction, we expected that the answer to both questions would be "yes" and that the specifics of this answer would follow from the results of Experiment 1. In particular, we expected

that measures of reduction that systematically varied with age group (especially, Det:V duration and Det:V amplitude; see Experiment 1) would best predict listener ratings of goodness independently from a preference for adult speech over child speech.

## A. Methods

### *1. Participants*

A total of 100 adult listeners participated in Experiment 2. Their mean age was 35.5 years (*SD* = 10.7 years); 81 self-identified as female. Listeners were recruited using Amazon's Mechanical Turk crowdsourcing platform (MTurk; Buhrmester et al., 2011). Recruitment was limited to self-reported native speakers of English residing in either the United States or Canada. It was further restricted to just those MTurkers who had previously completed a minimum of 5000 HITs ("Human Intelligence Tasks") with an acceptance rate of at least 95%. Each listener was compensated for their time upon completing the task.

### *2. Stimuli*

The sentences that were elicited from children and adults in Experiment 1 were used to investigate the effect of age group and determiner vowel reduction on listener preference. Stimuli were those sentences with maximally contrastive vowel contexts (*I shock the geek today* and *I tweak the god today*). The Verb-*the*-Noun sequence from these sentences was excised and amplitude normalized to 70dB. This resulted in 72 sequences for the [ɑ]_[i] and [i]_[ɑ] vowel context (= 1 sentence × 3 elicitations × 24 speakers).

### *3. Procedure*

One group of 50 adult listeners rated the goodness of all [ɑ]_[i] stimuli, another group rated the goodness of all [i]_[ɑ] stimuli. Each stimulus was presented 3 times to the listener for a total of 216 stimuli (= 24 speakers × 3 elicitations × 3 repetitions). Stimuli were blocked by speaker, and speaker was blocked by age group. Stimuli within speaker and speaker within age block were randomized for each listener, as was the order of the age blocks. Listeners were instructed to wear headphones set to a comfortable listening volume based on a preliminary task, which was to listen to and then type 3 different

words. They were then instructed on the main task. The instructions, given in full below, sought to draw listeners' attention to the rhythmic aspects of the sequence and thus away from other features that are specific to child versus adult speech (e.g., pitch):

> *This experiment is broken into 2 parts. You will be prompted at the start of Part 1 to begin, and then again at the start of Part 2. Feel free to take a break between parts if you need one. In both parts of the experiment, you will hear the same 3-word stretch of speech produced by different talkers. The sequence of words has been taken out of a single sentence context. Your task is to rate the rhythmic quality of the sequence. On each trial, a cross will appear on the screen and then disappear. Then you will hear audio play. After the audio plays, a scale will appear with numbers from 1 to 7. When the scale appears, rate the rhythmic quality of the 3-word sequence from 1 to 7: 1 = Sounds Weird. 7 = Sounds Great. Please press the keyboard button corresponding to your rating. Before the study begins, there will be practice trials to familiarize you with the task.*

The instructions were presented on the screen with breaks between thoughts and between steps (e.g., "Then you will hear audio play." ¶ "After the audio plays, a scale will appear…" ¶ "When the scale appears…"). Listeners were given a few practice trials with the task-specific manipulation using a sham sequence. These trials were not meant to teach listeners about rhythm; they were meant to familiarize them with the task. After the practice trials, listeners were presented with the experimental stimuli. The task took an average of 18 minutes to complete (+/– 5.7 minutes).

***4. Analyses***

Ratings that were provided too quickly (< 300 ms) or too slowly ( > 2400 ms) were excluded from the analyses (= 14% of the data), following best practices (see Ratcliff, 1993). Ratings were then averaged within listener across repetitions to generate a single score for each unique stimulus that the listener heard. This procedure resulted in a total of 3,600 ratings per vowel context (50 listeners x 24 speakers x 3 unique elicitations per speaker). In a first set of analyses, a linear mixed-effects model was

used to test the fixed effects of age group (Group: 2 levels) and vowel context (Context: 2 levels) and their interaction on ratings. The model was built using the lme4 package (Bates et al., 2015) in R (R Development Core Team, 2019), and included a random intercept for every combination of the levels of listener and speaker. The lmerTest package (Kuznetsova et al., 2017) was used to estimate the degrees of freedom with Satterthwaite's method (Satterthwaite, 1946).

In a second set of analyses, multiple linear regression was used to evaluate the specific influence of determiner vowel reduction on goodness ratings. The models were implemented in SPSS (IBM SPSS Statistics Version 27). The ratings for the [ɑ]_[i] and [i]_[ɑ] contexts were fit separately. To maximize the independence of residuals in the model, ratings were standardized across listeners within a context based on individual listener means and standard deviations (i.e., z-scored). The Durbin-Watson statistic indicated a minor case of positive autocorrelation in each model ($d$ = 1.50 in the [ɑ]_[i] model; d = 1.43 in the [i]_[ɑ] model). This was corrected by adding a lag-1 of the dependent variable, which was entered first in the model. The assumption of homoscedasticity was also met in the data: scatterplots of predicted values versus residuals showed no relationship in either the [ɑ]_[i] or [i]_[ɑ] model. Q-Q plots of the residuals indicated that the assumption of normality was also met in both models.

