



Published in final edited form as:

J Speech Lang Hear Res. 2008 December ; 51(6): 1424–1437. doi:10.1044/1092-4388(2008/07-0020).

The Breadth of Coarticulatory Units in Children and Adults

Lisa Goffman,

Purdue University, West Lafayette, IN

Anne Smith,

Purdue University, West Lafayette, IN

Lori Heisler, and

Purdue University, West Lafayette, IN

Michael Ho

Boston University

Abstract

Purpose—To assess, in children and adults, the breadth of coarticulatory movements associated with a single rounded vowel.

Method—Upper and lower lip movements were recorded from 8 young adults and 8 children (aged 4–5 years). A single rounded versus unrounded vowel was embedded in the medial position of pairs of 7-word/7-syllable sentences.

Results—Both children and adults produced movement trajectories associated with lip rounding that were very broad temporally (i.e., movement duration lasting 45% to 56% of the sentence). Some effects appeared to extend across the entire utterance. There were no differences between children and adults in the extent of the coarticulatory effect. However, children produced relatively variable movements associated with lip rounding.

Conclusions—These data support the hypothesis that, for young children and adults, broad chunks of output have been planned by the onset of implementation of a sentence. This implies that, based on a change in a single phoneme, the motor commands to the muscles are altered for the production of the entire sentence.

Keywords

coarticulation; speech production; developmental speech production; syllable

Coarticulation effects—that is, the influence of the production of one phonetic segment on surrounding units—have been widely documented in studies of adult speakers. The classic study of Daniloff and Moll (1968) provided clear kinematic evidence that movements associated with a single vowel could be initiated up to four consonants before that vowel and that such coarticulatory effects extend across syllable and word boundaries. MacNeilage (1970) remarked on the ubiquity of variability in speech production, with the highly variable acoustic and kinematic expression of a single segment arising in large part because of coarticulatory effects. Indeed, speech articulation is coarticulation. Numerous studies (e.g., Benguerel & Cowan, 1974; Daniloff & Moll, 1968; Perkell & Matthies, 1992; Recasens,

2002), using both kinematic and acoustic analyses, have shown that adults interleave articulatory movements across adjacent segments.

Despite the ubiquity of coarticulation in speech, theoretical accounts of this phenomenon are quite varied. This state of affairs probably reflects our lack of a comprehensive model that provides a mapping between abstract linguistic units and their physiological realization in muscle activity and movements, which then determine the acoustic output. Daniloff and Moll (1968) suggested that their findings best fit within a *feature look-ahead model*, which posited that the speech production planning process scanned ahead in time (Henke, 1966). A feature (e.g., lip rounding) that was not in opposition to other features required for the sequence would be inserted as early as possible, and it would carry over as long as no conflict with another upcoming feature occurred. This type of feature-based model is computationally demanding, and it makes no attempt to incorporate larger stored units (e.g., syllables) that could be used to limit online computational demands. Using kinematic data from tongue movement, Recasens (2002) showed that there are phoneme-specific effects on the extent of coarticulation. These effects are broad and, in some instances, discontinuous, in that they can be interrupted by competing gestures. Such adaptive effects are in contrast to the temporally fixed coarticulatory effects proposed by Fowler and Saltzman (1993). These investigators argue that coarticulatory effects are “temporally limited, and do not extend very far backward in time from the period of a gesture’s own predominant interval” (p. 185). Although there is some evidence for adaptive effects in adults (Recasens, 2002), very few experiments have considered coarticulation beyond the word, especially in children’s speech.

Recent models of adult speech production, such as Levelt’s (1989; Levelt, Roelofs, & Meyer, 1999), posit the syllable as a core production unit. Speech output processes are organized into syllabic units, and adult speakers store an inventory of frequently occurring syllables, termed a *syllabary*, which can be drawn on during online speech production (Cholin, Levelt, & Schiller, 2006; Levelt & Wheeldon, 1994). This hypothesis is attractive in that it simplifies the overwhelmingly complex task of real-time speech production. In Levelt’s language production model, there is a syllabification stage of planning, and presumably, coarticulatory effects arise when adjacent syllables are planned for co-production.

