Grammatical word production across metrical contexts

in school-aged children's and adults' speech

Melissa A. Redford

University of Oregon

ACCEPTED FOR PUBLICATION IN *JSLHR* FEBRUARY 6, 2018

Keywords: acoustic measures; vowel reduction; coarticulation; prosodic words.

Please address correspondence to:

Melissa A. Redford
Linguistics Department
1290 University of Oregon
Eugene, OR 97403

email: redford@uoregon.edu
tel. 541-346-3818

ABSTRACT

Purpose: Test whether age-related differences in grammatical word production are due to differences in how children and adults chunk speech for output or to immature articulatory timing control in children.

Method: Two groups of 12 children, 5 and 8 years old, and one group of 12 adults produced sentences with phrase-medial determiners. Preceding verbs were varied to create different metrical contexts for chunking the determiner with an adjacent content word. Following noun onsets were varied to assess the coherence of determiner-noun sequences. Determiner vowel duration, amplitude, and formant frequencies were measured.

Results: Children produced significantly longer and louder determiners than adults regardless of metrical context. The effect of noun onset on F1 was stronger in children's speech than in adult speech; the effect of noun onset on F2 was stronger in adults' speech than in children's. Effects of metrical context on anticipatory formant patterns were more evident in children's speech than in adults' speech.

Conclusion: The results suggest that both immature articulatory timing control and age-related differences in how chunks are accessed or planned influence grammatical word production in school-aged children's speech. Future work will focus on the development of long-distance coarticulation to reveal the evolution of speech plan structure over time.

Introduction

Many children with developmental disabilities, including those with childhood apraxia of speech (CAS) and autism spectrum disorders (ASD), produce speech with atypical rhythm patterns (Paul, Augustyn, Klin, & Volkmar, 2005; Shriberg, Paul, McSweeny, Klin, Cohen, & Volkmar, 2001). Atypical rhythm increases the perception of disorder (Olejarczuk & Redford, 2013; Paul, Shriberg, McSweeny, Cicchetti, Klin, & Volkmar, 2005) and decreases speech intelligibility (Munro & Derwing,1995; Weismer & Martin, 1992). Perceived difference and poor intelligibility result in negative social evaluations (McCabe & Meller, 2004; Redford, Kapatsinski, Cornell-Fabiano, 2017; Rice, Sell, & Hadley, 1991), which can undermine access to positive social interactions. To effectively intervene and ameliorate atypical rhythm patterns, we must understand how rhythm typically emerges. Current understanding is largely based on a metrical theoretic approach to explaining weak syllable omissions in early child language (e.g., Demuth, 2001; Gerken, 1996; Gerken, Landau, & Remez, 1990; Wijnen, Krikhaar, & Os, 1994). As a result of this focus, current understanding of speech rhythm acquisition does not accommodate the classic observation that school-aged children's speech is more equally-timed than adults' speech (Allen & Hawkins, 1978). Nor does it incorporate into theory ideas about how the protracted development of speech motor skills, including planning and control, may contribute to the emergence of English rhythm. The research reported here addresses these limitations by testing between specific hypotheses that follow from a more general working hypothesis for why rhythm production is still immature during the school-age years.

The working hypothesis

The rhythm of American English is defined in large part by the alternation of unstressed and stressed vowels in running speech. Unstressed vowels are "reduced" relative to stressed vowels, which is to say that they are shorter, quieter, and more coarticulated with adjacent speech sounds than stressed vowels (Fowler, 1981; Fourakis, 1991; Plag, Kunter, & Schramm, 2011). When the durational and amplitude differences between unstressed and stressed vowels are less pronounced, speech is perceived as more equally stressed or syllable-timed (Ling, Grabe, & Nolan, 2000; Barry, Andreeva, & Koreman, 2009; Tilsen & Arvaniti, 2013). This is the case in child language (Allen & Hawkins, 1978; 1981). Findings from developmental studies using interval-based metrics suggest that the perception of English-speaking children's speech as (quasi) syllable-timed is due to smaller differences between successive vowel durations (Bunta & Ingram, 2007; Grabe, Post, & Watson, 1999; Payne, Post, Astruc, Prieto, & del Mar Vanrell, 2012; Polyanskaya & Ordin, 2015; Sirsa & Redford, 2011). Since children are known to produce adult-like patterns of relative vowel duration and amplitude in multisyllabic lexical words as early as age 2 years (Kehoe, Stoel-Gamon, & Buder, 1995; Pollock, Brammer & Hageman, 1993; Schwartz, Petinou, Goffman, Lazowski, & Cartusciello, 1996), our working hypothesis for the prolonged acquisition of speech rhythm is as follows: young school-aged children's speech is more equally timed than adults' speech because of age-related differences in the production of supralexical prosodic structures, which emerge when unstressed grammatical words are reduced and prosodified (i.e., cliticized or chunked) with adjacent content words to form prosodic words.

Grammatical words, such as auxiliaries and determiners, serve mainly to indicate grammatical relations between other words in a sentence; they are structural elements with abstract meaning that is often better described with respect usage than to a specific concept. Grammatical words are also acquired later than content words, such as verbs and nouns, even though they are very high frequency lexical items (Caselli, et al., 1995). Our working hypothesis for the slow acquisition of English speech rhythm assumes that grammatical words also have a special production status in young school-aged children's speech. This assumption follows from Allen and Hawkin's (1978) classic observation based on transcribed data that children produce fuller grammatical words than adults. The aim of the present study was to determine whether age-related differences in grammatical word production is due to differences in how children and adults chunk speech for execution (= chunking hypothesis) or is simply a consequence of children's immature articulatory timing control (= articulatory timing hypothesis).

The chunking hypothesis

Psycholinguistic theory assumes that prosodic words are the principle units of speech planning and production (Levelt, Roelofs, & Meyer, 1999; Wheeldon & Lahiri, 1997; 2002; see also, Sternberg, Knoll, Monsell, & Wright, 1988). As units of production, we expect the component sounds and syllables to cohere more tightly compared to sounds and syllables that are separately planned. Coherence can be defined with reference to anticipatory coarticulation (see, e.g., Whalen, 1990; Ma, Perrier, & Dang, 2015); for example, a grammatical word that is chunked with a subsequent content word during speech planning will be produced with greater influence from the subsequent word than one that is planned as an independent prosodic unit. Grammatical words that are independent prosodic units

will not only be less coarticulated with adjacent content words, they will also be longer and louder than those that are part of a prosodic word (see Selkirk, 1996).

According to linguistic theory, prosodic word formation is highly constrained by the metrical foot, which is a language-specific rhythmic grouping of syllables (Hayes, 1995). In most standard dialects of English, the foot groups strong and weak syllables together in that order to form trochees (Hayes, 1982). This is also true when the weak syllable is a grammatical word and so an independent lexeme. For example, the determiner in the sequence "Tom's a cat" is footed with "Tom" to form a prosodic word. Note that, in this instance and others like it, metrically-based chunking means that the prosodic word boundary is misaligned with respect to syntactic constituency (i.e., $[\textit{Tom's a}]_{\text{PW}}$ $[\textit{cat}]_{\text{PW}}$ versus $[\textit{Tom's}]_{\text{NP}}[\textit{a cat}]_{\text{NP}}$). This misalignment is frequently cited as evidence for separate syntactic and prosodic grammars (see, e.g., Shattuck-Hufnagel & Turk, 1996).

Although syntax and prosody may be distinct, syntax influences the prosodic chunking of grammatical words when these cannot be footed (Selkirk, 1996); for example, when it is the second weak syllable in a sequence of weak syllables. In such cases, the grammatical word is chunked with reference to constituent structure. In this way, syntactic structure can help to create production units that violate the preferred rhythm pattern of a language. For example, an unfooted determiner will be chunked with the noun it modifies to form an iambically-stressed prosodic word (e.g., $[\textit{Tommy's}]_{\text{F}}]_{\text{PW}}$ $[\textit{a}$ $[\textit{cat}]_{\text{PW}}$).

Violations of the trochaic pattern in English may slow speech rhythm acquisition. In particular, a number of child language researchers have argued that children are slower to acquire prosodic words that have iambic stress compared to those that have trochaic stress (e.g., Demuth & Fee, 1995; Fikkert, 1994; Gerken, 1996). Their evidence is very young

children's omission of word-initial weak syllables in multisyllabic words (e.g., ba'nana → 'nana) and their variable realization of grammatical words. For example, Gerken (1996) showed in a series of elegant experiments that 2-year-old children were more likely to produce a determiner (i.e., "the") in contexts where it could be footed with a preceding strong syllable than in contexts where it was unfooted. In the experiments most relevant to our interest in grammatical word production (Experiments 1-3), Gerken manipulated metrical structure in simple 4 and 5 word SVO sentences by manipulating verbal inflections to achieve either a stressed monosyllabic word (e.g., "pushed") or a trochaically-stressed disyllabic word (e.g., "pushes"). In the critical cases, the object was a noun phrase (e.g., "the pig"): children were asked to produce sentences such as "Tom pushed the pig" and "Tom pushes the pig". In experiment after experiment, 2-year-olds were more likely to produce "the" when it was preceded by a monosyllabic verb (e.g., "pushed") than when it was preceded by a disyllabic verb (e.g., "pushes"); a result that suggests slower acquisition of prosodic words with unfooted determiners compared to those with footed determiners. Could this effect persist into the school-age years to influence how children chunk speech for output?

Intriguingly, school-aged children also have a bias for producing trochaic patterns (Goffman & Westover, 2013; Redford & Oh, 2016). Redford and Oh (2016) showed that 5- and 8-year-old children were much more likely to blend two separately-presented, equally-stressed syllables into a single nonce word if the phonological structure of the sequence encouraged trochaic stress. Conversely, if phonological structure encouraged iambic stress, children would either alter the pronunciation of the syllable sequence to produce a trochaically-stressed nonce word or they would produce the syllables with equal stress and

therefore as a prosodic word sequence. Although older children also preferred to produce trochees, they were sensitive to the biasing influence of a syntactic manipulation (noun frame → trochees; verb frame → iambs). Their nascent sensitivity to syntactic influences on lexical stress suggests a developmental hypothesis with respect to rhythm acquisition: as speech and language skills develop, the influence of speech sound patterning on production processes is demoted relative to the influence of language structure and meaning.

If the later stages of speech rhythm acquisition involve overcoming rhythmic preferences so that speech is prosodically chunked with reference to constituent structure, then we might expect school-aged children's speech to reflect language influences that go beyond simple rhythmic distinctions. This expectation is supported by Goffman and colleagues extensive study of different metrical effects on children's motor speech behavior (e.g., Goffman & Malin, 1999; Goffman, 2004; Goffman, Heisler, & Chakraborty, 2006; Goffman & Westover, 2013). A major conclusion of this work is that children's motor speech rhythmicities provide evidence for differentiated prosodic and/or morphosyntactic categories. For example, Goffman & Westover (2013) investigated lip+jaw movement patterns during production of differently stressed disyllabic nouns (trochaic = *baby*, iambic = *baboon*) in two different sentence frames: one with a free standing determiner (*Em's a* ___) and one with a word-final weak syllable (*Emma's* ____). Adults, 4-year-old children with typical development, and 4- to 6-year-old children with specific language impairment (SLI) produced smaller movement amplitudes when the noun followed a determiner than when it followed another noun regardless of stress pattern type. All speakers, but especially children, had more trouble producing sequences of weak syllables compared to producing

an alternating pattern of strong-then-weak syllables. This was even more true when the weak syllable sequence included an unfooted syllable (i.e., with the determiner frame).

