



Coarticulatory effects on “the” production in child and adult speech

Jeffrey E. Kallay, Melissa A. Redford

University of Oregon, Eugene, Oregon, USA
jkallay@uoregon.edu, redford@uoregon.edu

Abstract

If prosodic words are the principle units of speech planning and production, then the production of an unstressed grammatical word should be especially influenced by the adjacent context word with which it is chunked. The current study tested this prediction in child and adult speech to investigate the development of the speech plan. Anticipatory and perseveratory influences on determiner vowel production were investigated in simple SVO sentences produced by 5-year-old children and college-aged adults. Although children’s productions indicated greater perseveratory influences than adults’ productions, anticipatory effects were consistently stronger than perseveratory effects across age groups. The results suggest that, by age 5 years, children chunk determiners along morphosyntactic boundaries just like adults.

Index Terms: chunking, coarticulation, development

1. Introduction

Psycholinguistic theory assumes that prosodic words are the principle units of speech planning and production [1-3]. As units of production, we expect the component sounds and syllables of a prosodic word to cohere more tightly compared to sounds and syllables that are planned as separate production units. In production, coherence manifests as coarticulation. Thus, a grammatical word that is chunked with an adjacent content word for production should be more coarticulated with that adjacent word than one that is chunked as an independent prosodic unit.

In theory, coarticulatory patterns can allow us to assess whether a grammatical word is the initial or final unstressed syllable of a prosodic word: when a grammatical word is chunked with a subsequent content word, and so is the initial syllable of the prosodic word, it will be especially influenced by the shape of that content word compared to when it is chunked with a previous content word. This prediction is consistent with the view that anticipatory coarticulation is planned (e.g., [4]). If the grammatical word is chunked with a preceding syllable, and so is the final syllable of the prosodic word, perseveratory influences may be especially strong and anticipatory ones fairly weak. Here, we test these predictions in child and adult speech by investigating the acoustic signatures of anticipatory and perseveratory coarticulation on determiner (*the*) production. The investigation is motivated by an interest in the development of speech planning processes. These processes are especially opaque in development in part because children’s speech has been alternatively described as more, less, or equally coarticulated as adults’ speech (see, e.g., [5-9]).

The present study extends our prior work on grammatical word production in child and adult speech to investigate coarticulatory effects in more detail. Redford [10] found that, relative to the determined noun, 5-year-old children’s productions of *the* were longer and louder than adults’ productions. There were also age-related differences in the

effect that a subsequent noun onset had on the production of a determiner vowel. In particular, children’s determiner vowels varied more along the F1 dimension than the F2 dimension as a function of noun onset. The opposite was true of adults’ determiner vowels. While this finding provides some suggestion of age-related coarticulation differences, the study had several limitations that should be addressed if we are to use coarticulation to understand the development of speech chunking. First, the study did not investigate perseveratory effects, and so provided limited information about the direction of chunking. Second, the study only investigated V-to-C effects and not V-to-V effects, thus providing limited information about unit coherence. Finally, F1 and F2 were analyzed separately. These analyses indicated age-related differences in coarticulatory patterns, but little information about the effect of age on overall coarticulatory strength. The current study addresses these limitations by investigating both perseveratory and anticipatory V-to-V influences on determiner vowel production in simple SVO sentences. The target determiner, *the*, was in object position and modified a monosyllabic noun. The preceding verb was also monosyllabic. The nuclei of both the verb and noun were varied. The onset of the object noun was also varied to test for previously found V-C anticipatory effects on determiner production. Finally, overall coarticulatory strength was estimated as the mean Euclidian distance in normalized F1 x F2 space between determiner vowels in different phonological contexts, where strength is indexed by greater distances between determiner vowels as a function of context.

2. Methods

2.1. Participants

A total of 20 speakers participated in the study: 11 American-English speaking 5-year-old children (6 female) and 9 American English-speaking college-aged adults (4 female). All had typical hearing, as determined with a pure tone hearing screen at 1000, 2000, and 4000 Hz in each ear at 25 dB HL, and typical speech-language development, as determined by self or parental report for undergraduate students and children, respectively. All participants were recruited from the broader community in Eugene, Oregon, and spoke a west coast dialect of American English.