The predictor variables in the multiple regression analyses were the acoustic correlates of temporal and amplitude reduction from Experiment 1 (Det:V duration, Det:V amplitude, Det:N duration, Det:N amplitude) and a single measure of schwa coarticulation. This measure was based on the mean formant frequency measures from Experiment 1. It was the Euclidean distance of each schwa produced by the speaker from the mean schwa for that speaker. All predictor values were log transformed to minimize the influence of extreme values on the results. Tests for correlations between the acoustic predictor variables entered into the models indicated expected significant pairwise correlations between relative duration and amplitude for Det:V and Det:N as well as an unsurprising relationship between relative duration and amplitude. The very strongest relationship, which was between Det:N and Det:V duration, had a coefficient of nearly 0.8 (Pearson's r = 0.797) in the [a]_[i] data and of nearly 0.7 in the [i]_[a] data

(Pearson's r = 0.660). Despite this, collinearity in the models was low; the largest VIF value was 2.51 in the [a]_[i] model and it was 2.30 in the [i]_[a] model).

After the predictor variables of interest were included in the model, age group was added to control for age-related effects that were not of interest in the experiment (e.g., the effect of F0). Speaker was initially included for the same reason, but then eliminated because its effect was not significant.

**B. Results**

The data in Figure 5 show that goodness ratings were lower overall for child speech than for adult speech [$F(1, 22) = 6.10$, $p = 0.022$], but this effect interacted with vowel context [$F(1, 7076) = 35.87$, $p < 0.001$].

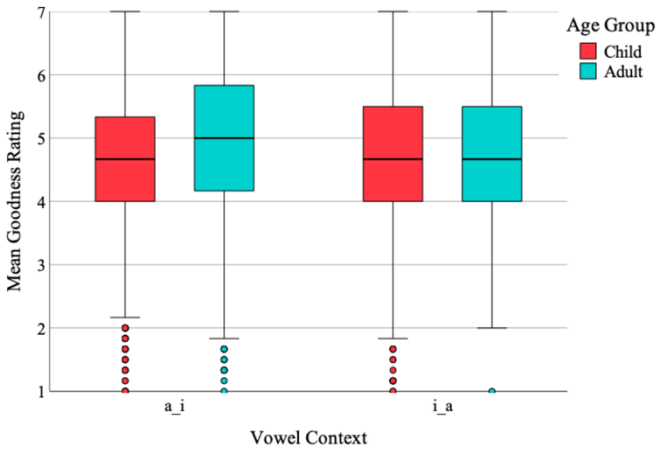


*Figure 5. Goodness ratings on V-the-N sequences in child and adult speech as a function of vowel context.*

The second set of analyses tested for the independent influence of reduction on listener ratings. As expected based on the mixed effects model results, age group accounted for a significant proportion of the variance in the full [ɑ]_[i] model [b = −.143, $t(3591) = -7.207$, $p < 0.001$], but not in the [i]_[ɑ] model [b = −0.018, $t(3591) = -1.03$, $p = 0.302$]. The strongest predictor variable of interest in both models was schwa coarticulation [ɑ_i model: b = −.125, $t(3591) = -7.70$, $p < 0.001$; i_ɑ model: b = −0.055, $t(3591) = -3.36$, $p < 0.001$]. Det:N duration was also significant in the [ɑ]_[i] model [b = −0.076, $t(3591) = -3.28$, $p$

< 0.001] and nearly so in the [i]_[a] model [b = −0.043, $t(3591) = -1.96$, $p = 0.05$]. Det:N amplitude was significant in the [i]_[ɑ] model [b = 0.042, $t(3591) = 2.15$, $p = 0.032$]. The overall model of ratings on stimuli from [ɑ]_[i] elicitations accounted for 10% of the variance (R = 0.317; Adjusted $R^2 = 0.099$), which represents a significant improvement over the null model [$F(7,3591) = 57.33$, $p < 0.001$]; the overall model of ratings on stimuli from [i]_[a] elicitations accounted for 9% of the variance (R = 0.302; Adjusted $R^2 = 0.089$), which was also significant [$F(7,3591) = 51.43$, $p < 0.001$].