Although there is clear evidence that the syllable serves as a critical unit at the interface between phonological and phonetic processing, it is also likely that the syllable is not a privileged unit—or, at least, it is not the only unit used in the planning process. An alternative is that multiple units of different lengths and linguistic types are co-produced, with motor commands generated that synthesize simultaneous movement demands for a variety of upcoming linguistic goals. A. Smith and Goffman (2004) have suggested that linguistic units of many different sizes, ranging from the feature to the phrase or sentence, influence the production planning and implementation process. A. Smith and Goffman have also argued that adults, with their many years of speech production experience, possess multilayered and complex mappings among linguistic units, auditory targets, and movements. Young children have the task of learning these complex, multilayered mappings. It makes sense, then, to examine the parameters of young children’s productions for evidence of the operation of various units involved in the planning process.

With regard to coarticulation, children’s speech has been the subject of numerous studies. These studies have been largely framed around the question of whether children possess broader, less segmentally specified units of speech production compared with adults. Phonetic transcription and acoustic results have suggested that very young children acquire movement patterns associated with whole word or even phrase units (Ferguson & Farwell, 1975; Vihman, 1996) or syllables (Nitttrouer, 1993), rather than with segments. Thus, young children initially learn poorly coordinated articulatory routines (Vihman, 1996) that become more precisely

controlled with maturation (Nittrouer, 1993; Nittrouer, Studdert-Kennedy, & McGowan, 1989). A related, movement-based hypothesis has been proposed by MacNeilage and Davis (1990, 2000) in which the basic organizing unit is the frame, which corresponds to an open–close cycle of the jaw. The developmental task is in filling in the local details, or the content, which corresponds with specific segments. All of these ideas have in common a developmental sequence that begins with larger syllable or word units and transitions to smaller segmental units.

To date, transcription and acoustic evidence have served as the primary data sources supporting the idea that children’s speech production units are organized at the level of the syllable or the word rather than the segment (Nittrouer, 1993, p. 960). In several acoustic studies of children’s speech production, usually of disyllabic sequences, greater gestural overlap has been inferred in young children than in older children and adults (Goodell & Studdert-Kennedy, 1993; Nittrouer et al., 1989). For example, Nittrouer et al. (1989) observed that the F2 frequency components in the fricatives /s/ and /ʃ/ were influenced by whether the following vowel was an /i/ or /u/. Furthermore, children were imprecise in their articulatory constrictions for these fricative consonants. These results together suggest that children’s speech movements are both broader and less tied to the phoneme than are adults’ speech movements.

Other investigators argue that coarticulatory effects are actually reduced in children and that developmental differences emerge because of higher variability of speech production output processes rather than because of the units themselves (Katz, Kripke, & Tallal, 1991; Sereno, Baum, Marean, & Lieberman, 1987). For example, Katz and colleagues (1991) examined intrasyllabic coarticulatory effects in CV sequences and found that, for 3-, 5-, and 8-year-old children, factors related to the precision of speech production rather than to the organization of the basic linguistic unit lead to the developmental differences observed in coarticulation. Certainly, children’s speech production differs from adults, as observed in their increased motor and acoustic variability and their slower speech rates (Goffman, 2004; Kent & Forner, 1980; Sharkey & Folkins, 1985; A. Smith & Goffman, 1998; A. Smith & Zelaznik, 2004; B. L. Smith, 1995; Walsh and Smith, 2002). Thus, it seems plausible that poor speech motor precision may interfere with the assessment of the organization of speech production units.