In sum, a metrical theoretic approach to prosodic word formation highlights a tension between basic rhythmic preferences and syntactic constituency for delimiting units of speech production in English. Unstressed grammatical words are preferentially chunked with an adjacent content word to form trochees. When the speech sequence does not allow for this preferred rhythmic chunking, these grammatical words are chunked with reference to syntactic constituency. The importance of both metrical structure and syntax on prosodic word chunking could help explain the slow acquisition of English rhythm. In particular, later stages of speech rhythm acquisition may involve calibrating the strength of a trochaic bias in production and attending more to syntactic structure.

The articulatory timing hypothesis

The chunking hypothesis assumes that prosodic words are recoverable from speech acoustics because they represent production units that are delimited in the speech plan. The theory behind the hypothesis is that the structure of the speech plan emerges with lexical acquisition and production practice over developmental time (Redford, 2015). In mainstream theory, however, speech plans are derived from the phonology. Patterns of syllable omission and other systematic transformations of adult target words reflect a child phonology. When children produce all elements of target words in sentences, the acquisition of phonology is complete. Still, speech sound distortions persist. This is because of the slow development of speech motor skills; namely, the ability to coordinate and sequence articulatory movements. This ability, known as articulatory timing control, is not fully acquired until middle adolescence (Smith & Zelaznik, 2002; Walsh & Smith, 2002).

Low-level speech sound distortions due to immature articulatory timing control include longer and more variable segment durations (Lee, Potamianos, & Narayanan, 1999; Redford, 2014). Longer segment durations are in turn due to children's larger amplitude and lower velocity speech movements (Smith & Goffman, 1998; Riely & Smith, 2002). Variability is due the slow development of stable articulatory synergies (Smith & Goffman, 1998; Smith & Zelaznik, 2002).

Intriguingly, many markers of immature speech motor control are more pronounced in children's production of weak syllables than in their production of strong syllables (Goffman, Gerken, & Lucchesi, 2007; Goffman, Heisler, et al., 2007). This may have less to do with stress per se and more to do with the relative incompressibility of weak syllables. We know, for example, that adults increase overall articulatory rates by compressing the duration of vowels in strong syllables more than those in weak syllables (Gay, 1978; 1981), presumably because the latter are already short relative to the former and cannot be produced any more rapidly without sacrificing intelligibility. It could be that, like adults, children's weak syllables are equally compressed, but immature articulatory timing control means that it takes them longer than adults to produce short syllables that are intelligible. Relatedly, Goffman and colleagues have shown that adults compress unfooted grammatical words in an iambic prosodic word context more than footed ones in a trochaic prosodic word context (Goffman & Malin, 1999; Goffman, 2004; Goffman et al., 2006). This suggests that unfooted grammatical words may be especially challenging for children to produce in an adult-like fashion. More generally, it suggests an articulatory timing hypothesis: children fail to reduce grammatical words to the same extent as adults simply because they cannot. No difference in the underlying production unit (i.e., the prosodic word) need be assumed;

longer and louder grammatical words simply index immature articulatory timing control in school-age children.

The articulatory timing hypothesis is consistent with the view that children acquire an adult-like phonology before they acquire adult-like articulatory timing control. This view predicts a dissociation between the representation and execution of speech timing patterns. Redford and Oh (2017) provide some evidence for such a dissociation at the level of the lexical word: English-speaking school-aged children were found to produce the same language-specific, within-word temporal patterns as adults, but their productions were often slower and always more variable; in contrast, adult second-language learners of English produced different within-word temporal patterns than native English-speaking adults, but their productions were as stable as native speakers' productions. These findings were interpreted to reflect a double dissociation; specifically, "immature control (execution) over adult-like specification of sequential motor goals (representation) in children, and mature control (execution) over the realization of non-native specification of sequential goals (representation) in adult second language learners. (pp. 135-136)."

Current study

To summarize, the chunking hypothesis explains that age-related differences in grammatical word production emerge due to children's immature calibration of metrical and syntactic pressures on prosodic word formation, which renders grammatical words more variable in their acoustic realization across metrical contexts. In particular, school-age children may fail to chunk grammatical words with adjacent content words if a trochee is not possible. This would result in the production of unfooted grammatical words as independent prosodic words. Children's unfooted grammatical words should therefore be

longer, louder, and less coarticulated with the following content word than unfooted grammatical words in adult speech.

Insofar as the articulatory timing hypothesis allows for early acquisition of adult-like prosodic structures, it predicts that children and adults will chunk grammatical words with adjacent content words in the same way across different metrical contexts. Under the assumption that anticipatory coarticulation indexes coherence, the specific prediction is that children and adults will show a pattern of strong anticipatory coarticulation when unfooted grammatical words are chunked with a following content word and a pattern of weak anticipatory coarticulation when they are footed and chunked with the preceding content word. This prediction regarding the similar effect of metrical context on anticipatory coarticulation across age groups is made here absent a prediction about the effect of age on these patterns. It is unclear from the literature whether long-distance (heterosyllabic) anticipatory coarticulation in children's speech is stronger, weaker, or the same as in adults' speech (see, e.g., Nijland, Maassen, Van der Meulen, et al., 2002; Nittrouer, Studdert-Kennedy, & Neely, 1996; Noiray, Cathiard, Abry, & Menard, 2010; Repp, 1986).

The predictions from the chunking and articulatory timing hypotheses were tested in the current study. We investigated child and adult productions of a determiner (i.e., "the") as a function of metrical context and following noun onset. Metrical context was varied to create different chunking patterns. Noun onsets were varied to assess determiner + noun coherence. Children's age was also manipulated: younger children were 5 years old; older children were 8 years olds. Prior work suggests significant development change in articulatory timing control and speech rhythm between the ages of 5 and 8 years (see, e.g.,

Lee, Potamianos, & Narayanan, 1999; Sirsa & Redford, 2011). The chunking hypothesis predicts an interaction between metrical context and age group on measures of reduction and coherence; the articulatory timing hypothesis does not.

Methods

Participants

A total of 36 speakers participated in the study: two groups of 12 American English-speaking children, aged 5 and 8 years old; one group of 12 American English-speaking adults. The mean age of children in the 5-year-old group was five years, seven months (= 5;7). The range was from 5;2 to 6;3. The mean age in the 8-year-old group was 8;1. The range was from 7;7 to 8;8. The mean age of the adults was 19;2 years. The range was from 18 to 21 years. All study participants spoke a west coast dialect of American-English. All had typical hearing and typical speech-language development for their age, as determined by self-report in the adults and by parental report in the children. In addition, children passed a pure tone hearing screen at 1000, 2000, and 4000 Hz in each ear at 20 dB HL and had average to above average standard scores on the Peabody Picture Vocabulary Test (*M* = 125.75, *SD* = 11.24; Dunn & Dunn, 2007) and on the Recalling Sentences subtest (*M* = 13.54, *SD* = 2.04) from the CELF-4 (Semel, Wiig, & Secord, 2006) [1]. Half of the twelve 5-year-olds were female as were 7 of the twelve 8-year-olds and 7 of the 12 adults.

Stimuli

Sentence stimuli were designed to elicit the production of utterance medial noun phrases made up of a determiner and noun. Sentences were always 9 syllables in length.

[1] Note that these data were collected in 2013 before the CELF-5 was published.

The target noun phrase (NP) was always the direct object in the sentence. The preceding verb was in the 3rd person. The determiner was the definite article (i.e., "the"). The nouns were monosyllabic and had the same rhyme, but began either with a labial or a velar consonant (i.e., "bat" versus "cat"). Sentences with a third noun, "the rat," were also elicited, but are excluded from the study because of problems inherent to segmenting intervocalic liquids from adjacent vowels. The target NPs occurred in one of 3 metrical contexts: footed, unfooted, and ambiguous. The verb was monosyllabic in the footed context, disyllabic and trochaically-stressed in the unfooted context. Phrasal verbs made up of two monosyllabic words created the ambiguous context. There were 3 verbs per context. These were "hates", "hits", and "loves" in the footed context, "pushes", "watches", and "shushes" in the unfooted context, and "glares at", "sneers at", and "pleads with" in the ambiguous context. Although controlled for phonological and prosodic factors, the verbs were not controlled for lexical frequency. Instead, lexical frequency was statistically controlled (see Measurement and Analysis). The average normalized frequency for verbs used to create the footed context was 16.69 (*SD* = 11.15) in the spoken portion of the Corpus of Contemporary American English (COCA). This was higher than the average frequency of the verbs used to created the unfooted context (*M* = 3.71, *SD* = 3.37) or the frequency of phrasal verbs used to create the ambiguous context (*M* = .06, *SD* = .05). The verbs and nouns were crossed for a total of 18 sentences. Example are shown in Table 1 below.

Table 1 about here.

Elicitation Procedure

Sentences were recorded by a female speaker of west coast American English. Care was taken to produce each sentence under a single intonational contour. Multiple repetitions of each were produced, but only a single good rendition was excised from the recording and saved as its own audio file. The individual audio files were then randomized and aggregated to serve as stimuli in the elicitation task.

Child participants were introduced to the characters they would be talking about with cartoon pictures: a bat wearing a hat and a cat laying on a mat. During the task, each stimulus sentence was played in turn, and the participant repeated it back. Auditory presentation and repetition were used to control for age-dependent differences in reading level. The experimenter controlled the pace of sentence elicitation and provided feedback on productions. The experimenter would also replay a stimulus sentence to elicit a new production if a prior production was deemed errorful or disfluent. The whole set of sentences was elicited twice in random order to obtain 36 tokens for measurement per speaker (3 metrical contexts x 3 verbs per context x 2 nouns x 2 repetitions). Participants' speech was digitally recorded onto a Marantz PMD660 (with a sampling rate of 44,100 Hz) using a Shure ULXS4 standard wireless receiver and a lavaliere microphone, which was attached to a baseball hat or headband that the speaker wore. If the experimenter deemed that the child needed a break, the task was paused so that children could color on their sheet of paper or to have a drink of water or juice.

Error and Disfluency Coding

Since the task was embedded in a larger protocol, the experimenters often accepted productions that were later deemed disfluent and, in any case, a stimulus sentence was

only ever elicited twice before the experimenter moved on to the next sentence. For this reason, a number of sentences were produced with a pause or other disfluency that disrupted the prosodic structure of the sentence. The target noun was also sometimes not realized correctly. These sentences were excluded from acoustic analysis. The error and disfluency coding criteria are described below.

Post-V break. All participants, even the youngest, did a good job of imitating the intonation contour of the recorded model, which often included a pitch accent on the verb. Still, a number of participants sometimes produced long pauses (> 250 milliseconds) between the offset of the verb and the onset of the target NP. These pauses, perceived as disfluencies or hesitations, were coded as post-V breaks. Occasionally, participants produced the wrong NP after the verb and then corrected themselves by restarting the sentence with the correct NP. These restart repetitions were also coded as post-V breaks.

Noun substitution. Children sometimes substituted *bat* or *cat* with another noun (e.g., *hat* or *mat*) from the closed set that were being elicited. In most cases, experimenters caught the error and elicited a new production of the sentence in order to replace the errorful production. However, an experimenter would move on after a second incorrect repetition in order not to tire or frustrate the child participant. This meant that some sentences were never produced with the target noun. Incorrect noun production was coded as a noun substitution error.