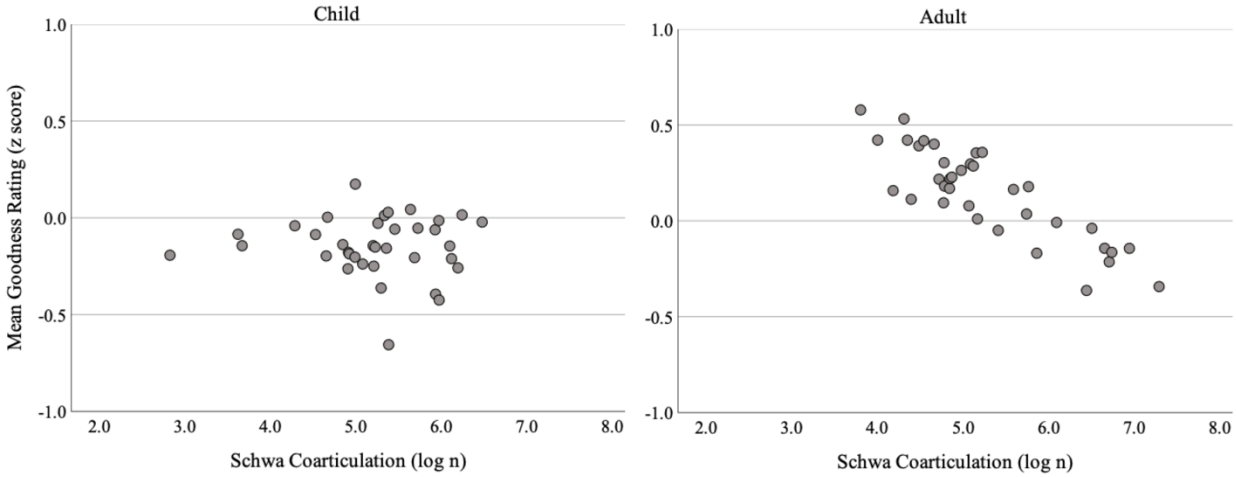


*Figure 6. The effect of schwa coarticulation on model predicted goodness ratings for [a]_[i] sequences as a function of age group.*

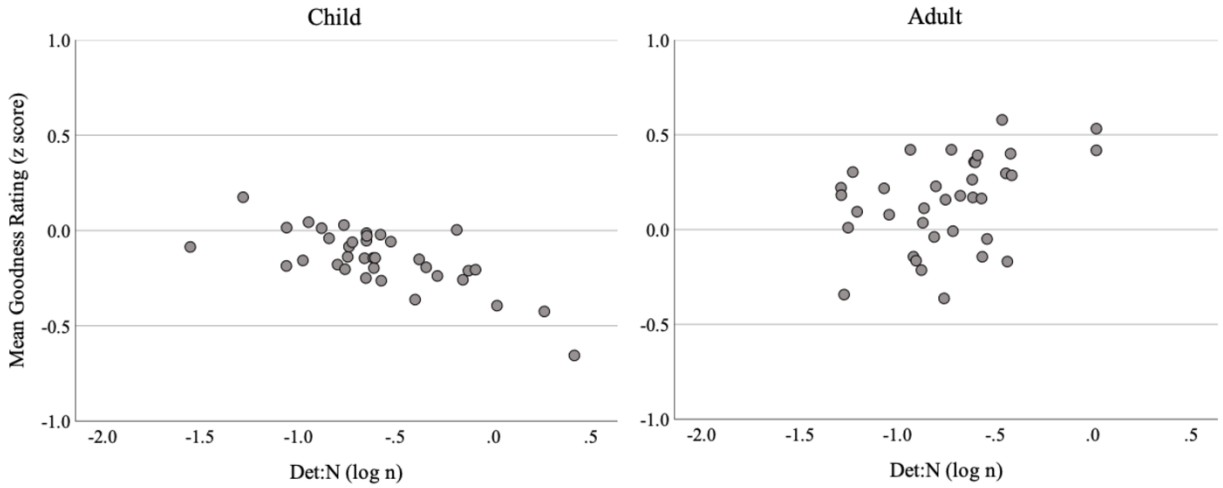


*Figure 7. The effect of relative schwa duration (Det:N) on model predicted goodness ratings for [a]_[i] sequences as a function of age group.*

When the analyses on ratings of [ɑ]_[i] sequences were conducted separately by age group, the relationship between the acoustic predictors and goodness ratings was found to differ for child and adult speech. For example, Figure 6 shows the relationship between determiner vowel coarticulation (log transformed) and goodness ratings (z-scored); Figure 7 shows the relationship for Det:N duration and goodness ratings. In adult speech (right panels), the data show that goodness rating varied inversely with degree of coarticulation in adult speech, but not with relative duration. Specifically, when the determiner vowel was further away from an adult speaker's average schwa in F1 × F2 × F3 space, the overall sequence was rated as less good than when it was closer to their average schwa; this relationship accounted for a significant proportion of the variance in the adult [ɑ]_[i] model [$b = -0.170$, $t(1793) = -7.36$, $p < 0.001$]. In child speech (left panels), goodness rating was uncorrelated with schwa coarticulation, but it was higher when schwa duration was shorter than when it was longer; this relationship also accounted for a significant proportion of the variance in the child [ɑ]_[i] model [Det:N duration, $b = -0.211$, $t(1792) = -6.01$, $p < 0.001$]. The significant inverse relationship in child speech but not adult speech is likely dependent on the greater range of relative durations, at both extremes, in the child speech data.