It was our objective to assess the breadth and stability of lip rounding movement trajectories across an utterance. We collected movement data rather than making inferences from the acoustic signal. In adults, coarticulatory effects have been observed to cross up to six consonant segments (Benguerel & Cowan, 1974) and multiple syllables (Recasens, 2002). Consistent with these results, in adults we hypothesized extremely broad phrase-level coarticulatory effects that cross syllable and word boundaries. Coarticulatory effects in children have seldom been investigated in units larger than the syllable. If children at this age are using word- or syllable-based planning units, coarticulation would not cross word boundaries. If, on the other hand, they are already using phrase-level planning, coarticulation would cross word boundaries and influence the phrase.

Method

Participants

Sixteen individuals participated in this study, 8 young adults and 8 children ($M = 5;1$ (years; months), $SD = 1.5$ months, range = 4;10–5;3). Adults filled out a case history in which they reported normal educational; developmental; medical; and speech, language, and hearing histories. Children showed scores within the normal range on the Columbia Mental Maturity Scale (Burgemeister, Blum, & Lorge, 1972) as well as on the Reynell Developmental Language Scales–U.S. edition (Reynell & Gruber, 1990). They also passed a hearing screening in which pure tones were presented bilaterally at 25 dB HL at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz.

Structural and functional oral motor skills were determined to be at expected levels using the Clinical Assessment of Oropharyngeal Motor Development (Robbins & Klee, 1987).

Data Recording

Video- and audiorecordings were collected from all participants. These were used for confirmation that selected utterances were phonetically accurate. The Optotrak (Northern Digital, Waterloo, Ontario, Canada) was used to record oral movements. Small (6-mm) infrared light-emitting diodes (IREDs) were placed on the structures to be tracked: the upper lip, lower lip, and jaw. Lip markers were attached at the vermilion border at midline. To record head motion, four IREDs were attached to a pair of modified sports goggles. An additional marker was placed on the center of the forehead. Data from these markers were used to compute the three-dimensional axes of the head. To correct for head motion artifact, lip movements were computed in reference to the head coordinate system that was generated for each participant. (See A. Smith & Zelaznik, 2004, for details regarding this method.) The kinematic data were collected at a sampling rate of 250 Hz. Displacement data were low-pass filtered in the forward and backward directions with a 10-Hz cutoff prior to the computation of velocity. An acoustic signal (collected at a sampling rate of 8000 Hz) was recorded time locked to the kinematic data so that we could confirm that the selected movement record aligned with the appropriate utterance.

Procedures and Stimuli

This experiment was completed in a 20- to 30-min testing session. After participating in the experiment, children also completed standardized tests, usually across two sessions. During the experimental task, children were seated in a stable Rifton chair that includes a play tray and adjustable head and foot rests. Adults sat in a wooden chair with a headrest.

Pairs of stimuli were constructed, such that only the utterance medial vowel (rounded or unrounded) differentiated each pair. For all productions, the identical sentence frame was produced: "Mom has the ___ in the box." Embedded word pairs were: *goose/geese*, *moon/man*, and *boot/beet*. These stimuli were selected because they included minimal pairs that differed in relation to lip rounding. We note that lip rounding is not the only aspect that differentiates the vowels (e.g., jaw height). The unrounded item in each pair served as a critical control for ensuring that the observed differences could be attributed to the individual vowel. Each utterance was produced 12–15 times in succession, so that 10 accurate utterances that contained no overt disfluencies could be extracted for analysis. Sentences were elicited in the context of a play routine. Toy props, including a doll, a box, and objects associated with each of the stimuli, were introduced. Children and adults were initially familiarized with the names of each object. Imitation prompts were used to teach the sentence frame. When the doll placed a toy in the box, the child or adult produced the utterance "Mom has the ___ in the box." This routine was repeated until a sufficient number of productions were obtained. Because each sentence was repeated multiple times in succession, speakers were not observed to use contrastive stress to highlight minimal pairs. For adults, the same routine was used, and participants were told that this was a study also being completed with children.

Analyses

Perceptual judgments—Videotapes were coded by two observers, and any utterances that contained phonetic errors or speech disfluencies were excluded. At the level of transcription, only error-free utterances were included. The first 10 error-free productions from each participant were used in the analyses.