Disfluent NP. Target NPs were sometimes produced either with "the" prolongation and/or with a pause that intervened between the determiner and noun. Prolongations and pauses were defined objectively within speaker based on outliers in the data such that all

tokens with schwa or stop closure duration values that were greater than 1.5 times the interquartile range for a particular speaker were coded as disfluent NP productions.

Post-NP break. Participants occasionally inserted a pause after the target noun and before the prepositional phrase. When this pause was equal to or greater than 250 milliseconds, it was identified as a break. A post-NP break meant that the target noun was in phrase-final position and the noun therefore subject to final lengthening. Accordingly, NPs from sentences with a post-NP break were excluded from the analyses of relative duration and amplitude.

Segmentation

Recordings were displayed both as oscillograms and spectrograms in Praat (Boersma & Weenink, 2013), the target NP was located in its verb context and its vowels segmented based on repeated listening, visible abrupt changes in the oscillogram, the presence of formant structure and periodicity. Stressed vowels in the monosyllabic nouns were typically produced in such a way that all visible cues were robustly present. Determiner vowels were often identified based on some subset of the cues. Figure 1 provides an example of highly reduced adult speech to illustrate the segmentation decisions made in the most difficult cases. Note that children's speech was less reduced, if also less crisply articulated.

Figure 1 about here.

The figure shows the oscillogram, spectrogram, and textgrid tiers associated with an adult's production of "the cat" in the sentence "The bad bat likes the cat on the mat."

Fricative energy from the third person marking on the verb can be seen at the beginning of the interval displayed in the spectrogram, followed by an abrupt change in the frequency distribution of the noisy energy due to the onset of the determiner. Schwa segmentation was based on the presence of periodicity in the waveform and formant structure on the spectrogram. Closure duration associated with /k/ production was marked separately from the offset of the schwa vowel to the moment of release, evident as an abrupt increase in energy in the oscillogram and across frequencies in the spectrogram. In this example, there was no clear closure associated with the consonantal offset in "cat." Instead, the speaker compressed the vocal folds to signal "t" using a glottal gesture. The speaker also shifted immediately from an [æ] configuration to a nasal consonant, thus eliding the vowel in the preposition. The section might thus be transcribed as [ðə.k^{h}æ̃.n̩], though this does not adequately capture the timing relations evident in Figure 1.

Segmentation reliability was assessed by randomly selecting recordings from 3 speakers per age group (= 25%) and segmenting anew (i.e., blindly) all target determiner and noun vowels in the nine files. The correlation between vowel durations extracted based on the original segmentation and those extracted from the blindly resegmented tokens was extremely high: $r(646) = .97$.

Measurement and Analysis

The dependent variables were acoustic correlates of reduction and determiner-noun coherence. The measured correlates of reduction included the absolute and relative duration and amplitude of the schwa in the determiner. Interval durations and amplitudes were extracted automatically based on the segmentations. Relative values were obtained by dividing values from the target noun by those from the determiner (i.e., N/DET).

The measured correlate of determiner-noun coherence was the influence of noun onset on schwa formant frequencies; that is, coherence was indexed as V-to-C coarticulation. All determiner and noun vowel formants were tracked using linear predictive coding (LPC). The maximum number of formants was set to 5, with the maximum formant frequency set to 6000 Hz in children's speech and to 5000 Hz in adults' speech. All tracks were visually inspected. If deemed accurate, the corresponding F1, F2, F3 values were automatically extracted at vowel midpoint. If the tracks were off, they were hand-corrected by adjusting either the maximum number of formants or maximum formant frequency before F1, F2, and F3 values were extracted.

Once extracted, formant values were normalized for vocal tract size using Thomas and Kendall's (2007) modified version of Syrdal and Gopal's (1986) bark difference metric. Specifically, formant values in Hertz were bark transformed using Traunmüller's (1997) formula. The transformed values were then used to derive measures indicating production along the front-back and height dimension. The value for the front-back dimension was obtained by subtracting bark transformed F2 (Z2) from bark transformed F3 (Z3); the value for the height dimension by subtracting bark transformed F1 (Z1) from Z3.

Linear mixed effect modeling was used to test for effects of age group, metrical context, noun onset and their interaction on the dependent variables. Speaker was treated as a random intercept with random slopes for verb frequency and for the number of repetitions per item. A compound symmetry covariance structure was used. The output of the analyses included ANOVA tables for the fixed effects in the model. The *F* values from those tables are reported here.

Results

Acoustic analyses were based on sentences that were produced fluently and correctly by the 36 speakers in the study: 5-year-olds produced 305 of these, 8-year-olds produced 337, and adults produced 401. The elicitation goal was 432 sentences per age group. The exclusion rate was therefore 29.4% in the 5-year-old group, 22% in the 8-year-old group, and 8.7% in the adult group. A binary regression analysis on excluded/included confirmed the apparent effect of age group on exclusions, $W(2) = 18.26$, $p < .001$, but found no effect of metrical context. The effect of metrical context was also not significant when the analyses were conducted by error type, but there was a significant metrical context by age group interaction on post-V breaks, $W(4) = 10.13$, $p = .038$. The interaction was due to different patterns in the younger and older children's data. Younger children introduced more breaks in the unfooted context compared to older children, who introduced more breaks in the footed context. When analyses were split by age group, the effect of metrical context on post-V breaks was found to be significant only in the older children's data, $W(2) = 6.75$, $p = .034$. This effect is opposite the metrical theoretic expectation that unstressed grammatical words are preferentially chunked with a preceding strong syllable. Table 2 summarizes all the data as a function of age group and context.

Table 2 about here.

Note that individual sentences were often produced with an error *and* a disfluency or with more than one disfluency. This means that the total number of errors and disfluencies produced was greater than the total number of sentences excluded from the analyses. Note

also that the effect of noun on exclusions was not included in the analyses because this factor indexes determiner-noun coherence; it does not condition the chunking pattern.

Reduction

Recall that both the chunking and articulatory hypotheses predict main effects of age and metrical context on grammatical word reduction, but only the chunking hypothesis predicts an interaction between these factors.

Schwa duration. As predicted, absolute schwa duration, measured in milliseconds, varied systematically with age, $F(2, 100) = 6.88$, $p < .001$, as is evident from the boxplots shown in Figure 2. Schwa duration was significantly longer in 5-year-old children's speech compared to adults' speech (mean difference = 14.93, $p = .006$) [2]. Mean schwa duration in 8-year-old productions were intermediate to those produced by 5-year-olds and adults (5-year-olds' $M = 61.79$, $SD = 18.30$; 8-year-olds' $M = 54.37$, $SD = 18.79$; adults' $M = 47.02$, $SD = 6.39$), and not significantly different from either.

Figure 2 about here.

The effect of metrical context on absolute schwa duration was also significant, $F(2, 896) = 6.77$, $p = .001$ (see Figure 2), but did not interact with age. Mean comparisons indicated that schwa durations were shorter in the footed context compared to the unfooted context (mean difference = −3.76, $p = .010$). Durations in the control ambiguous context were not significantly different from the other two contexts, but the data in Figure 2 suggests that

[2] Alpha was set at $p = .05$, which means that the Bonferroni corrected alpha is $p = .017$.

schwa durations in the ambiguous context were more similar to those in the footed context compared to the unfooted context.

Whereas the data on absolute duration are consistent with the metrical theoretic assumption that footed and unfooted grammatical words are realized differently, the data on relative schwa duration are not. In particular, the effect of metrical context was not significant in the analysis on relative schwa duration (see Figure 3). Instead, there was a strong effect of noun onset, $F(1, 866) = 91.71$, $p < .001$, due to substantially shorter [æ] after the voiceless stop in "cat" than after the voiced stop in "bat" ("bat" $M = 180.74$ msec., $SD = 31.55$; "cat" $M = 155.99$ msec., $SD = 34.83$). More interestingly, the effect of noun onset interacted with age, $F(2, 947) = 5.33$, $p = .005$. The simple effect of age was also significant, $F(2, 118) = 10.38$, $p = .002$. This effect was the same across metrical contexts, as is evident from the data shown in Figure 3.

Figure 3 about here.

Mean comparisons indicated a significant difference between 5-year-olds' and adults' productions of determiner + noun sequences (mean difference = −.72, $p = .012$). On average, the [æ] in "cat" and "bat" was 3 times as long the schwa in "the" in 5-year-olds' target NPs ($M = 3.12$; $SD = .77$), but nearly 4 times as long in adults' NPs ($M = 3.82$; $SD = .03$). The patterns in the 8-year-old data did not differ significantly from the patterns in either the 5-year-old or adult data. Taken together, the results on relative duration are consistent with the assumption that younger children do not reduce "the" to the same degree as adults.

Schwa amplitude. The analysis on absolute schwa amplitude indicated significant effects of age group, $F(2, 109) = 11.74$, $p < .001$, metrical context, $F(2, 876) = 4.91$, $p = .008$, and noun onset, $F(1, 935) = 81.47$, $p < .001$. None of the interactions between the factors were significant. As predicted, schwa was somewhat louder (> dB) in younger children's speech than in older children's and adults' speech (5-year-olds' $M = 72.95$, $SD = 3.68$; 8-year-olds' $M = 70.08$, $SD = 4.74$; adults' $M = 68.64$, $SD = 3.04$). It was also louder in the footed context compared to the unfooted and ambiguous contexts (strong $M = 71.03$, $SD = 4.21$; weak $M = 70.60$, $SD = 4.44$; ambiguous $M = 70.00$, $SD = 4.13$), and before the voiced onset compared to the voiceless onset ("bat" $M = 71.46$, $SD = 4.07$; "cat" $M = 69.62$, $SD = 4.27$). Only the effect of age was preserved in the analysis of relative amplitude, $F(2, 176) = 22.81$, $p < .001$, as is evident from the data shown in Figure 4.

_____

Figure 4 about here.

_____

Post hoc tests indicated that 5-year-olds' productions were significantly different from 8-year-old productions (mean difference = −.036, $p < .001$) and from adults' productions (mean difference = −.037, $p < .001$). Younger children often produced the schwa in "the" with slightly greater amplitude than the [æ] in "bat" and "cat": the mean relative amplitude of the two vowels (N/Det) was = 0.99 ($SD = .04$). In contrast, older children and adults almost always produced louder [æ] than schwa (8-year-olds, $M = 1.02$; $SD = .05$, and adults, $M = 1.02$; $SD = .03$). A larger noun-determiner ratio, which is equivalent to a soft-then-loud pattern, is what we would expect if "the" is reduced relative to the noun it determines. Thus, these results parallel those on relative duration. Both sets of results are consistent

with the assumption that 5-year-old children do not reduce "the" to the same degree as adults.

Coherence

When the target noun phrase (NP) is a prosodic word it will be more coherent (=coarticulated) than when a prosodic word boundary intervenes between the determiner and noun. The effect of noun onset provides an index of this coherence. Recall that only the chunking hypothesis predicts an interaction between metrical context and age on chunking, which is equivalent here to a 3-way interaction between noun onset, metrical context, and age group on schwa formant frequencies.

Vowel height. The analysis of normalized F1 (Z3-Z1) indicated significant effects of noun onset, $F(1, 984) = 33.65$, $p < .001$, age group, $F(2, 198) = 19.47$, $p < .001$, and the 2-way interaction between these factors, $F(2, 984) = 4.24$, $p = .015$. Younger children's production of the determiner vowel was more influenced by the adjacent noun onset than older children's and adults' productions. The relevant data are shown in Figure 5. Note that larger normalized F1 values (Z3-Z1) indicate lower absolute F1 values and so greater vocal tract closure.