An unexpected positive relationship between Det:V and rating goodness was also found in child speech for [ɑ]_[i] sequences [b = 0.105, $t(1792) = 3.00$, $p = 0.003$]. Yet, a straightforward bivariate correlation between the variables shows the expected inverse relationship [r(1800) = −0.05, p = 0.03]. Also, Det:N and Det:V duration are positively correlated (see Methods). This results is therefore interpreted to indicate that Det:N duration captured all of the shared variance due to vowel reduction in the child [ɑ]_[i] model. Some additional variance in ratings was then accounted for by an increase in the determiner vowel relative to [ɑ] in the child model – a finding that could also indicate an influence on ratings from the stressed vowel itself. Overall, the adult [ɑ]_[i] model accounted for 13% of the variance in goodness ratings (R = 0.362; Adjusted $R^2 = 0.131$) and the child [ɑ]_[i] model accounted for 7% of the variance (R = 0.265, Adjusted $R^2 = 0.067$). Both models were significantly different from the null models [Child [ɑ]_[i] model: $F(6,1793) = 22.50$, $p < 0.001$; Adult [ɑ]_[i] model: $F(6,1793) = 44.94$, $p < 0.001$].

Consistent with the child [ɑ]_[i] model results, both Det:N duration and Det:N amplitude were significant predictors of goodness rating in the [i]_[ɑ] model when the non-significant effect of age group was removed [Det:N duration, b = −0.05, $t(3591) = -2.36$, $p = 0.019$; Det:N amplitude, b = 0.046, $t(3591) = 2.38$, $p = 0.017$]. The signs of the coefficients indicated that, as expected, goodness ratings were higher when the determiner vowel was shorter and of lower amplitude relative to the noun vowel. The effect of schwa coarticulation remained strong in the partial model as well [b = −0.056, $t(3591) = -3.37$, $p < 0.001$]. As in the full model, when schwa was further away from its mean value in F1 × F2 × F3 space listeners rated the sequence as less good than when it is closer to its mean value. The model $R^2$ is as before: R = 0.301; Adjusted $R^2 = 0.089$. Overall, the change in coefficients from the full [i]_[ɑ] model that controlled for age group and the partial [i]_[ɑ] model where this nonsignificant variable was removed suggests that listeners were sensitive to the overlap between the temporal and amplitude correlates of determiner vowel reduction and age group.

**C. Discussion**

The results from Experiment 2 upheld the expectation that adults would rate child speech as less good than adult speech. Also upheld was the expectation that function word reduction would predict

goodness ratings independently of the preference for adult speech. Although duration and amplitude measures of reduction were expected to predict goodness ratings better than schwa coarticulation, schwa coarticulation was the stronger overall predictor. Relative schwa duration and amplitude did not combine to explain additional variance in ratings within each model. Instead, relative schwa duration combined with schwa coarticulation to predict goodness ratings on [ɑ]_[i] sequences on top of the effect of age, and relative schwa amplitude combined with schwa coarticulation to predict ratings on [i]_[ɑ] sequences. On the other hand, the analyses by age group on ratings of [ɑ]_[i] sequences suggested that listeners attended to different correlates of reduction depending on the speaker's age: the global measure of coarticulation predicted goodness ratings of adult speech only; the relative duration of schwa in [ɑ]_[i] sequences was the only predictor of rating variance in child speech.

## V. GENERAL DISCUSSION

The strong-weak pattern of English rhythm leads to the reduction of function words, which are typically unstressed even when stranded in prosodic positions where they are considered extra-metrical (Selkirk, 1996). The high frequency with which function words occur in spoken language coupled with their minimal semantic weight contributes to their reduction (Bell et al., 2009; Jurafsky et al., 1998). But function words also provide critical grammatical information – to the point where they are treated as heads of phrases in some current theories of syntax (Müller, 2016). For this reason, if function words are reduced to the point of not being heard, the speaker's message may either be ungrammatical or the information they wish to communicate may be compromised (see, e.g., Baese-Berk et al., 2016). We argue that the slow development of adult-like function word reduction results from this tension between a rhythmic and semantic pressure towards maximal reduction and a listener-oriented pressure towards maximal intelligibility. The present results are consistent with findings from prior acoustic and kinematic studies of immature function word production in child speech: results from Experiment 1 were that 5-year-old speakers do not reduce determiners to the same extent as adult speakers in all contexts. More specifically, when child and adult determiners differs, they differs in the direction of relatively longer and higher amplitude schwa in child speech compared to adult speech.