Kinematic analysis—Utterances were also excluded if the IREDs were out of the cameras' view during data collection because of extreme head movement. In all but one case, 10

productions for each speaker in each condition met perceptual and kinematic criteria and were included in the analyses (1 child had only 9 productions of the word *beet*).

Extraction of movement sequences—Figure 1 illustrates the movement data extraction. The sentence frame “Mom has the ___ in the box” was used because the onset and offset of this utterance are easily identifiable based on superior–inferior (SI) lower lip displacement. Lower lip movement (SI dimension) is shown in the top panel of Figure 1, and the derived velocity is shown in the middle panel. Peak velocity of the lower lip opening movement for the initial [m] in “mom” was used to mark the onset of the utterance, and peak velocity of the opening movement for the [b] in “box” was used to mark the offset. Following visual inspection, a MATLAB (The MathWorks, 2001) algorithm selected the precise location of each velocity peak. These start and end points from the lower lip signal were chosen because they are easy to pick reliably, ensuring that the same kinematic events were included in every participant’s record. The bottom panel of Figure 1 shows the associated anterior–posterior upper lip movement associated with the lip rounding—in this example, for the word *boot*. As has been observed in previous work (Perkell & Matthies, 1992), lip rounding was more evident in the upper than the lower lip, and we therefore used the upper lip signal for all measures of lip rounding in this experiment. We acknowledge, however, that reliance on upper lip movement is a methodological decision and that lip rounding is complex and involves multiple dimensions of movement. Of interest in the present study was the timing and amplitude of the upper lip rounding movement relative to the whole sentence. Thus, the sentence containing the corresponding unrounded vowel (in this case, *beet*) served as a control because the two sentences differed only in the single vowel. Following extraction of the movement trajectories for each sentence (shown between the dotted lines, Figure 1, bottom trace), several analyses were completed, as described later in this article.

Lip protrusion measure—The purpose of this analysis was to verify that sentences in each pair containing a rounded vowel were produced with larger anterior–posterior upper lip movement compared with the control sentence. Thus, for each sentence, the difference between maximum and minimum values of upper lip movement in the anterior–posterior dimension was measured (see Figure 2).

Spatiotemporal stability of lip rounding—This analysis was designed to assess spatiotemporal stability of upper lip movements associated with lip rounding. Examples of 10 upper lip movement records extracted for the sentences “Mom has the *goose/geese* in the box” are shown in Figures 3 and 4. The spatiotemporal index (STI; A. Smith, Goffman, Zelaznik, Ying, & McGillem, 1995; Smith, Johnson, McGillem, & Goffman, 2000) was used to assess variability across conditions and groups. The analytic purpose of the STI is to quantify the stability of underlying movement patterns when absolute differences in duration (e.g., rate) and amplitude (e.g., loudness) are eliminated. As illustrated in Figures 3 and 4 (middle graph), in this analysis movement trajectories corresponding to the entire utterance were linearly amplitude- and time-normalized (for a detailed description of this analysis, see A. Smith et al., 1995; A. Smith et al., 2000). Following normalization, standard deviations were computed at 2% intervals across the 10 time- and amplitude-normalized displacement records. The STI is the sum of these 50 standard deviations. The STI values (as illustrated in the bottom panels of Figures 3 and 4) were used to compare spatiotemporal aspects of movement stability across groups (adult, child) and conditions (utterances containing a rounded or unrounded vowel; individual words *goose/geese*, *boot/beet*, *moon/man*).

Total area of difference between rounded and un-rounded conditions—This analysis was designed to provide a global measure of temporal and spatial differences between the rounded versus the unrounded vowel conditions. For each sentence pair (i.e., *goose/geese*,

boot/beet, moon/man), grand averages of the 10 normalized un-rounded upper lip anterior–posterior trajectories and the 10 normalized rounded trajectories were computed. These two grand averages, or templates, are illustrated in Figure 5. The rounded versus unrounded templates were compared by computing the total area of difference between unrounded and rounded conditions within each participant for each of the three sentence pairs—that is, the values of the point-by-point differences between rounded and unrounded upper lip anterior–posterior templates were determined by summing the absolute value of the difference at every point through the record, yielding a normalized area difference score. Because both positive and negative differences were included, this analysis allowed for the incorporation of lip retraction and protrusion differences.