Figure 5 about here.

Although the analysis indicated no simple effect of metrical context on schwa height, there was a significant noun onset by metrical context interaction, $F(2, 1024) = 3.22$, $p = .040$. The age group by context interaction was not significant, $F(4, 608) = 1.48$, p = NS, but there was a significant 3-way interactions between noun onset, metrical context, and age

group, $F(4, 1024) = 2.58$, $p = .036$. Figure 6 shows that this interaction was due to a larger effect of metrical context on V-to-C coarticulation in younger children's speech compared to older children and adults' speech. In particular, younger children's schwa height was more influenced by the adjacent noun onset in the footed and ambiguous metrical contexts than in the unfooted metrical context. Adult productions only varied systematically with noun onset in the ambiguous metrical context in adult speech. Older children's productions were not at all affects by context.

Figure 6 about here.

Overall, the results on normalized schwa F1 indicate greater coherence of determiner-noun sequences in younger children's speech than in adult speech. They also indicate greater coherence in the ambiguous context than in the unfooted context in younger children's and adults' speech. The determiner-noun sequence was also more coherent in the footed context than in the unfooted context in younger children's speech.

Vowel frontedness. Turning now to the front-back dimension, the analysis of normalized schwa F2 indicated effects of noun onset, $F(1, 928) = 1605.74$, $p < .001$, age group, $F(2, 124) = 12.71$, $p < .001$, and metrical context, $F(2, 882) = 21.72$, $p < .001$, on schwa production. Only two of the 2-way interactions were significant (noun onset x age group: $F(2, 927) = 58.22$, $p < .001$; noun onset x metrical context: $F(2, 933) = 4.37$, $p = .013$). The 3-way interaction was not significant.

Smaller Z3-Z2 values indicate that F2 is closer to F3; larger values indicate that it is farther. Thus, the effect of noun onset could indicate either that participants produced

more fronted versions of schwa before /k/ or that F2 and F3 were particularly close together in the /-Vk/ context, which would be consistent with a velar pinch configuration. Analysis of mean bark values for F2 and F3 support both interpretations: all participants produced schwa with much higher Z2 values before "cat" than before "bat," $F(1, 164) = 514.07$, $p < .001$, and significantly lower Z3 values before "cat" than before "bat", $F(1, 164) = 13.96$, $p < .001$.

The 2-way interaction between noun onset and age reflected stronger V-to-C effects on the frontedness of adults' schwa compared to children's, as shown in Figure 7. The 2-way interaction between noun onset and metrical context was less striking. It nonetheless reflected weaker V-to-C effects on frontedness in the unfooted context compared to the footed and ambiguous context, contrary to expectations based on a metrical theoretic approach to prosodic word formation. The mean difference in normalized F2 before "bat" versus "cat" was 1.37 (*SD* = .07) in the unfooted context, 1.48 (*SD* = .07) in the footed context, and 1.56 (*SD* = .06) in the ambiguous context. The relevant data are shown in Figure 8.

Figures 7 and 8 about here.

Overall, the results on normalized schwa F2 suggest stronger coherence of the determiner-noun sequences in adults' speech compared to children's speech regardless of metrical context, and somewhat stronger coherence of determiner-noun sequences in the footed and ambiguous contexts compared to the unfooted context regardless of age.

Discussion

The study findings confirm Allen and Hawkin's (1978) perceptually-based observation that children do not reduce grammatical words to the same extent as adults. Compared to adults, children produced "the" in target noun phrases (NPs) with longer and louder vowels than adults in prosodic contexts where reduction is expected. In addition, the stressed and unstressed vowels in the target NPs were nearer in duration and amplitude in younger children's speech compared to older children's and adults' speech, consistent with previous findings (see Sirsa & Redford, 2011). Assuming these effects of age on grammatical word production generalize to spontaneously produced speech, the stress-based rhythm pattern of English will be measurably less pronounced in younger children's speech compared to older children's and adults' speech. This implication is consistent with the working hypothesis that persistent age-related rhythm differences are due to differences in how children and adults produce grammatical words in running speech.

Grammatical word reduction is a phonological phenomenon that references prosodic words in English. Prosodic words are units of production: chunks delimited in the speech plan. The current study investigated metrical context effects on the production of "the" to determine whether children's longer and louder grammatical words are due to age-related differences in how speech is chunked for output (i.e., the chunking hypothesis) or to children's slower, larger amplitude, and less tightly coordinated speech movements (i.e., articulatory timing hypothesis). The chunking hypothesis predicted an interaction between the fixed effects of age group and metrical context on schwa duration and amplitude, and an interaction between age group and metrical context on determiner-noun coherence, measured as the effect of noun onset on schwa formant frequencies. The articulatory

timing hypothesis did not. Overall, the results suggest that age-related differences in grammatical word production likely stem both from immature timing control and from differences in chunking.

Timing control

Consistent with the articulatory timing hypothesis, there was a strong effect of age on schwa duration and amplitude, but little effect of metrical context on the measures and no interaction between age and metrical context.

Although absolute schwa duration was shortest in the footed context, metrical context had no effect on relative schwa duration. Taken together, these two results suggest that speakers produced the deteriminer-noun sequences more quickly in the footed context than in the unfooted context. Is this a prosodic effect? Maybe, but not necessarily.

One possibility is that the verb itself impacted the speed with which determiner-noun sequences were produced. The footed context was created using higher frequency verbs than those used to create the unfooted and ambiguous contexts. Higher frequency items have stronger (more entrenched) speech plan representations than lower frequency items (see, e.g., Pluymaekers, Ernestus, & Baayen, 2005), which could facilitate not only that item's production but also the planning and production of adjacent items.

Of course, this explanation is largely undercut in the current study. Frequency effects were statistically controlled: it was entered as a within speaker random factor (i.e., random slope) in the mixed effects model. Also, the results in the ambiguous context were not statistically different from the footed context. Moreover, the phrasal verbs used to create the ambiguous context were not only less frequent than those used to create the footed context, they were also less frequent than those used to create the unfooted context.

Although likely not relevant to the results of the current study, lexical frequency effects are worth keeping in mind when interpreting the results from developmental studies on rhythm production. In particular, phonotactic nor lexical frequency may provide an alternative explanation to prosodic structure for syllable omission in young children's speech. Afterall, these effects has well known effects on other aspects of children's lexical and phonological acquisition (see Curtin & Zamuner, 2014, for a review) as well as on their production of newly presented words (e.g., Richtsmeier, Gerken, Goffman, & Hogan, 2009). Effects of high-frequency versus low-frequency phonotactics on children's productions are especially well documented (Munson, 2001; Storkel, 2001; 2003; Edwards, Beckman, & Munson, 2004; Zamuner, Gerken, & Hammond, 2005).

Another possible explanation for the effect of metrical context on absolute duration in the present study is has to do with sentence-level rhythmicities. The footed context extended a strong-weak alternation set up at sentence onset with "Maddy"—the two syllable proper name in subject position; the unfooted context disrupted this pattern. Extended alternations of strong and weak syllables generate a rhythm pattern that may help speakers better coordinate articulatory movements. Greater synchronic control over movement coordination will allow sequential motor goals to be attained more quickly. On this view, the metrical structure of the specific chunks executed is irrelevant to the speed of execution. What matters is the extended rhythm pattern of a stretch of speech that is planned for execution.

Following this line of thinking further, it is worth noting that the effect of metrical context on absolute amplitude was opposite the expectation based on a metrical theoretic view of chunking. Grammatical words in the footed context were louder in absolute terms

than those produced in the unfooted context. Importantly, this main effect would appear to be consistent with Goffman and colleagues' finding that movement amplitudes are *less* reduced in a footed context compared to an unfooted context (Goffman & Malin, 1999; Goffman, 2004; Goffman et al., 2006). Thus, the finding lends further credibility to the argument that shorter grammatical word durations in the footed context arise from an overall increase in articulatory rate rather than from metrically-conditioned differences in chunking. It also casts doubt on the assumption that a metrical theoretic approach to prosodic word formation is relevant to speech planning. This doubt is amplified by the results on V-to-C coarticulation.

Chunking

The effects of metrical context on V-to-C coarticulation, our measure of coherence, were consistently opposite to those expected based on metrical theory. Recall that metrical theory requires unstressed grammatical words to be prosodified (= chunked) with the preceding word when footed, and according to syntactic constituency when unfooted. Here, when metrical context effects were significant, the determiner was found to cohere more tightly with the noun it modified in the footed context than in the unfooted context. This result may underscore the importance of syntactic constituency for chunking. Still, the duration results suggest an alternative explanation: greater determiner-noun coherence in the footed context may have been an epiphenomenon of the faster rates at which these sequences were produced. If this explanation is correct, then—like the results on schwa duration and amplitude—metrical context effects on schwa formant structure may be better explained with reference to lower-level speech production factors than to speech plan structure.

Consider, for example, the finding that schwa was least fronted in the unfooted metrical context. This context was created using verbs with stem final palatal-alveolar consonants. These consonants are argued to have stronger "degree of articulatory" constraints (DAC) than, for example, dentals and vowels (Recasens, Pallarès, & Fontdevila, 1997; Recasens, 2015). Coarticulatory influence of one segment on adjacent segments is directly proportional to DAC strength. If we assume that "the coarticulatory fields for specific phonetic segments may be resized depending on prosodic condition (Recasens, 2015:1422)," coupled with the very weak DAC values of schwa, then it is possible that the perseveratory effects of /ʃ/ on schwa in "the" account for the effect of metrical context on F2 values in the present experiment. In other words, perseveratory effects of the phonotactic environment along with articulation rate likely provide a better explanation than chunking for the interaction between metrical context and noun onset on schwa formant structure.

Overall, though, the effects of age on schwa formant structure were stronger than the effects of metrical context. For example, younger children's vowels were more closed overall than adults' vowels (Figure 6). This result could suggest either stronger V-to-C effects in children's speech compared to adults' speech or weaker V-to-V effects, since [æ] is more open than [ə]. The possibility that the result indicates stronger V-to-C effects is undermined by the direction of the effect of target onset on F1. If we assume that children were producing [k] as a dorsal consonant. In children's speech, [ə] was more closed (Z3-Z1 was larger) before [k] than before [b]. This is opposite of the coarticulatory effects we might expect based on adult data, which show that voiceless velar stops are produced with

a more open jaw position than voiced bilabial stops (Keating, Lindblom, Lubker, & Kreiman, 1994).

Overall, analyses of normalized F1 and F2 indicate different patterns of interaction between noun and age group. A limitation of the present study is that the design does not allow us to account for this difference. One possibility is that even school-age children have better control over jaw positioning than tongue positioning (see, e.g., Sorensen, Toutios, Goldstein, & Narayanan, 2017), and thus poorer control over the rapid movements required for consonantal articulation compared to vowel articulation. Slower lingual articulation could in turn bias children towards stronger V-to-V coarticulation. Measurement and analysis of formant frequencies associated with [æ] in the following noun showed that only children's production of the determiner vowel were overlapped with their production of [æ] in the front-back dimension, albeit only before "cat. This result, though consistent with stronger V-to-V coarticulation in children's speech compared to adults' speech, is hardly definitive. To test for age-related differences in V-to-V coarticulation the upcoming vowel must be manipulated. Only the upcoming consonant was manipulated in the current study.