2.2. Stimuli

Participants produced sentences designed to elicit utterance-medial *the* in the context of two real monosyllabic content words: a verb preceding *the* and a noun following. The content word vowels were either /æ/ or /oʊ/. Noun onsets were /b/, /s/, /g/. Noun offsets were voiceless stop consonants. Each sentence began with the proper name *Maddy*. Example sentences are as follows:

<i>Maddy packs <u>the</u> bat.</i>	<i>Maddy pokes <u>the</u> bat.</i>
<i>Maddy packs <u>the</u> boat.</i>	<i>Maddy pokes <u>the</u> boat.</i>
<i>Maddy packs <u>the</u> sack.</i>	<i>Maddy pokes <u>the</u> sack.</i>
<i>Maddy packs <u>the</u> soap.</i>	<i>Maddy pokes <u>the</u> soap.</i>

2.3. Procedure

A child or adult speaker sat across from an experimenter in a quiet observation room. The experimenter first introduced a set of picture cards to the speaker. The pictures corresponded to each of the 6 nouns (*bat/boat, sack/soap, gak/goat*), which were to be elicited in a *pack* or *poke* frame sentence. A picture of a green cartoon figure was used for the nonce word *gak*. Elicitation was blocked by verb frame. Nouns were randomized within the block by shuffling the pictures. Each noun was elicited 6 times in each of the 2 verb frames. During the first few elicitation blocks, experimenters would produce the whole target sentence, introduce a pause by counting to three, then ask the speaker to produce the sentence. Once the speaker was comfortable with what they were to produce, the experimenter would simply provide the verb frame at the beginning of a block and then show each picture card to elicit the appropriate sentence. If the experimenter detected a disfluency or error during production, the card was placed to the back of the stack and the sentence re-elicited.

Elicitations were audio-visually (AV) recorded for later analysis. Speakers wore a hat with a Shure ULX wireless microphone. The wireless receiver input to a Panasonic HC-V770 camcorder.

2.4. Segmentation and Measurement

Audio tracks were exported from the AV files for coding, segmentation, and measurement in Praat [11]. Recall that if the experimenter detected a disfluency or error during elicitation, the sentence was re-elicited. Still, experimenters were sometimes less attuned to the speech elicited than to getting through the entire protocol and so sentences that were produced disfluently or with an error were first labeled so that they could be excluded from acoustic analyses. Disfluencies were defined as noticeable pauses between the verb and determiner or between the determiner and noun, unusually elongated words, and unusually slow productions. Errors were typically word or sound substitutions. Elicitations that included yelling, laughing, or background noise were also labeled and excluded. Roughly 12% of the data were excluded in this way. A total of 699 out of a possible 792 child sentences were segmented for acoustic analysis. A total of 583 out of a possible 648 adult sentences were segmented for analysis.

All vowels in the medial verb-*the*-noun sequence were 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.

Following segmentation, F1 and F2 measurements were extracted semi-automatically. For each vowel, the spectrogram was displayed so that the researcher could inspect the formant tracks. If it did not appear that the formants were being tracked accurately, formant settings were changed and the tracks redrawn. Once formant tracking was accurate, measurements were extracted from 10 evenly-spaced time points within the vowel interval. The analyses reported here focus on the determiner vowel; more specifically, on average F1 and F2 values calculated from the middle 5 measurement points.

2.5. Estimating Coarticulatory Strength

Coarticulatory strength was estimated by calculating the Euclidian distance in F1 x F2 space for determiners in each of the anticipatory and perseveratory contexts. This space was normalized across speakers by using the min-max scaling

procedure [12]. The F1 and F2 measurements for each age group (5-year-olds and adults) were placed on a 0-100 scale, with 0 corresponding to the lowest value and 100 the highest value. We then calculated Euclidian distance between the mean determiner vowel locations at each level of a given context for each speaker. For example, in the anticipatory V-V context we calculated the by-speaker average Euclidian distance between the determiner vowel in the /æ/ noun context and /oʊ/ noun context. The Euclidian distance measurements for the 3-level noun onset context were also calculated for each of the pairwise combinations (i.e., /b/-/g/, /b/-/s/, /s/-/g/).

2.6. Statistical Analyses

Child and adult data were split and linear mixed effects modeling was used to assess the effects of context on determiner F1 and F2 values. The lme4 package in R was used for this purpose [13,14]. The fixed effects were noun onset, noun vowel, and verb. Speaker was entered as a random effect. Model comparison was used to test the significance of each of the fixed effects on the formant values. For this, a full model was constructed with all of the fixed and random effects. The full model was then compared to reduced models in which each of the random factors was removed in turn. The resulting χ^2 statistic and p -value are reported here for significant effects within each age group.

Independent samples t-tests were used to test for age-related group differences in coarticulatory strength.