Temporal extent of lip rounding effects—Another measure was used to assess the temporal extent of the rounding movement trajectory, from its onset to its offset. As shown in Figure 6, the average template of the upper lip anterior–posterior movement for the unrounded sentences was used, this time as a reference to aid in determining the onset and offset of the rounded upper lip trajectory. In this case, each of the 10 individual normalized trajectories of the rounded sentence for 1 participant was compared to his/her normalized template of the unrounded reference (e.g., individual exemplars of the utterance containing the word *goose* were compared to the average template of utterances containing the word *geese*). Individual tokens were included because, unlike the previous analysis, this was a simple temporal measure of the onset, offset, and duration of the rounding movement in relative time. The onset and offsets of the individual rounding movements were visually identified. An algorithm then selected the point corresponding to the minimum value associated with the onset and the offset of the rounding movement. Normalized (expressed as a percentage) time of divergence from onset to offset of the rounding movement was computed (see Figure 6 for an illustration of this method). A second judge, well trained in kinematic analysis and with minimal knowledge of the hypotheses of this study, independently assessed records from 3 randomly selected participants (2 children and 1 adult). For children, the mean difference of percent divergence between judges was 1.71%. For the adults, there was no between-judge difference in the selection of onsets or offsets. It is important to note that for 3 adults and 2 children, some records were excluded from this analysis because onset and offset points were not obvious. Specifically, for each of the 3 adults, a total of 1, 1, and 4 tokens were excluded for the sentence containing the word *goose*; 3, 7, and 9 tokens were excluded for *boot*; and 1 and 3 for *moon*. For each of the 2 children, a total of 5 and 2 tokens were excluded for the sentence containing the word *goose*.

Statistical Analyses

For all analyses, mixed analyses of variance (ANOVAs) were completed, with an alpha level of .01 used in all statistical tests. Group (adult, child) served as the between-subjects variable, and rounding (rounded vs. unrounded vowel) and word (*goose/geese, moon/man, boot/beet*) as the within-subject variables.

Results

Lip Protrusion Measure

As illustrated in Figure 7, both adults and children produced larger anterior–posterior upper lip movement in sentences containing a rounded vowel, $F(1, 14) = 37.94, p < .0001, \eta_p^2 = .73$. There were no group effects, $F(1, 14) < 1, p > .20$; children and adults showed equal amplitudes of upper lip rounding movements. This analysis revealed that all speakers used lip rounding to produce the vowel distinction. Differences were observed across word pairs, $F(2, 28) = 11.93, p = .0002, \eta_p^2 = .46$. Post hoc comparisons using the Tukey's honestly significant

difference (HSD) procedure showed that the *goose/geese* pair differed from both the *boot/beet* pair ($p = .04$) and the *moon/man* pair ($p = .0002$).

Spatiotemporal Variability of Rounded and Unrounded Oral Movements

The results of the STI analysis are summarized in Figure 8. As expected, group differences were observed, with adults producing more consistent movements than children, $F(1, 14) = 17.19$, $p < .0001$, $\eta_p^2 = .55$. Sentences containing rounded vowels were produced with less spatiotemporal variability than those containing unrounded vowels, $F(1, 14) = 39.44$, $p < .00001$, $\eta_p^2 = .74$. There were no word effects, $F(2, 28) < 1$, $p > .20$.

Total Area of Difference Between Rounded and Unrounded Conditions

As illustrated in Figure 9, the sum of all point-by-point difference scores (striped area in Figure 5) revealed no group effect, $F(1, 14) < 1$, $p > .20$. Children and adults produced the same degree of distinction between rounded and unrounded utterances. There was no effect of word, $F(2, 28) < 1$, $p > .20$.