Conclusions and future directions

The hypothesis that immature articulatory timing control helps to explain age-related differences in the realization of grammatical words is strongly supported in the present study. Of course, given the extensive work on the slow development of speech motor skills, it would be surprising if the hypothesis were not supported. For this reason, we deem the articulatory timing hypothesis of significantly less theoretical interest than the hypothesis that children and adults chunk speech differently for output. Yet, the chunking hypothesis

is not well supported in the present study. Our conclusion is that the hypothesis needs further reticulation and testing. This need dovetails with another that is more a comment on the state of the field rather than a specific limitation of the present study: we currently lack a sophisticated theory of speech production; one that incorporates what has been learned about the very protracted development speech motor skill with what we know of language acquisition. Lisa Goffman's work on the influences of metrical structure and syntax on movement patterns comes closest in our estimation to providing the relevant insights for building such a theory, but further work is needed if we are to detail a psycholinguistic model of speech production that can be used to understand development and disorder. Until such a model is articulated, those of us working on child speech and language will be obligated to borrow from established linguistic and psycholinguistic theories to understand children's speech patterns. The results of the present study demonstrate that this borrowing is inappropriate. Mainstream theories of production, even the most elegant ones, are built entirely from observations of adult behavior. Moreover, the behaviors of interest are rarely phonetic in nature.

The present study represents a first step towards acquiring the type of data that is useful for model building. In particular, the results suggest that careful study of long-distance anticipatory coarticulation in children's speech can provide important insights into how the speech plan is structured at different stages of development. Still, patterns of anticipatory coarticulation are only able to indicate whether a particular item is chunked or not with an upcoming one. In the case where it is not, there is no way to distinguish between the possibilities that it has been chunked with a preceding item or that it is planned and produced as its own prosodic word. Duration and amplitude measures

certainly provide some insight, but it is likely that unstressed items will always be relatively short in context. In future work, we will also examine patterns of perseveratory coarticulation to supplement our understanding of planned production units in child and adult speech.

Acknowledgments

This work was supported in part by the Eunice Kennedy Shriver National Institute of Child Health & Human Development under grants R01HD061458 and R01HD087452, and in part by a fellowship from the European Institutes for Advanced Study (EURIAS), co-funded by the European Commission (Marie-Sklodowska-Curie Actions COFUND Programme FP7) with administrative and further financial support provided by IMéRA at Aix-Marseille Université. The content is solely the author's responsibility and does not necessarily reflect the views of her sponsors.

REFERENCES

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

Barry, W., Andreeva, B., & Koreman, J. (2009). Do rhythm measures reflect perceived rhythm? *Phonetica*, 66(1-2), 78-94.

Boersma, P. & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.59, retrieved from http://www.praat.org/.

Bunta, F., & Ingram, D. (2007). The acquisition of speech rhythm by bilingual Spanish-and English-speaking 4-and 5-year-old children. *Journal of Speech, Language, and Hearing Research*, 50(4), 999-1014.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, 10(2), 159-199.

Curtin, S., & Zamuner, T. S. (2014). Understanding the developing sound system: interactions between sounds and words. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(5), 589-602.

Demuth, K. (2001). Prosodic constraints on morphological development. In J. Weissenborn & B. Höhle (eds.), *Approaches to Bootstrapping: Phonological, Syntactic and Neurophysiological Aspects of Early Language Acquisition* (pp. 3-21). Amsterdam: John Benjamins.

Demuth, K., & Fee, E. J. (1995). *Minimal words in early phonological development*. Ms., Brown University and Dalhousie University.

Dunn, D. M., & Dunn, L. M. (2007). *Peabody picture vocabulary test: Manual*. Pearson.

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, 47(2), 421-436.

Fikkert, P. (1994). *On the Acquisition o f Prosodic Structure*. PhD thesis, University of Leiden, The Netherlands.

Fourakis, M. (1991). Tempo, stress, and vowel reduction in American English. *Journal of the Acoustical Society of America*, 90(4), 1816-1827.

Fowler, C. A. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech, Language, and Hearing Research*, 24(1), 127-139.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*, 63, 223–230.

Gay, T. (1981). Mechanisms in the control of speaking rate. *Phonetica*, 27, 44–56.

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

Gerken, L., Landau, B., & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology*, 26(2), 204.

Goffman, L. (2004). Kinematic differentiation of prosodic categories in normal and disordered language development. *Journal of Speech, Language, and Hearing Research*, 47(5), 1088-1102.

Goffman, L., Gerken, L., & Lucchesi, J. (2007). Relations between segmental and motor variability in prosodically complex nonword sequences. *Journal of Speech, Language, and Hearing Research*, 50(2), 444-458.

Goffman, L., Heisler, L., & Chakraborty, R. (2006). Mapping of prosodic structure onto words and phrases in children's and adults' speech production. *Language and Cognitive Processes*, 21(1-3), 25-47.

Goffman, L., & Malin, C. (1999). Metrical effects on speech movements in children and adults. *Journal of Speech, Language, and Hearing Research*, 42(4), 1003-1015.

Goffman, L., & Westover, S. (2013). Interactivity in prosodic representations in children. *Journal of Child Language*, 40(5), 1032-1056.

Grabe, E., Post, B., & Watson, I. (1999, August). The acquisition of rhythmic patterns in English and French. In *Proceedings of the International Congress of Phonetic Sciences. International Congress of Phonetic Sciences* (Vol. 1999, pp. 1201-1204).

Hayes, B. (1982). Extrametricality and English stress. *Linguistic inquiry*, 13(2), 227-276.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.

Keating, P. A., Lindblom, B., Lubker, J., & Kreiman, J. (1994). Variability in jaw height for segments in English and Swedish VCVs. *Journal of Phonetics*, 22(4), 407-422.

Kehoe, M., Stoel-Gammon, C. & Buder, E.H. (1995). Acoustic correlates of stress in young children's speech. *Journal of Speech and Hearing Research*, 38, 338-350.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, 105(3), 1455-1468.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1-38.

Ling, L. E., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 43(4), 377-401.

Ma, L., Perrier, P., & Dang, J. (2015). Strength of syllabic influences on articulation in Mandarin Chinese and French: Insights from a motor control approach. *Journal of Phonetics*, 53, 101-124.

McCabe, P. C., & Meller, P. J. (2004). The relationship between language and social competence: How language impairment affects social growth. *Psychology in the Schools*, 41(3), 313-321.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73-97.

Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research*, 44(4), 778-792.

Nijland, L., Maassen, B., Meulen, S. V. D., Gabreëls, F., Kraaimaat, F. W., & Schreuder, R. (2002). Coarticulation patterns in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 16(6), 461-483.

Nittrouer, S., Studdert-Kennedy, M., & Neely, S. T. (1996). How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables. *Journal of Speech, Language, and Hearing Research*, 39(2), 379-389.

Noiray, A., Cathiard, M. A., Abry, C., & Ménard, L. (2010). Lip rounding anticipatory control: Crosslinguistically lawful and ontogenetically attuned. In B. Maassen & P. van Lieshout (eds.) *Speech Motor Control: New Developments in Basic and Applied Research* (pp. 153-171). OUP.

Olejarczuk, P., & Redford, M. A. (2013, June). The relative contribution of rhythm, intonation and lexical information to the perception of prosodic disorder. In Proceedings of Meetings on Acoustics ICA2013 (Vol. 19, No. 1, p. 060154). ASA.

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 35(2), 205-220.

Paul, R., Shriberg, L. D., McSweeny, J., Cicchetti, D., Klin, A., & Volkmar, F. (2005). Brief report: Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 35(6), 861.

Payne, E., Post, B., Astruc, L., Prieto, P., & Vanrell, M. D. M. (2012). Measuring child rhythm. *Language and Speech*, 55(2), 203-229.

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

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*, 118(4), 2561-2569.

Pollock, K.E., Brammer, D.M., & Hageman, C.F. (1993). An acoustic analysis of young children's productions of word stress. *Journal of Phonetics*, 21, 183–203.

Polyanskaya, L., & Ordin, M. (2015). Acquisition of speech rhythm in first language. *Journal of the Acoustical Society of America*, 138, EL199-EL204.

Recasens, D. (2015). The Effect of Stress and Speech Rate on Vowel Coarticulation in Catalan Vowel–Consonant–Vowel Sequences. *Journal of Speech, Language, and Hearing Research*, 58(5), 1407-1424.

Recasens, D., Pallarès, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America*, 102(1), 544-561.

Redford, M. A. (2014). The perceived clarity of children's speech varies as a function of their default articulation rate. *Journal of the Acoustical Society of America*, 135(5), 2952-2963.

Redford, M. A. (2015). Unifying speech and language in a developmentally sensitive model of production. *Journal of Phonetics*, 53, 141-152.

Redford, M. A., Kapatsinski, V., & Cornell-Fabiano, J. (2017). Lay Listener Classification and evaluation of typical and atypical children's speech. *Language and Speech*, https://doi.org/10.1177/0023830917717758.

Redford, M. A., & Oh, G. E. (2016). Children's abstraction and generalization of English lexical stress patterns. *Journal of Child Language*, 43(02), 338-365.

Redford, M. A., & Oh, G. E. (2017). The representation and execution of articulatory timing in first and second language acquisition. *Journal of Phonetics,* 63, 127-138. https://doi.org/10.1016/j.wocn.2017.01.004.

Repp, B. H. (1986). Some observations on the development of anticipatory coarticulation. *Journal of the Acoustical Society of America*, 79(5), 1616-1619.

Richtsmeier, P. T., Gerken, L., Goffman, L., & Hogan, T. (2009). Statistical frequency in perception affects children's lexical production. *Cognition*, 111(3), 372-377.

Riely, R. R., & Smith, A. (2003). Speech movements do not scale by orofacial structure size. *Journal of Applied Physiology*, 94(6), 2119-2126.

Schwartz, R. G., Petinou, K., Goffman, L., Lazowski, G., & Cartusciello, C. (1996). Young children's production of syllable stress: An acoustic analysis. *Journal of the Acoustical Society of America*, 99(5), 3192-3200.

Selkirk, E. (1996). The prosodic structure of function words. In J.L. Morgan & K. Demuth (eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, (pp. 187-214). Mahwah, NJ: Erlbaum.

Semel, E. M., Wiig, E. H., & Secord, W. (2006). *CELF 4: Clinical Evaluation of Language Fundamentals*. Pearson: Psychological Corporation.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25(2), 193-247.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research*, 44(5), 1097-1115.

Sirsa, H., & Redford, M. A. (2011, August). Towards understanding the protracted acquisition of English rhythm. In *Proceedings of the International Congress of Phonetic Sciences. International Congress of Phonetic Sciences* (Vol. 2011, pp. 1862-1865). NIH Public Access.

Smith, A., & Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research*, 41(1), 18-30.

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

Sorensen, T., Toutios, A., Goldstein, L., & Narayanan, S. (2017). Tracking developmental changes in articulatory strategy during childhood. *Journal of the Acoustical Society of America*, 142(4), 2584.

Sternberg, S., Knoll, R. L., Monsell, S., & Wright, C. E. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica*, 45(2-4), 175-197.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, 44(6), 1321-1337.

Storkel, H. L. (2003). Learning new words II: Phonotactic probability in verb learning. *Journal of Speech, Language, and Hearing Research*, 46(6), 1312-1323.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, 79(4), 1086-1100.

Thomas, E.R. & Kendall, T. 2007. NORM: The vowel normalization and plotting suite. [Online Resource: http://ncslaap.lib.ncsu.edu/tools/norm/]

Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *Journal of the Acoustical Society of America*, 134(1), 628-639.