Although schwa was relatively longer in child speech compared to adult speech, it was still much shorter than the vowels in the adjacent content words within the same sequence. In fact, it was often just half the duration of the adjacent content word vowel. This finding is consistent with children's relatively early acquisition of lexical stress patterns (Ballard et al., 2012). A possible implication of the finding is that children reduce monosyllabic function words to the same degree that they reduce weak syllables in content words, whereas adults reduce function words to a greater degree than they reduce weak syllables in content words. This possibility is consistent with empirical findings (Goffman, 2004; Fuchs, 2016; van Bergem, 1993). For example, Goffman (2004) compared child (4 to 7 years old) and adult production of weak syllables in two-syllable sequences. The syllables were embedded in a discourse context that led speakers to produce them either as a determiner-noun sequence or as an iambically stressed content word. Vertical movements of the lip-jaw complex were measured. In adult speech, the results were that movement duration and amplitude was smaller when the syllables were treated as a determiner-noun sequence than when it was treated as a disyllabic content word. In child speech, the results were that the weak syllable was produced with shorter and lower amplitude than the strong syllable, but there was no effect of morphosyntactic environment. This again suggests that children do not need to learn to reduce function words; they need to learn to reduce them to the same extent as adults.

The results from Experiment 1 also showed that both children and adults coarticulated the determiner with a following noun. If vowel-to-vowel coarticulation can be used to index chunking in the speech plan, this result is consistent with chunking along morphosyntactic lines rather than along metrical ones. Recall that the elicited sequences had a strong-weak-strong stress pattern; that is, strong monosyllabic content words separated by a weak function word. In this context, the weak function word should adhere to the preceding strong syllable to form a trochaic foot (Selkirk, 1996). Instead, the coarticulatory pattern found here suggests an adherence to the following strong syllable. This adherence pattern promotes a coherent determiner noun phrase. Given that listeners are known to use coarticulatory cues to aid in speech segmentation (Mattys, 2004), chunking along morphosyntactic lines has functional value. Of course, the distributional patterns of lexical stress in English are such that listeners also use a

trochaic pattern to aid in speech segmentation (Cutler and Butterfield, 1992; Mattys et al., 2005; Dilley and McAuley, 2008). This again suggests competing pressures on production: the chunking of determiner-noun sequences will (usually) result in the production of an iambic pattern; the most common disyllabic nouns and adjectives in English are produced with a trochaic pattern (see Cutler and Carter, 1987). Conflicting pressures such as these may also influence the production of determiner noun phrases.

Although the noun vowel influenced the production of schwa in the determiner in both child and adult speech, the transition towards the noun was only evident in children's speech. In adult speech, schwa was simply produced with a different quality depending on the identity of the noun vowel. It could be that these differences are merely an epiphenomenon of age-related differences in articulation rate. In particular, children's slower rate of articulation could provide them with more time to adjust articulatory movements into and out of a vowel target compared to adults. Such an explanation assumes that schwa coarticulation in adult speech is due to target undershoot (i.e., hypoarticulation; Agwuele et al., 2008; Moon and Lindblom, 1994). Target undershoot increases as the time allotted for target attainment is decreased so long as articulatory effort is held constant. If we explain the effect of age on coarticulation in these terms, it would suggest that children's slower articulation rates reflect a production strategy designed to maximize sequential target attainment. There is some evidence to support this possibility. Recall that children's speech movements are in fact larger relative to their oral-facial size (Riely and Smith, 2003), suggesting that they expend more effort in speaking compared to adults – presumably, in order to achieve acoustic targets.

Yet, if we explain the effect of age on coarticulation as an epiphenomenon of speech rate, we must still explain why the relative duration and amplitude of schwa is (typically) less than that of an adjacent full vowel. Saying that it is "reduced" only says that, at some level of representation, its timing is due to language factors, which is to say that its timing is planned before execution. And, if timing is part of the plan, then coarticulation may not be due to target undershoot, but instead to target overlap (Fowler, 1980). This view of coarticulation suggests an alternative explanation for the present results – one that is more consistent with the idea of competing rhythmic and morphosyntactic pressures on production: if

children are less able than adults to resolve these pressures, their determiner-noun sequences are less tightly bound than in adult speech. If the determiner-noun sequence is less tightly bound in child speech than in adult speech, then the vowels in children's sequences will also be less overlapped and schwa less subject to "truncation" allowing for its fuller realization (Harrington et al., 1995).

Regardless of exactly why children do not reduce function words to the same extent as adults, they eventually learn to do so. Our suggestion is that learning requires a honing process that depends on feedback in the form of communicative success or failure. This hypothesis is consistent with an adult listener preference for adult speech over child speech. It also predicts that listener ratings of speech rhythmicity will be most influenced by those measures of reduction that best distinguish between child and adult speech. The results from Experiment 2 confirmed the predicted listener preference for adult speech over child speech, but the influence of specific measures of reduction on listener preference was more complicated than we had expected. Based on the results from Experiment 1, we had expected that measures of relative duration and amplitude of schwa (especially, Det:V duration and Det:V amplitude) would predict listener goodness ratings on V-*the*-N sequences. Instead, the strongest predictor of goodness ratings was a global measure of schwa coarticulation.