Temporal Extent of Lip Rounding Effects

The mean duration in normalized time of temporal divergence (i.e., percentage of the record diverging from onset to offset of the rounding movement; see Figure 6) measures for each group was as follows: adults: *goose* = 56% ($SD = 8\%$), *boot* = 46% ($SD = 13\%$), *moon* = 45% ($SD = 10\%$); children: *goose* = 55% ($SD = 10\%$), *boot* = 46% ($SD = 5\%$), *moon* = 47% ($SD = 9\%$). That is, approximately half (45%–56%) of the sentence “Mom has the ____ in the box” contained lip rounding associated with the medial [u]. In the example illustrated in Figure 6, as verified acoustically, the onset of lip rounding began in the middle of the word *has* and the offset of lip rounding began at the end of the word *in*. According to a mixed ANOVA, there was no effect of group, $F(1, 14) < 1$, $p > .20$. There was an effect of word, $F(2, 28) = 7.14$, $p = .003$, $\eta_p^2 = .34$, with no Group \times Word interaction, $F(2, 28) < 1$, $p > .20$. Post hoc (Tukey’s HSD) testing showed that lip rounding movements for the production of *goose* included a greater proportion of the overall record than for *boot* ($p = .009$) and *moon* ($p = .007$).

The onset of the temporal divergence in normalized time was as follows: Adults: *goose* = 25% ($SD = 7\%$), *boot* = 34% ($SD = 13\%$), *moon* = 35% ($SD = 9\%$); children: *goose* = 27% ($SD = 7\%$), *boot* = 36% ($SD = 7\%$), *moon* = 37% ($SD = 7\%$). There were no group differences in this point of onset divergence, $F(1, 14) < 1$, $p > .20$. There were word effects, $F(2, 28) = 8.62$, $p = .001$, $\eta_p^2 = .38$. Post hoc testing revealed that the word *goose* was produced with an earlier onset of lip rounding compared with the words *boot* and *moon*.

Finally, the offset of the temporal divergence occurred at the following points in the movement record: adults: *goose* = 81% ($SD = 6\%$), *boot* = 81% ($SD = 7\%$), *moon* = 80% ($SD = 5\%$); children: *goose* = 82% ($SD = 6\%$), *boot* = 82% ($SD = 9\%$), *moon* = 84% ($SD = 8\%$). Again, children and adults showed similar points of temporal divergence, $F(1, 14) < 1$, $p > .20$. In this case, there were no phonetic effects on the point of offset divergence, $F(2, 28) < 1$, $p > .20$. The offset of the rounding movement for all three sentences occurred at a similar point in the relative trajectory.

Individual Variation

Although the statistical results were clear, there were individual differences across and within participants. Anterior–posterior upper lip movement is not the only option for producing a rounded vowel. The observed individual variability is expected as a result of articulatory trading relations.

Discussion

Surprisingly broad coarticulatory effects were observed for 4- to 5-year-old children and for adult speakers. Utterances that contained a rounded vowel (i.e., [u]) were produced with an anterior–posterior lip movement that exerted a broad influence on the upper lip movement trajectory for the sentence. A control condition that included the identical sentence frame but with an un-rounded vowel (e.g., *man* instead of *moon*) demonstrated that the observed effect was the consequence of the change in a single segment. Consistent with the findings reported for adults (Recasens, 2002), the extent of these effects is broader than has previously been reported in most studies (e.g., Boyce, 1990; Benguerel & Cowan, 1974; Daniloff & Moll, 1968; Perkell & Matthies, 1992). Both the breadth (temporal measure; relative duration of the rounding movement) and the extent (absolute anterior and posterior displacement difference measure) of these effects were similar in adults and in children; only articulatory movement variability distinguished child from adult speakers. Importantly, when embedded in a seven-word, seven-syllable utterance, coarticulatory effects crossed word and even phrase boundaries for both child and adult speakers.