Traunmüller, H. (1997). Auditory scales of frequency representation. [Online: http://www.ling.su.se/staff/hartmut/bark.htm]

Walsh, B., & Smith, A. (2002). Articulatory Movements in AdolescentsEvidence for Protracted Development of Speech Motor Control Processes. *Journal of Speech, Language, and Hearing Research*, 45(6), 1119-1133.

Weismer, G., & Martin, R. (1992). Acoustic and perceptual approaches to the study of intelligibility. IN R.D. Kent (ed.), *Intelligibility in speech disorders*, (pp. 67-118). John Benjamins.

Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37(3), 356-381.

Wheeldon, L. R., & Lahiri, A. (2002). The minimal unit of phonological encoding: prosodic or lexical word. *Cognition*, 85(2), B31-B41.

Wijnen, F., Krikhaar, E., & Den Os, E. (1994). The (non) realization of unstressed elements in children's utterances: Evidence for a rhythmic constraint. *Journal of Child Language*, 21(01), 59-83.

Zamuner, T. S., Gerken, L., & Hammond, M. (2004). Phonotactic probabilities in young children's speech production. *Journal of Child Language*, 31(3), 515-536.

Table 1. Example stimulus sentences show how metrical context and target noun were varied while sentence length and the location of the target noun phrase (underlined) were held constant.

METRICAL CONTEXT	EXAMPLE SENTENCE	TARGET ONSET
footed	The fat cat **hits** the bat in the hat.	labial
	The bad bat **hits** the cat on the mat.	velar
unfooted	The cat **pushes** the bat in the hat.	labial
	The bat **pushes** the cat on the mat.	velar
ambiguous	The cat **glares at** the bat in the hat.	labial
	The bat **glares at** the cat on the mat.	velar

Table 2. Number of elicited sentences with a disqualifying error (= noun substitution) or disfluency after the verb (post-V break), within the noun phrase (disfluent NP), or between the noun phrase and prepositional phrase (post-NP break). The disqualifying errors/disfluencies are shown as a function of age group and metrical context. Note that a single sentence could contain an error and one or more disfluencies. This means that the total number of sentences excluded was less than the total number of errors and disfluencies produced.

		METRICAL CONTEXT			
AGE GROUP	CATEGORY	FOOTED	UNFOOTED	AMBIGUOUS	TOTALS
5-year-olds	post-V break	13	26	16	55
	N substitution	8	11	16	35
	disfluent NP	26	16	24	66
	post-NP break	10	9	8	27
8-year-olds	post-V break	14	4	6	24
	N substitution	1	6	6	13
	disfluent NP	21	21	27	69
	post-NP break	9	6	12	27
Adults	post-V break	0	0	1	1
	N substitution	1	0	0	1
	disfluent NP	10	10	10	30
	post-NP break	2	3	1	6

Figure Legends

Figure 1. Example segmentation of highly reduced speech produced by an adult participant. The section shows the interval from the offset of the verb to the beginning of the prepositional phrase in the target sentence, "The bad bat likes the cat on the mat". The first tier codes the verb (= metrical context), target determiner-noun sequence (DP context), and repetition number. The second tier is used to identify pauses between the noun and following preposition that would imply phrase-final position. The third tier shows the segmentation of the determiner schwa and stressed [æ] in the noun (i.e., acoustics Vs). The fourth tier marks out the silent interval between the determiner and noun; within speaker outliers in the duration of this interval were used to identify pause breaks between the determiner and noun.

Figure 2. Boxplots show the dispersion of absolute schwa durations around the median by age group and metrical context (footed, unfooted, and ambiguous). Both effects were significant, but the factors did not interact. The dashed reference line indicates the mean absolute duration of schwa in the adult data.

Figure 3. Boxplots show the dispersion of relative schwa durations around the median by age group and metrical context (footed, unfooted, and ambiguous). Only the effect of age group was significant. The dashed reference line indicates the mean relative duration of schwa in the adult data.

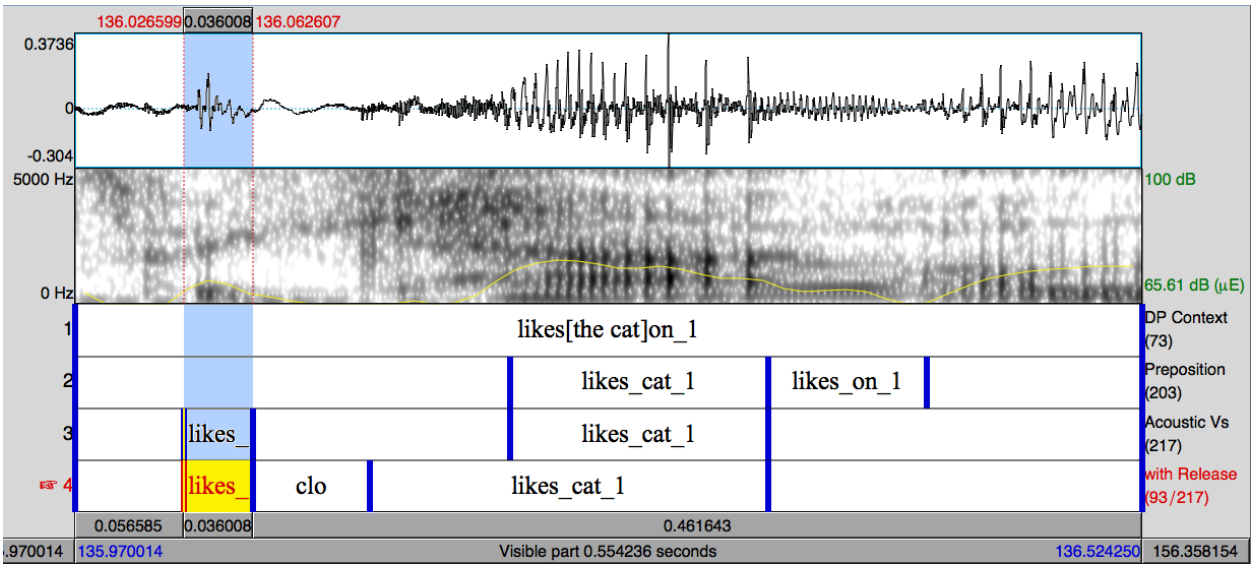
Figure 4. Boxplots show the dispersion of relative schwa amplitudes around the median by age group and metrical context (footed, unfooted, and ambiguous). Only the effect of age group was significant. The dashed reference line indicates the mean relative amplitude of schwa in the adult data.

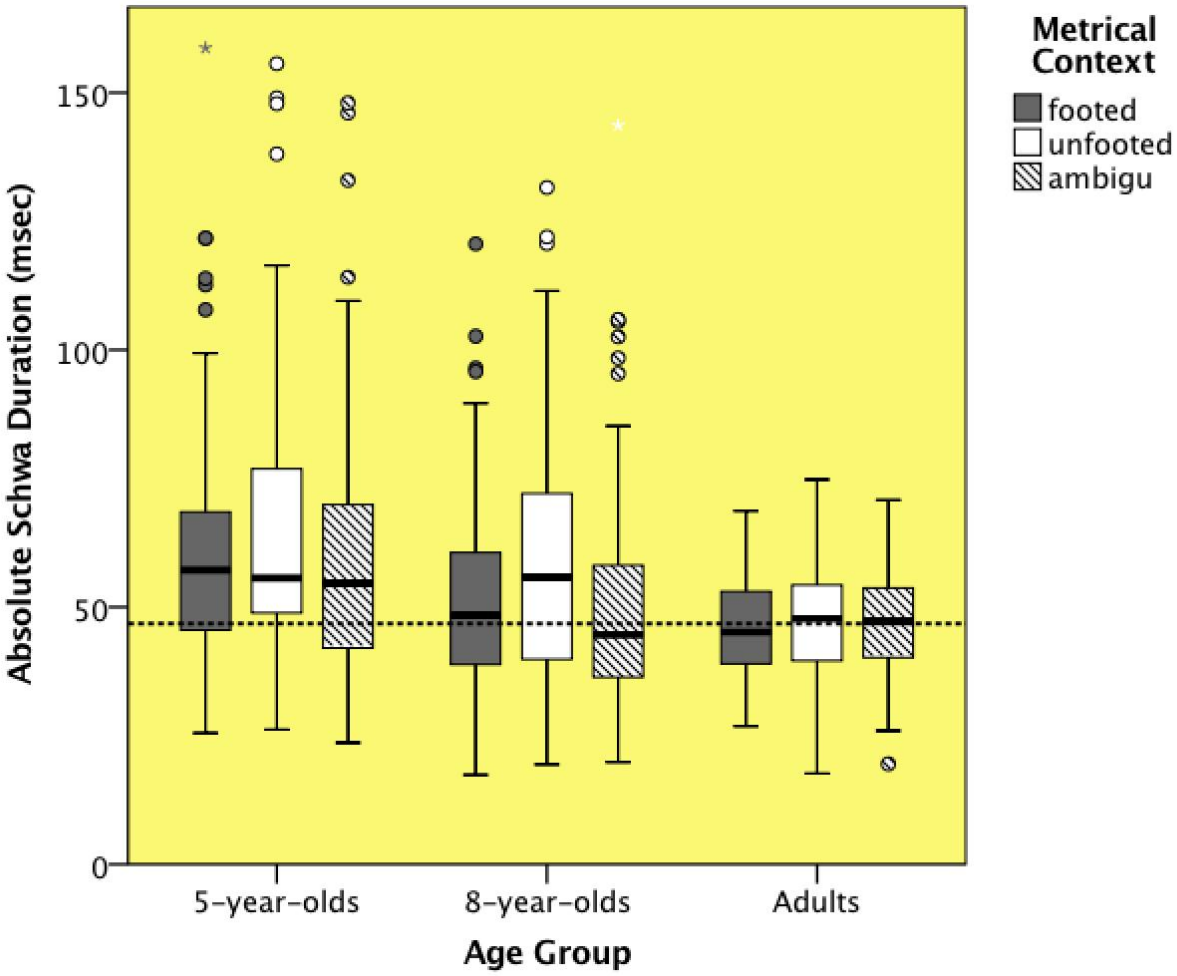
Figure 5. Boxplots show the dispersion of normalized F1 values for schwa around the median by noun onset and age group. The interaction between these two factors was significant. The dashed reference line indicates the mean value of normalized F1 in the adult data.