3. Results

3.1. Context Effects on Determiner Vowel Production

Recall that *the* was produced in the context of two monosyllabic content words; namely, a verb and a noun. The stressed vowel in the verb and noun vowel was either /æ/ or /oʊ/ so that perseveratory and anticipatory V-to-V effects on *the* production could be investigated. V-to-C anticipatory effects were investigated by varying noun onsets. Perseveratory effects are presented first. In all figures, the dark boxes represent the adult speakers and the light boxes the child speakers.

3.1.1. Perseveratory V-V Coarticulation

The determiner vowel did not vary systematically with the preceding verb in adults' speech, but the effect of verb on children's *the* was significant for both F1 ($\chi^2 = 4.84, p = .03$) and F2 ($\chi^2 = 7.96, p < .01$), as shown in Figure 1. These results indicate perseveratory effects of the preceding vowel on the determiner vowel in children's speech.

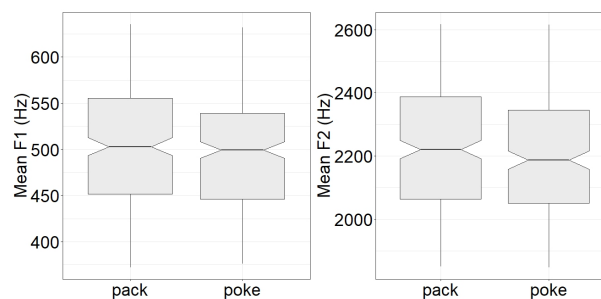


Figure 1: *V-to-V* perseveratory effects on determiner vowel F1 (left) and F2 (right) in children's speech. Perseveratory effects on determiner vowel production were not significant in adults' speech.

Children's *the* F1 was lower after *poke* than after *pack*, suggesting a more closed vocal tract configuration for the determiner following the mid vowel /oʊ/ compared to the low vowel /æ/. Children's *the* F2 was higher after *pack* than after *poke*, consistent with a more advanced tongue position for the determiner after the front vowel /æ/ compared to the back vowel /oʊ/. Note that, though statistically significant, the mean differences by verb context are very small and not substantially different than those observed for adult speakers. Mean F1 and F2 values are given in Table 1 by speaker and verb context.

Table 1: Mean (SD) F1 and F2 in adults' and children's determiner vowels by verb vowel context.

		Verb /æ/	Verb /oʊ/
F1	Adult	391.85 (42.80)	388.20 (42.70)
	Child	505.56 (65.16)	495.66 (63.43)
F2	Adult	1646.28 (155.41)	1638.63 (154.15)
	Child	2229.49 (206.32)	2197.60 (194.28)

3.1.2. Anticipatory V-V Coarticulation

The second set of analyses tested for effects of the noun vowel on the determiner vowel F1 and F2. Mixed effects modeling indicated significant effects on the noun vowel on F1 in adults' speech ($\chi^2 = 24.62, p < .001$), but not in children's speech. In contrast, noun vowel had a significant effect on F2 in both adults' ($\chi^2 = 47.35, p < .001$) and children's speech ($\chi^2 = 12.91, p < .001$). These results indicate anticipatory V-V coarticulation in both groups, with more reliable effects in adults' speech. The results are shown in Figure 2.

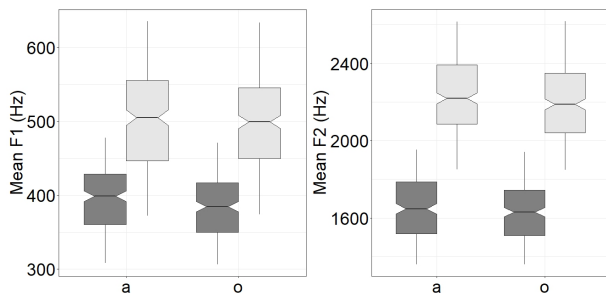


Figure 2: V-to-V anticipatory effects on determiner vowel F1 (left) and F2 (right) in adults' (dark boxes) and children's (light boxes) speech.

Adults produced the determiner vowel with a more open vocal tract configuration before nouns with the low vowel /æ/ than before those with the mid vowel /oʊ/. Both groups tended to produce the determiner vowel with a more advanced tongue position before nouns with the front vowel /æ/ than with the back vowel /oʊ/. Mean F1 and F2 values are given in Table 2 by speaker and noun vowel.

Table 2: Mean (SD) F1 and F2 in adults' and children's determiner vowels by noun vowel context.