Production Units

Coarticulation, particularly anticipatory coarticulation, has previously been considered by some investigators as an adaptive and varying index of the extent of planning units (Benguerel & Cowan, 1974; Daniloff & Moll, 1968; Nittrouer et al., 1989; Recasens, 2002). An alternative view is that coarticulatory effects arise from overlapping speech gestures (Fowler & Saltzman, 1993) or from physiological constraints (Bell-Berti & Harris, 1981; Fowler & Saltzman, 1993). These investigators posit that gestures are tied to the “gestural frame” (Fowler & Saltzman, p. 185). Thus, rounding gestures are produced in a relatively short and invariant time window. The present results clearly do not support this alternative model.

Interestingly, in many earlier studies, the window of analysis was narrow, focusing on syllables or sequences of consonants that crossed a single word boundary. Similar to Recasens (2002), we considered coarticulatory effects across a sentence. Also, as has been previously suggested in a research note by Gelfer, Bell-Berti, and Harris (1989), we used a minimal pair as a control for the rounded vowel. That is, pairs of sentences were selected that differed only in the rounding of a vowel embedded in utterance medial position. In this way, differences in coarticulation could be attributed to the single rounded segment. Rounding that was associated with competing gestures, such as the [b] in “boot” and “beet,” were controlled for.

Analyzing lip rounding across a sentence and comparing it with an unrounded minimal pair control, we found that both anticipatory and perseveratory effects were extremely broad. The rounding gesture itself subsumed 45%–56% of the utterance: When the lip retraction movement—which often preceded and followed the protrusion movement—was taken into account, an even greater extent was affected (see Figure 5 for an example of this analysis). From our data, we cannot determine whether these extremely broad coarticulatory effects reflect distinctive anticipatory and/or perseveratory effects. Electromyographic measures would be required to determine whether the retraction movement is actively controlled or is a passive biomechanical effect. Such measures would demonstrate whether there is active and systematic perioral activity during the entire utterance. What is particularly striking about these data—as illustrated in Figure 5 and which requires more exploration—is that some coarticulatory effects cross the entire six syllables analyzed and influence both protrusion and retraction. It would be important to pursue other segmental influences on these coarticulatory effects. However, because all other aspects of the sentence frame were controlled, it is evident that other segmental variables, such as the production of the [m] in “Mom,” do not drive this effect. Overall, co-articulatory effects are broad and influence movement timing and amplitude similarly in adults and children.

These data support the hypothesis that children at the age of 4–5 years are already using multiple planning units. For young children and adults, speech production is not only incrementally planned at the level of the syllable, but broad chunks of output have been planned by the onset of implementation of a sentence. This implies that, based on a change in a single phoneme, the motor commands to the muscles are altered for the production of the entire sentence.

In their speech production model, Levelt and colleagues (Cholin et al., 2006; Levelt et al., 1999; Levelt & Wheeldon, 1994) propose that the syllable serves as the basic unit linking language processing with motor implementation. For example, syllables that are highly frequent are produced with more rapid response times than those that are low in frequency (Cholin et al., 2006; Levelt & Wheeldon, 1994). The present results suggest that for lip rounding, at least some aspects of speech production are organized at a broader level than the syllable; higher levels of language processing (e.g., prosodic words and phrases) also link directly to speech motor implementation. The inclusion of varying units such as syllables, prosodic words, and phrases is consistent with the theory of prosodic phonology and provides another line of evidence that units of speech production span multiple levels (Demuth, 1996; Gerken, 1996). It seems likely that multiple units are mapped at the interface between language formulation and motor production processes (A. Smith & Goffman, 2004).

Although anticipatory and perseveratory effects were not explicitly assessed in this experiment, the present findings speak indirectly to these phenomena. There were differential effects of phonetic content at the onset, but not the offset, of the rounding movement. That is, as reported in the results about the temporal extent of lip rounding, the onset of lip rounding for the word *goose*, which has no competing labial activity, occurred earlier than the onset of lip rounding for the words *boot* and *moon*. No such differences were observed in the offset of lip rounding, where there were no labial consonants for any of the target words. It is important to design an experiment explicitly to test how phonetic stimuli with competing features influence the timing of the rounding gesture. Also, as discussed by Whalen (1981) and Boyce (1990), it is crucial to consider cross-linguistic effects on coarticulation.