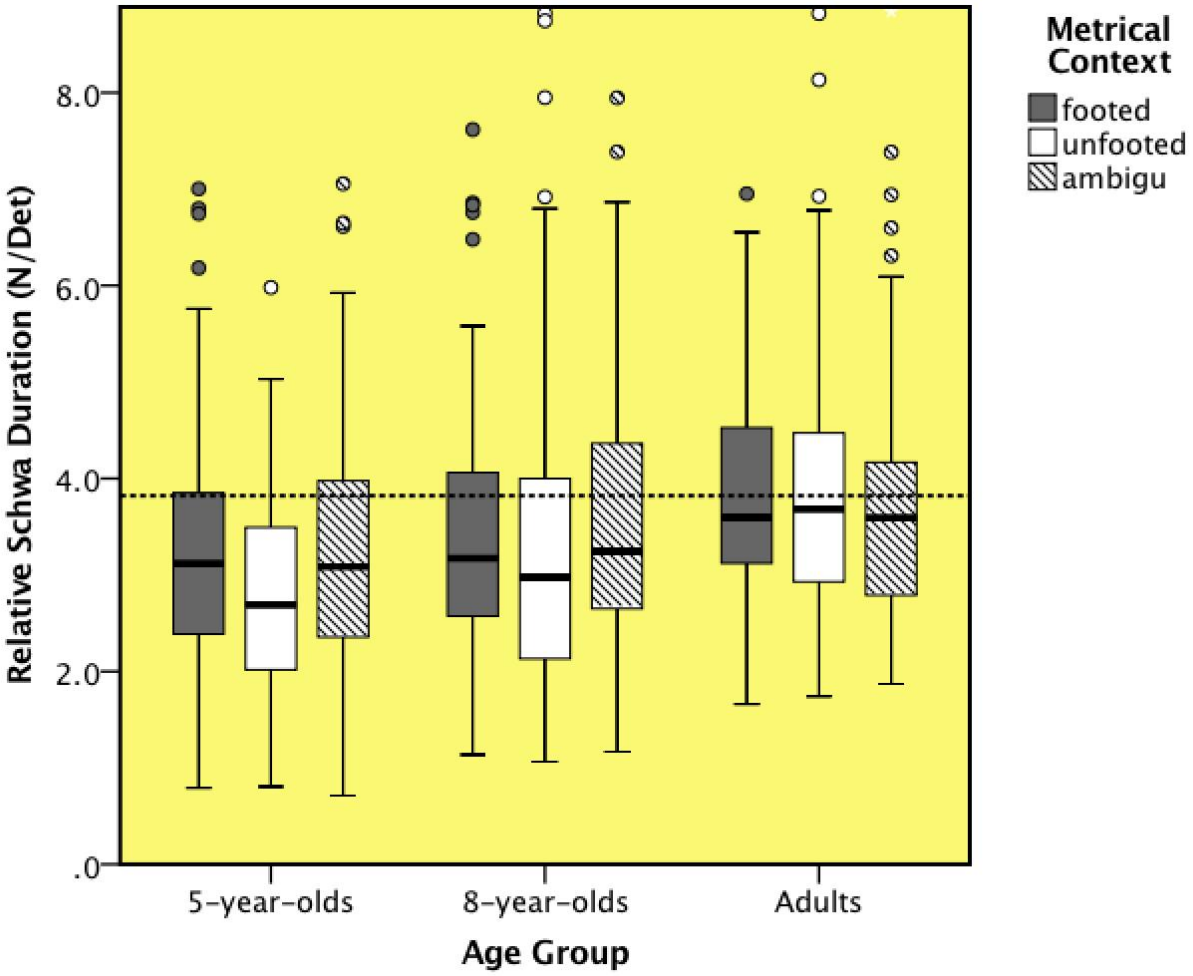
Figure 6. The significant 3-way interaction between noun onset, age group, and metrical context on normalized F1 values for schwa is shown. The grey dashed line and solid black line show the mean values for "cat" and "bat", respectively. The error bars indicate the 95% confidence interval.

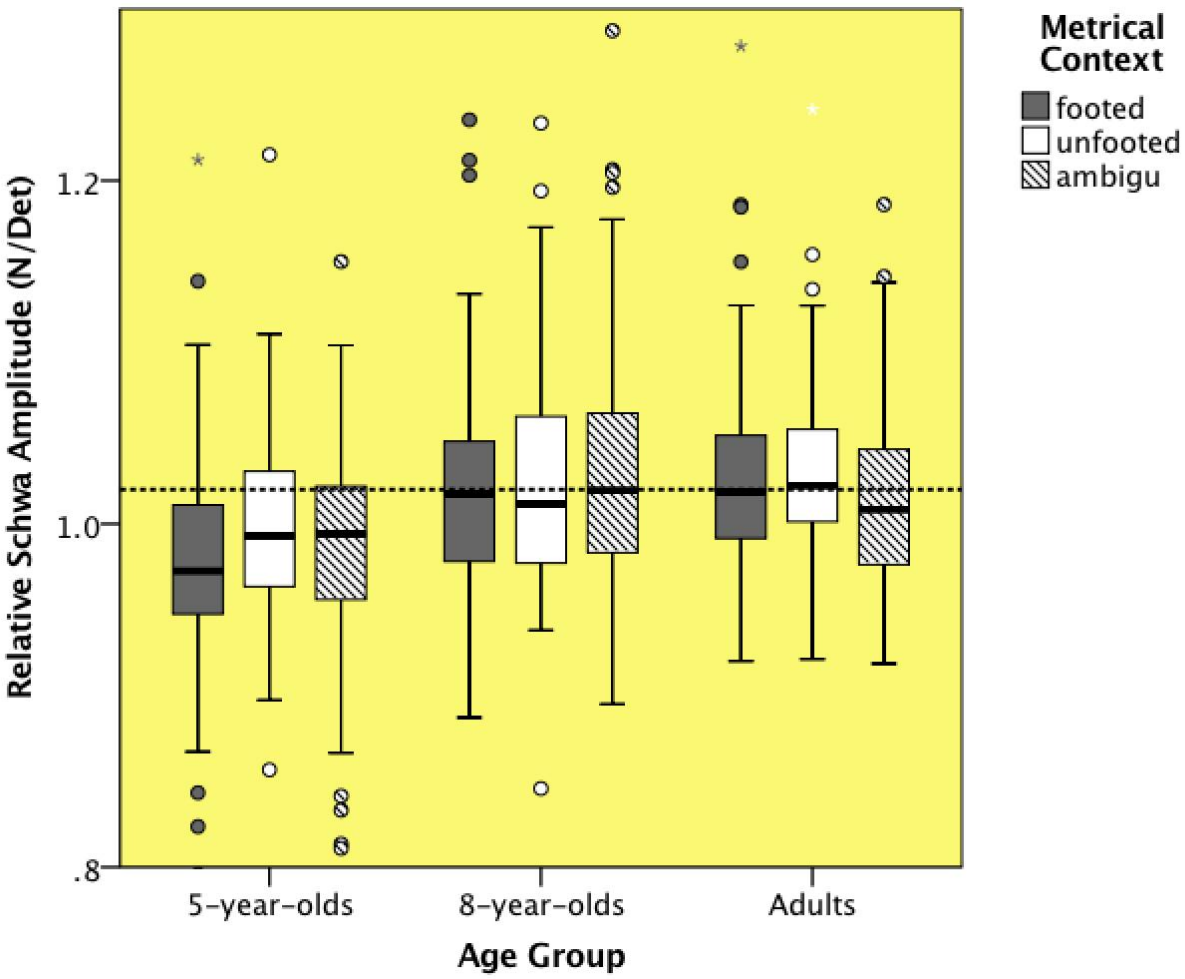
Figure 7. Boxplots show the dispersion of normalized F2 values for schwa around the median by noun onset and age group. The interaction between these two factors was significant. The dashed reference line indicates the mean value of normalized F2 in the adult data.

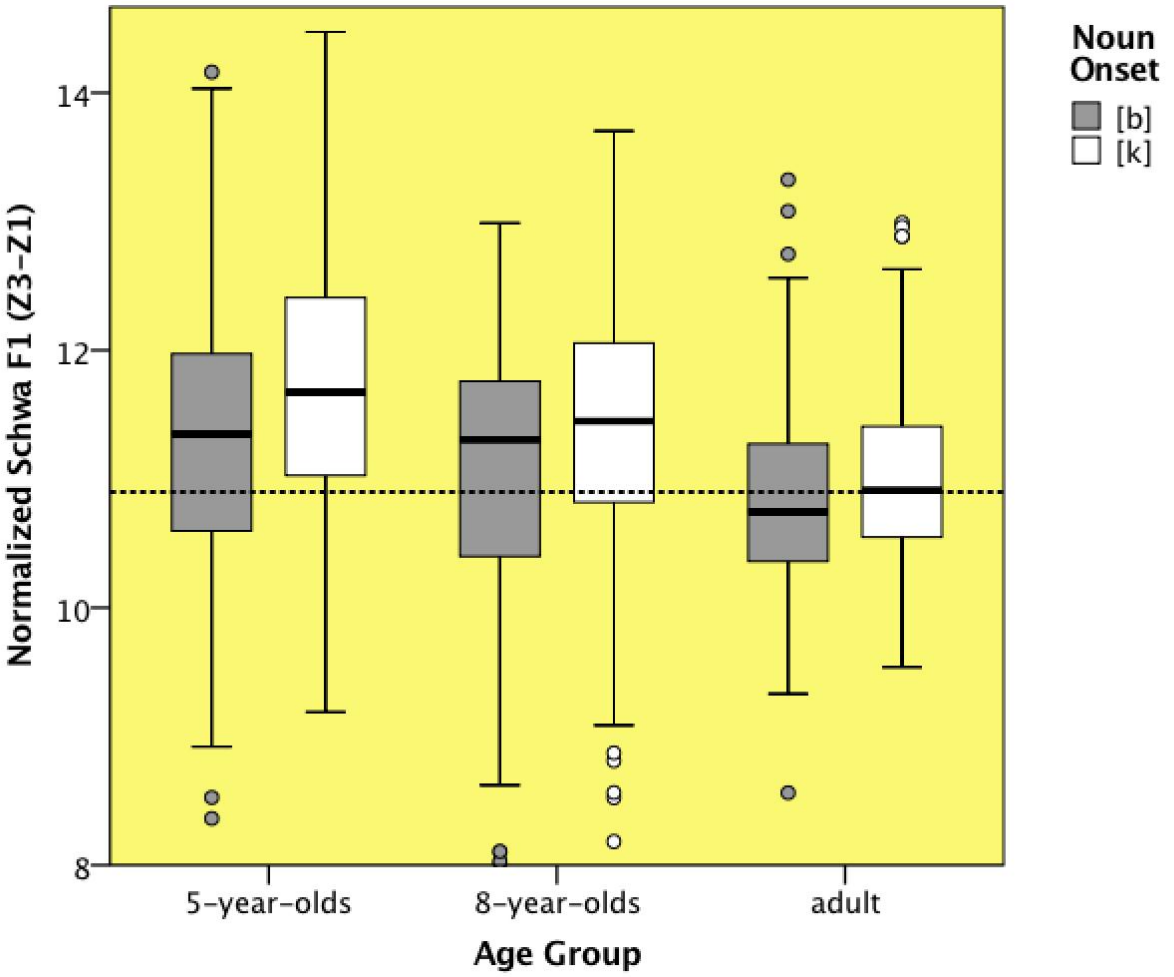
Figure 8. Boxplots show the dispersion of normalized F2 values for schwa around the median by noun onset and metrical context. The interaction between these two factors was significant. The dashed reference line indicates the mean value of normalized F2 in the ambiguous metrical context.