		Noun /æ/	Noun /oʊ/
F1	Adult	395.87 (42.49)	384.40 (42.32)
	Child	503.25 (66.81)	498.36 (62.15)
F2	Adult	1652.82 (163.83)	1631.69 (144.03)
	Child	2238.50 (199.73)	2192.53 (200.58)

3.1.3. Anticipatory V-C Coarticulation

The final set of analyses tested for effects of the noun onset (/b/, /g/, or /s/) on F1 and F2 of the determiner vowel in adult and child speech. Mixed effects modeling indicated significant effects of onset on both F1 ($\chi^2 = 37.81, p < .001$) and F2 ($\chi^2 = 194.87, p < .001$) in children's speech. Likewise, effects of onset were also significant on F1 ($\chi^2 = 88.65, p < .001$) and F2 ($\chi^2 = 595.29, p < .001$) in adults' speech. These results are shown in Figure 3.

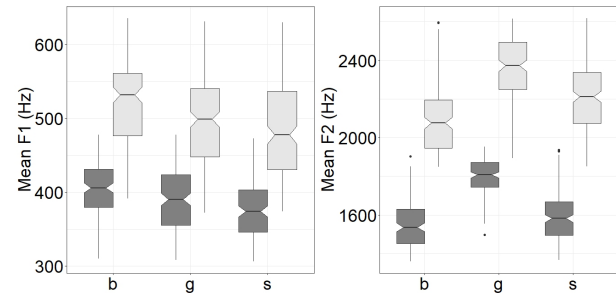


Figure 3: V-to-C anticipatory effects on determiner vowel F1 (left) and F2 (right) in adults' (dark boxes) and children's (light boxes) speech.

Both adults and children produced the determiner vowel with a more open vocal tract before /b/ than before /s/. F1 values in the /g/ context were intermediate between those in the /b/ and /s/ contexts. The values associated with F2 suggest that both groups of speakers produced *the* with a more retracted tongue position before /b/ than before /g/, and a more advanced tongue position before /s/ than before /b/. The surprising difference between /s/ and /g/ could suggest either that the velar was produced as a palatal-alveolar or that anticipatory coarticulation was blocked to some extent by /s/ [see 15]. Mean F1 and F2 values are given in Table 3 by speaker group and noun onset context.

Table 3: Mean (SD) F1 and F2 in adults' and children's determiner vowels by noun onset context.

		Noun /b/	Noun /g/	Noun /s/
F1	Adult	403.13 (40.87)	389.59 (42.67)	376.94 (40.86)
	Child	521.21 (58.96)	498.41 (62.42)	486.25 (66.89)
F2	Adult	1550.22 (123.19)	1802.40 (88.18)	1592.64 (118.62)
	Child	2092.22 (170.78)	2347.36 (179.30)	2211.55 (175.31)

3.2. Age Effects on Coarticulatory Strength

The preceding results suggest some group differences in the production of determiner vowels by context. In spite of this, independent 2-group Mann-Whitney U tests on our age-normalized measure of coarticulatory strength revealed no significant differences between children's and adults' productions. Of course, Mann-Whitney U is a conservative test. It was chosen here based on speaker sample sizes, which were relatively small and so not normally distributed. Perhaps with larger sample sizes, child and adult differences in coarticulatory strength would emerge. After all, the mean values presented in Table 4 indicate a trend towards such a difference, at least for the V-to-V contexts.

Table 4: Mean (SD) age-normalized distance between determiner vowels (i.e., context 1 – context 2) is shown as a function of context type and speakers' age group.

	Verb /æ/-/oʊ/	Noun /æ/-/oʊ/	Noun /b/-/g/	Noun /b/-/s/	Noun /s/-/g/
Adult	1.82 (1.53)	4.43 (2.07)	18.32 (5.59)	6.30 (2.35)	13.30 (3.88)
Child	2.52 (1.69)	5.82 (2.84)	17.55 (6.57)	7.94 (4.47)	11.18 (5.49)

The values shown in Table 4 also indicate that anticipatory effects on *the* production were consistently stronger than perseveratory effects for both groups of speakers. Noun onsets also had a greater effect on *the* production than noun vowels. These results are consistent with the by-group analyses on F1 and F2 presented in the preceding sections.