The global measure of schwa coarticulation that was used as a predictor variable in analyses of goodness ratings in Experiment 2 was calculated as the Euclidean distance of schwa from the speaker mean schwa in F1 × F2 × F3 space. Goodness ratings on V-*the*-N sequences decreased when schwa was more distant from the speaker mean. The assumption is that the larger the distance from the mean, the more schwa varied as a function of context. But production variability may be due to other sources as well, including immature motor skills (Smith and Zelaznik, 2004). The result could therefore indicate that listeners were attending to segmental target attainment, that schwa has a context-sensitive target (Browman and Goldstein, 1992), and that reduction per se was not relevant to the judgments that were made. Then again, the possibility that listeners rated sequences based on how similar the realization of schwa was to an expected schwa target could also be due to the experimental design.

The speech stimuli were blocked both by speaker and by age group. The goal of blocking was to attune listeners to the rhythmicity of the sequence and away from global differences in production due to individual or group differences (e.g., differences in F0, differences in segmental articulation). But the blocked design also likely attuned listeners to that which was most variable within a speaker and age group. It could be that temporal and amplitude patterns are simply more stable than the realization of schwa within a speaker and age group. But this explanation does not account for why relative schwa duration was the only significant predictor of ratings on [ɑ]_[i] sequences elicited from children or why the global measure of coarticulation was the only significant predictor of ratings on [ɑ]_[i] sequences in adult speech. Also, child speech is notoriously variable and this variability is especially significant in the spatial-temporal domain (Smith and Zelaznik, 2004). It therefore seems likely that listener expectations for reduction provide a better explanation for why their ratings on adult sequences were influenced by different correlates than their ratings on child sequences. For example, it could be that listeners expect some average amount of reduction in the temporal and amplitude domains and variability around that amount is less important than variability around an average that is already higher than the amount expected.

Although the results conform in several ways to the predictions made, the gap between the present results and the working hypothesis that communication success shapes speech motor learning is still admittedly large. Much more work is needed to firmly establish a relationship between child speech production and adult speech processing. Future research could start by confirming the relationship between function word reduction and children's speech rhythm. This relationship is assumed based on perceptual analyses of children's speech (e.g., Allen and Hawkins, 1978) and on a correlation between interval-based rhythm measures and the relative duration measures reported here (e.g., Sirsa and Redford, 2011). Combined acoustic and perceptual studies on large samples of spontaneous speech would go some way towards confirming the relationship. If confirmed, it would also be worth directly testing the link between violations in rhythm due to more or less reduced function words in adult speech processing. More generally, the relationship between child speech production and adult speech perception requires

further study. Although the link between immature motor skills and speech intelligibility is well-established, the relationship is rarely investigated in detail. For example, it is not clear how adult listeners weigh the relative contribution of variable segmental articulation and immature speech rhythm when processing child speech. Finally, ecologically-valid descriptions of adult listener response to child speech and child responses to adult behaviors is needed to better characterize the details of how communicative interactions may drive speech motor learning.

**ACKNOWLEDGMENTS**

This research was wholly supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) under grant R01HD087452 (PI: Redford). The content is solely the authors' responsibility and does not necessarily reflect the views of NICHD.

## REFERENCES


Abu-Akel, A., Bailey, A. L., and Thum, Y. M. (**2004**). "Describing the acquisition of determiners in English: A growth modeling approach," J. Psycholinguist. Res. **33**(5), 407-424.

Agwuele, A., Sussman, H. M., and Lindblom, B. (**2008**). "The effect of speaking rate on consonant vowel coarticulation," Phonetica. **65**(4), 194-209.

Allen, G., and Hawkins. S. (**1978**). "The development of phonological rhythm," in *Syllables and Segments*, edited by A. Bell and J. Bybee Hooper (North-Holland Publishing: New York), pp. 173–185.

Baese-Berk, M. M., Dilley, L. C., Schmidt, S., Morrill, T. H., and Pitt, M. A. (**2016**). "Revisiting Neil Armstrong's moon-landing quote: implications for speech perception, function word reduction, and acoustic ambiguity," PloS one. **11**(9), e0155975.

Ballard, K. J., Djaja, D., Arciuli, J., James, D. G., and van Doornd, J. (**2012**). "Developmental trajectory for production of prosody: lexical stress contrastivity in children ages 3 to 7 years and in adults," J. Speech Lang. Hear. Res. **55**, 1822-1835.

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

Bell, A., Brenier, J. M., Gregory, M., Girand, C., and Jurafsky, D. (**2009**). "Predictability effects on durations of content and function words in conversational English," J. Mem. Lang. **60**(1), 92-111.