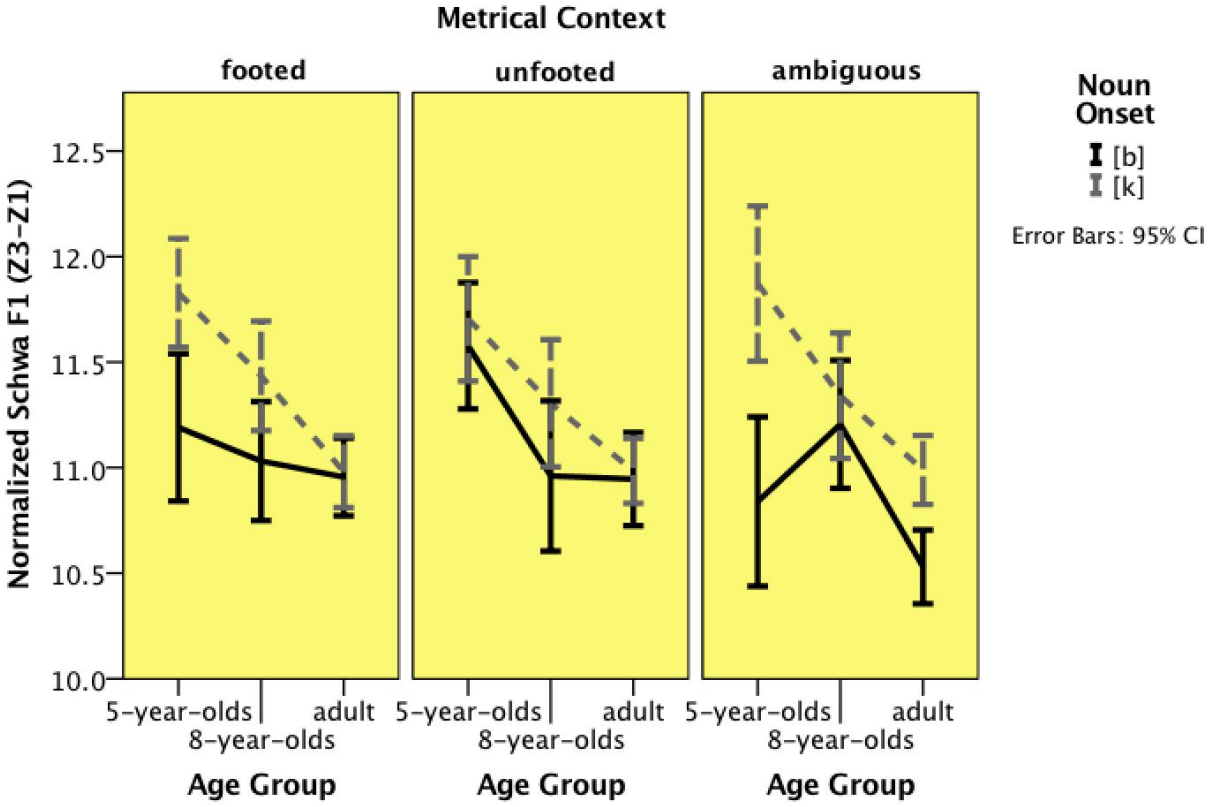

136.026599
0.036008
136.062607
0.3736
0
-0.304
5000 Hz
0 Hz
100 dB
65.61 dB (µE)
1
likes[the cat]on_1
DP Context (73)
2
likes_cat_1
likes_on_1
Preposition (203)
3
likes_
likes_cat_1
Acoustic Vs (217)
4
likes_
clo
likes_cat_1
with Release (93/217)
0.056585
0.036008
0.461643
.970014
135.970014
Visible part 0.554236 seconds
136.524250
156.358154

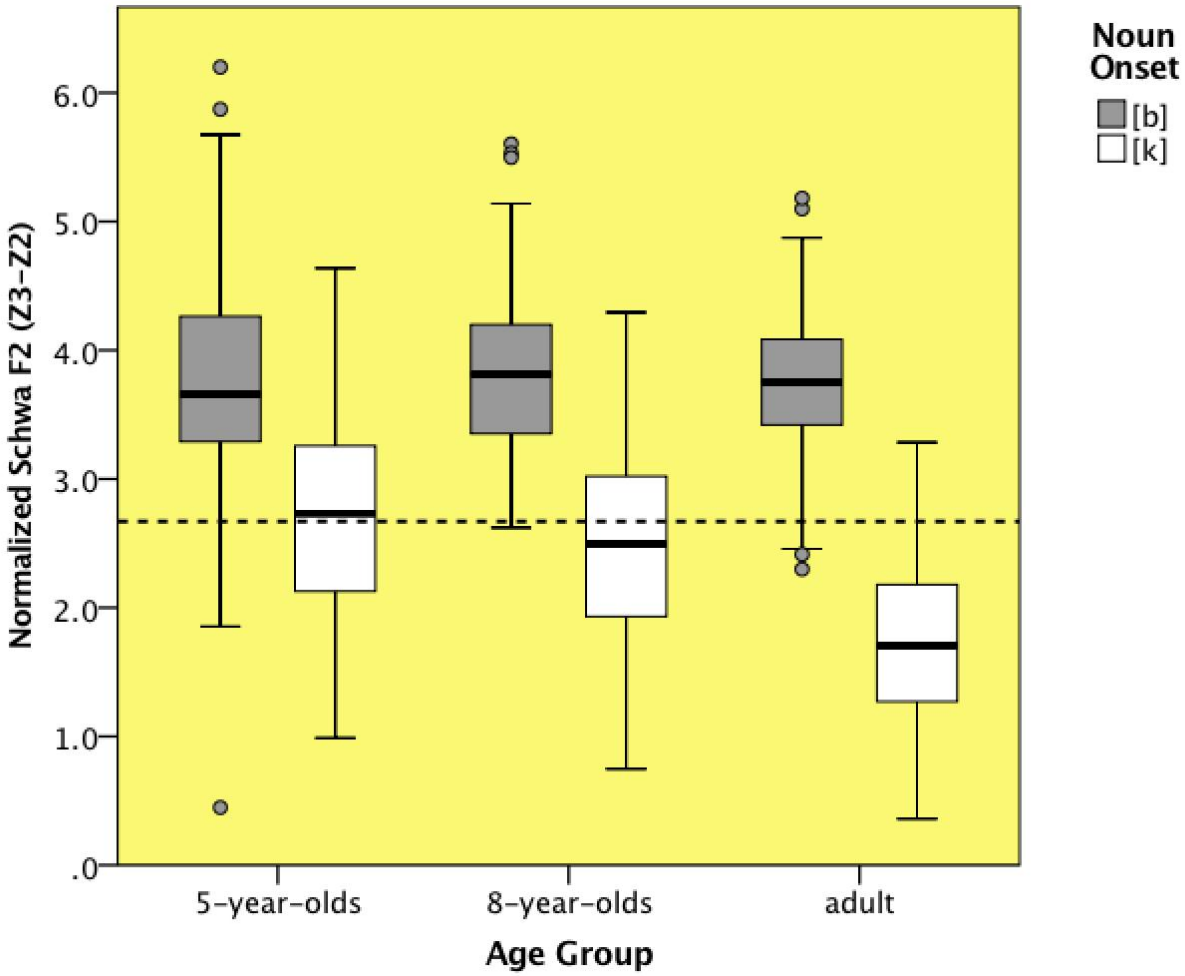

Metrical Context
footed
unfooted
ambigu
Absolute Schwa Duration (msec)
150
100
50
0
5-year-olds
8-year-olds
Adults
Age Group

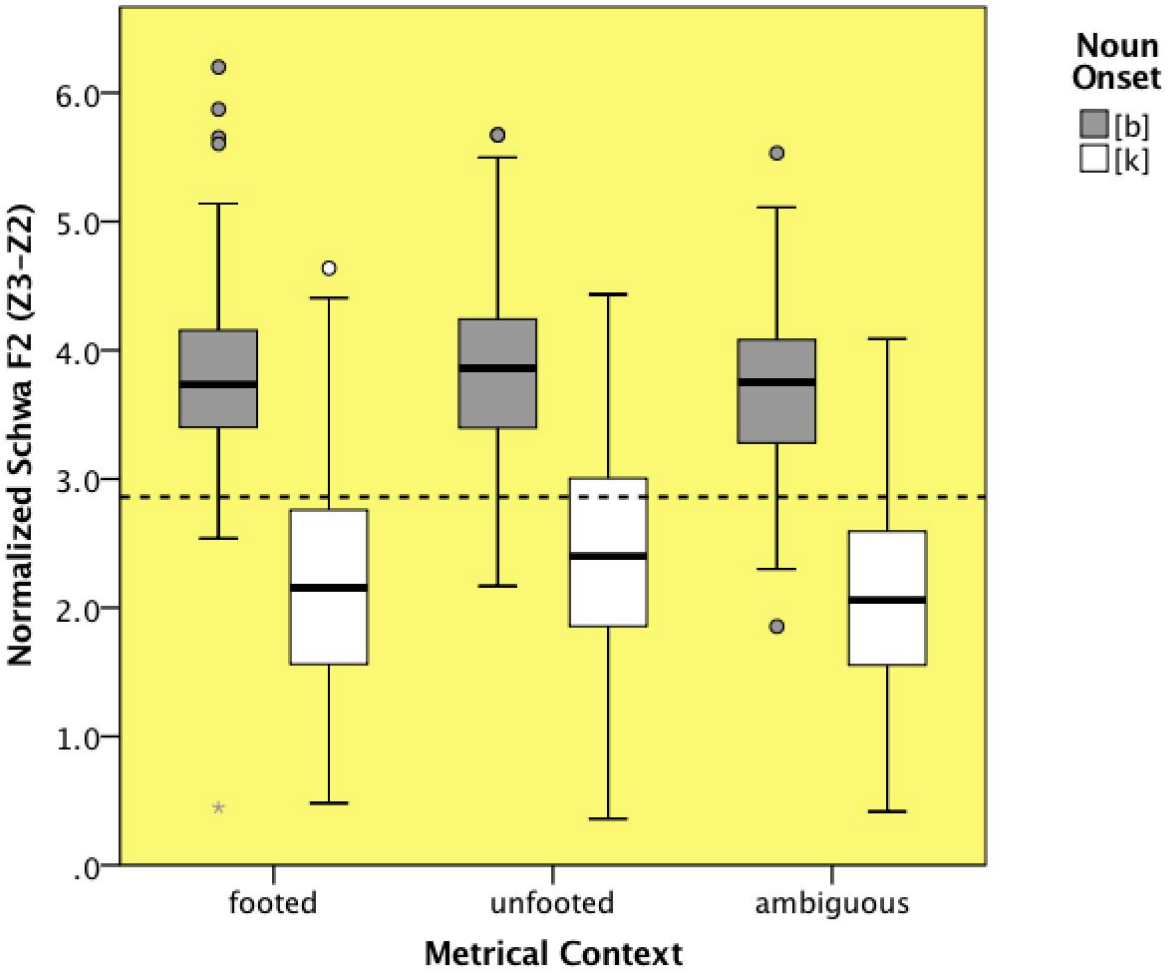

Relative Schwa Duration (N/Det)
8.0
6.0
4.0
2.0
.0
5-year-olds
8-year-olds
Adults
Age Group
Metrical Context
footed
unfooted
ambigu


Metrical Context
footed
unfooted
ambigu
Relative Schwa Amplitude (N/Det)
1.2
1.0
.8
5-year-olds
8-year-olds
Adults
Age Group


Noun Onset
[b]
[k]
Normalized Schwa F1 (Z3-Z1)
14
12
10
8
5-year-olds
8-year-olds
adult
Age Group


Metrical Context
footed
unfooted
ambiguous
Noun Onset
[b]
[k]
Error Bars: 95% CI
Normalized Schwa F1 (Z3-Z1)
12.5
12.0
11.5
11.0
10.5
10.0
5-year-olds
8-year-olds
adult
Age Group


Noun Onset
[b]
[k]
Normalized Schwa F2 (Z3-Z2)
6.0
5.0
4.0
3.0
2.0
1.0
.0
5-year-olds
8-year-olds
adult
Age Group


Noun Onset
[b]
[k]
Normalized Schwa F2 (Z3-Z2)
6.0
5.0
4.0
3.0
2.0
1.0
.0
footed
unfooted
ambiguous
Metrical Context