4. Discussion

Results from the current study indicate subtle differences in how children and adults coarticulate an unstressed grammatical word with adjacent content words. Only children's *the* productions indicated perseveratory influences, though these influences were fairly weak (see Figure 1). Anticipatory influences on *the* production were much stronger. Moreover, these influences appeared equally strong across speaker groups, if perhaps more systematic in adults' speech (see section 3.1.2.). Taken together, the results suggest that both children and adults chunked *the* with the noun it determined. Note that this chunking pattern suggests a weak-strong prosodic word unit that is in-line with morphosyntactic structure, if not with the rhythmic preferences of English.

We regard our conclusion about chunking as tentative since it relies on coarticulatory patterns that may have been influenced by other factors in the present study. For example, the anticipatory context varied more significantly than the perseveratory context, which may have biased speakers towards noun-focused productions. Moreover, the verb was always in third person and so produced with a final /s/. In so far as the alveolar fricative has a strong degree of coarticulatory resistance, it could have blocked perseveratory effects (see [15]). The fact that these effects were still significant in children's speech might indicate an implicitly learned sensitivity to prosodic boundaries or poorly defined boundaries. For example, it could be that adults actively inhibit inertial movement of articulators at the right-most edge of a prosodic unit and that children either lack the skill to do this or that their plans have less well-defined prosodic boundaries than adults'. Future work will use more perfectly symmetrical phonological contexts to rigorously investigate prosodically-related developmental changes in anticipatory and perseveratory coarticulation.

Acknowledgements

The authors are grateful to Julia Conway, Aubrianne Carson, and Jill Potratz for help with data collection. The work reported here was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) under grant R01HD087452. The content is solely the authors' responsibility and does not necessarily reflect the views of NICHD.

References

- [1] W.J.M. Levelt, A. Roelofs, and A.S. Meyer. "A theory of lexical access in speech production," *Behavioral and Brain Sciences* vol. 22, pp. 1-75, 1999.
- [2] L. Wheeldon and A. Lahiri. "Prosodic units in speech production," *Journal of Memory and Language*, vol. 37, no. 2, pp. 356-381, 1997.
- [3] L. Wheeldon and A. Lahiri. "The minimal unit of phonological encoding: prosodic or lexical word," *Cognition*, vol. 85, no. 2, pp. B31-41, 2002.
- [4] D.H. Whalen. "Coarticulation is largely planned," *Journal of Phonetics*, vol. 18, pp. 3-35, 1990.
- [5] G. Barbier. *Contrôle de la production de la parole chez l'enfant de 4 ans : l'anticipation comme indice de maturité motrice*. Ph.D. thesis. GIPSA-Lab, Université de Grenoble.
- [6] L. Nijland, B. Maassen, S. Van Der Meulen, F. Gabreels, F.W. Kraaimaat, and R. Schreuder. "Planning of syllables in children with developmental apraxia of speech," *Clinical Linguistics and Phonetics*, vol. 17, no. 1, pp. 1-24, 2003.
- [7] S. Nittrouer, M. Studdert Kennedy, and S.T. Neely. "How children learn to organize their speech gestures: further evidence from fricative-vowel syllables," *Journal of Speech and Hearing Research*, vol. 39, no. 2, pp. 379-389, 1996.
- [8] A. Noiray, M.A. Cathiard, C. Abry, and L. Menard. "Lip rounding anticipatory control: Crosslinguistically lawful and ontogenetically attuned," In: B. Maassen and P. van Lieshout (eds.) *Speech Motor Control: New Developments in Basic and Applied Research*, pp. 153-171, OUP, 2010.
- [9] B.H. Repp. "Some observations on the development of anticipatory coarticulation," *Journal of the Acoustical Society of America*, vol. 79, no. 5, pp. 1616-1619, 1986.
- [10] M.A. Redford. "Grammatical word production across metrical contexts in school-aged children's and adults' speech". *Journal of Speech, Language, and Hearing Sciences*, in press.
- [11] Boersma, Paul & Weenink, David. Praat: doing phonetics by computer [Computer program]. Version 6.0.35, retrieved February 2017 from <http://www.praat.org/>
- [12] L. Gerstman. "Classification of self-normalized vowels". In: *IEEE Transactions on Audio and Electroacoustics* vol. 16, no. 1, pp. 78-80, 1968.
- [13] D. Bates, M. Maechler, B. Bolker, and S. Walker. "Fitting linear mixed-effects models using lme4," *Journal of statistical software*, vol. 67. No. 1, pp. 1-48, 2015.
- [14] R Core Team. *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.
- [15] D. Recasens, M.D. Pallarès, & J. Fontdevila. "A model of lingual coarticulation based on articulatory constraints," *The Journal of the Acoustical Society of America*, vol. 102, no. 1, pp. 544-561, 1997.