Boersma, P. and Weenink, D. (**2019**). Praat: doing phonetics by computer [Computer program]. Version 6.0.49. http://www.praat.org/

Browman, C. P., and Goldstein, L. (1992). "Targetless schwa: An articulatory analysis." In *Papers in laboratory phonology II: Gesture, segment, prosody*, edited by G.J. Docherty and D.R. Ladd (Cambridge University Press, Cambridge), pp. 26–56.

Buhrmester, M., Kwang, T., and Gosling, S. D. (**2011**). "Amazon's mechanical turk: a new source of inexpensive, yet high-quality, data?" Perspect. Psychol. Sci. **6**(1), 3-5.

Cutler, A. and Butterfield, S. (**1992**). "Rhythmic cues to speech segmentation: Evidence from juncture misperception," J. Mem. Lang. **31**(2), 218-236.

Cutler, A. and Carter, D. (**1987**). "The predominance of strong initial syllables in the English vocabulary," Comput. Speech Lang. **2**, 133-142.

Dauer, R. M. (**1983**). "Stress-timing and syllable-timing reanalyzed," J. Phon. **11**, 51-62.

Deterding, D. (**2001**). "The measurement of rhythm: A comparison of Singapore and British English," J. Phon. **29**(2), 217-230.

Dilley, L. C. and McAuley, J. D. (**2008**). "Distal prosodic context affects word segmentation and lexical processing," J. Mem. Lang. **59**, 294-311.

Dilley, L. C. and Pitt, M. A. (**2010**). "Altering context speech rate can cause words to appear or disappear," Psychol. Sci. **21**(11), 1664-1670.

Dodd, B., Zhu, H., Crosbie, S., Holm, A., and Ozanne, A. (**2002**). *Diagnostic evaluation of articulation and phonology (DEAP)*. (Psychological Corporation, London).

Fikkert, P., Liu, L.,and Mitsuhiko, O. (**2021**) "The acquisition of word prosody," in *The Oxford Handbook of Language Prosody*, edited by C. Gussenhoven and A. Chen (Oxford University Press, London), pp. 541-552.

Fourakis, M. (**1991**). "Tempo, stress, and vowel reduction in American English," J. Acoust. Soc. Am. **90**(4), 1816-1827.

Fowler, C. A. (**1980**). "Coarticulation and theories of extrinsic timing," J. Phon. **8**(1), 113-133.

Fowler, C. A. (**1981**). "Production and perception of coarticulation among stressed and unstressed vowels," J. Speech Hear. Res. **24**(1), 127-139.

Fuchs, R. (**2016**). "The acoustic correlates of stress and accent in English content and function words," Proc. Speech Prosody, **8**, pp. 290-294.

Gay, T. (**1981**). "Mechanisms in the control of speech rate," Phonetica. **38**(1-3), 148-158.

Gerken, L. (**1996**). "Prosodic structure in young children's language production," Language. **72**(4), 683-712.

Goffman, L. (2004). "Kinematic differentiation of prosodic categories in normal and disordered language development," J. Speech Lang. Hear. Res., **47**(5), 1088-1102.

Grabe, E., and Low, E. L. (**2002**). "Durational variability in speech and the rhythm class hypothesis," in C. Gussenhoven and N. Warner (eds.), *Laboratory Phonology 7* (Mouton de Gruyter, Berlin), pp. 515-546.

Gu, C. (**2014**). "Smoothing spline ANOVA models: R package gss," J. Stat. Softw. **58**(5), 1-25.

Harrington, J., Fletcher, J., and Roberts, C. (**1995**). "Coarticulation and the accented/unaccented distinction: evidence from jaw movement data," J. Phon. **23**(3), 305-322.

He, L. (**2012**). "Syllabic intensity variations as quantification of speech rhythm: evidence from both L1 and L2," Proc. Speech Prosody, pp. 466-469.

Jurafsky, D., Bell, A., Fosler-Lussier, E., Girand, C.,and Raymond, W. (**1998**). "Reduction of English function words in switchboard," in ICSLP-98, 3111-3114.

Kedar, Y., Casasola, M.,and Lust, B. (**2006**). "Getting there faster: 18-and 24-month-old infants' use of function words to determine reference," Child Dev. **77**(2), 325-338.

Kehoe, M., Stoel-Gammon, C. and Buder, E.H. (**1995**). "Acoustic correlates of stress in young children's speech," J. Speech Hear. Res. **38**, 338-350.

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (**2017**). "lmerTest package: test in linear mixed effects models," J. Stat. Softw. **82**(13), 1-26.

Lee, S., Potamianos, A.,and Narayanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**(3), 1455-1468.

Lindblom, B., Sundberg, J., Branderud, P., and Djamshidpey, H. (**2011**). "Articulatory modeling and front cavity acoustics," Proc. Fonetik TMH-QPSR. **51**(1), 17-20.

Mattys, S. L. (**2004**). "Stress versus coarticulation: toward an integrated approach to explicit speech segmentation," J. Exp. Psychol. Hum. Percept. Perform., **30**(2), 397.

Mattys, S.L., White, L., and Melhorn, J.F. (**2005**). "Integration of multiple speech segmentation cues: a hierarchical framework," J. Exp. Psychol. Gen. **134**(4), 477.

Mielke, J. (**2015**). "An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons," J. Acoust. Soc. Am. **137**(5), 2858-2869.

Moon, S. J.,and Lindblom, B. (**1994**). "Interaction between duration, context, and speaking style in English stressed vowels," J. Acoust. Soc. Am. **96**(1), 40-55.

Müller, Stefan (**2016**). *Grammatical Theory: From Transformational Grammar to Constraint-based Approaches*. (Language Science Press, Berlin).

Nittrouer, S. (**1993**). "The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study," J. Speech Lang. Hear. Res. **36**(5), 959-972.

Payne, E., Post, B., Astruc, L., Prieto, and Vanrell, M. (**2011**). "Measuring child rhythm," Lang. Speech **55**(2), 203-229.

Plag, I., Kunter, G., and Schramm, M. (**2011**). "Acoustic correlates of primary and secondary stress in North American English," J. Phon. **39**(3), 362-374.

Polyanskaya, L.and Ordin, M. (**2015**). "Acquisition of speech rhythm in first language," J. Acoust. Soc. Am. **138**(3), EL199-EL204.

R Core Team. (**2019**). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.

Ratcliff, R. (**1993**). "Methods for dealing with reaction time outliers," Psychol. Bull. **114**(3), 510-532.

Redford, M. A. (**2018**). "Grammatical word production across metrical contexts in school-aged children's and adults' speech," J. Speech Lang. Hear. Res. **61**, 1339-1354.

Redford, M.A. (**2014**). "The perceived clarity of children's speech varies as a function of their default speech rate," J. Acoust. Soc. Am. **135**, 2952-2963.

Redford, M. A., Kapatsinski, V., and Cornell-Fabiano, J. (**2018**). "Lay listener classification and evaluation of typical and atypical children's speech," Lang. Speech, **61**, 277-302.

Riely, R.R., Smith, A. (**2003**). "Speech movements do not scale by orofacial structure size," J. Appl. Physiol. **94**, 2119-2126.

Satterthwaite, F. E. (**1946**). "An approximate distribution of estimates of variance components," Biometrics **2**, 110-114.

Schwartz, R. G., Petinou, K., Goffman, L., Lazowski, G. and Cartusciello, C. (**1996**). "Young children's production of syllable stress: An acoustic analysis," J. Acoust. Soc. Am. **99**(5), 3192-3200.

Selkirk, E. (**1996**). "The prosodic structure of function words," in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, edited by J.L. Morgan and K. Demuth (eds.), (Erlbaum, Mahwah), pp. 187-214.

Shriberg, L. D., Campbell, T. F., Karlsson, H. B., Brown, R. L., McSweeny, J. L., and Nadler, C. J. (**2003**). "A diagnostic marker for childhood apraxia of speech: The lexical stress ratio," Clin. Linguist. Phon. **17**(7), 549-574.

Smith, A., and Zelaznik, H. N. (**2004**). "Development of functional synergies for speech motor coordination in childhood and adolescence," Dev. Psychobio. **45**(1), 22-33.

Sirsa, H. and Redford, M. A. (**2011**). "Towards understanding the protracted acquisition of English rhythm," Proc. Int. Congress Phon. Sci. 1862-1865.

Tilsen, S., and Arvaniti, A. (**2013**). "Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages," J. Acoust. Soc. Am. **134**(1), 628-639.

Van Bergem, D. R. (**1993**). "Acoustic vowel reduction as a function of sentence accent, word stress, and word class," Speech Commun. **12**(1), 1-23.

Wiig, E. H., Secord, W., and Semel, E. (**2013**). *Clinical Evaluation of Language Fundamentals (CELF-5)*, (Harcourt Brace Jovanovich, San Antonio, TX).

## FIGURE LEGENDS

**Figure 1.** Example of vowel segmentations for a Verb-the-Noun sequence embedded in a simple S-V-O sentence elicited from a 5-year-old child.

**Figure 2.** Vowel duration ratios for Det:V (left) and Det:N (right) shown by age group and the vowel context within which the determiner appeared.

**Figure 3.** Vowel amplitude ratios for Det:V (left) and Det:N (right) are shown by age group and the vowel context within which the determiner appeared.

**Figure 4.** Best-fit curves for schwa F1, F2, and F3 are shown as a function of the stressed vowel context within which the determiner was produced in child (left) and adult (right) speech. Dotted lines represent the 95% confidence interval.

**Figure 5.** Goodness ratings on V-the-N sequences in child and adult speech as a function of vowel context.

**Figure 6.** The effect of schwa coarticulation on model predicted goodness ratings for [ɑ]_[i] sequences as a function of age group.

**Figure 7.** The effect of relative schwa duration (Det:N) on model predicted goodness ratings for [ɑ]_[i] sequences as a function of age group.