Implications for Developmental Theories

That speech production organization transcends the word has been previously reported in adults but has not been assessed in children. Much prior work has focused on the syllable or the word as the organizing unit for children as old as 3–8 years (Nittrouer, 1993; Nittrouer et al., 1989). However, many researchers have proposed that both processing (Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003) and production (e.g., Snow, 1994) units may be broader than the word. The finding that young children's coarticulation extends far beyond the syllable is compatible with other developmental results, in that infants are perceptually sensitive to a range of linguistic units, including segments, syllables, phrases, and clauses (see Jusczyk, 1997). Some investigators have suggested that infants first begin processing large prosodic units associated with phrases or sentences, with smaller units emerging later in development (Soderstrom et al., 2003). When toddlers begin to speak, a range of production units are thought to be available, including the syllable, the word, and the phrase (Bates, Bretherton, & Snyder, 1988; Demuth, 1996; Gerken, 1996). It seems plausible that production mirrors perception in that multiple units become available, including segments, syllables, words, phrases, and clauses. Big units such as phrases may developmentally co-occur with or precede small units such as segments and features. Thus, the present findings provide another source of evidence for multiple production units in children.

Variability and Extent of Movement

As expected, we did observe a developmental change in the precision, or variability, of speech movements, with children producing the simple anterior–posterior movement associated with

lip rounding more variably than adults. These results are consistent with prior hypotheses (Katz et al., 1991; Sereno et al., 1987). For example, it has been suggested that the differences in coarticulatory effects between children and adults “may reflect the development of automatized speech motor control programs (Sereno et al., 1987, p. 518).” Such automaticity is reflected in decreased speech motor variability. Thus, developmental differences observed between children and adults have been attributed variously to representational (Nitttrouer et al., 1989) and to performance (Katz et al., 1991; Sereno et al., 1987) factors. Similar to prior results (Goffman, 2004; Goffman & Malin, 1999; A. Smith & Zelaznik, 2004), the present findings suggest that performance factors play a significant role in developmental changes in speech production. Also, upper lip movement trajectories were less variable for both groups of speakers in the sentences containing rounded vowels. This is a predictable result in that the more constrained the articulator is for achieving an acoustic target, the less variable its movement pattern is on repeated productions.

We found that increased motor variability of the young children did not influence the breadth of coarticulatory effects. That is, even in the presence of increased variability, there were no differences between adults and children in any measure of the extent or timing of lip rounding coarticulation. Lip rounding amplitude was as prominent in children as adults. In terms of the real amplitude (in mm) of the upper lip protrusion movement for the sentences containing the rounded vowel, it may seem surprising that 4- to 5-year-olds and young adults produced anterior–posterior movements of the same extent. However, other work has demonstrated that the amplitude of lower lip and jaw movements for speech is the same in 5-year-olds and young adults (Goffman & Smith, 1999; Riely & Smith, 2003; B. L. Smith & Gartenberg, 1984), despite the obvious differences in oral facial structure size. It appears that young children use a relatively large proportion of their articulatory working space to achieve an acoustic goal.

The temporal extent and the overall distinction between rounded and unrounded movements (as measured by the difference score) revealed no developmental differences between adults and young children. Such a disassociation between performance factors and linguistic representations has been observed previously. Young children produce segmental distinctions, even in the face of increased variability (Goffman & Smith, 1999). Four- and 7-year-old children and adults produced lower lip movements that were similarly differentiated across phonetic segments (i.e., [m,b,p,f,v]). Similar to the present results, the relevant linguistic distinction was present even in these young children, but performance factors, in this case related to the ability to produce reliable movements, influenced production.

In earlier studies, children showed broader coarticulatory effects than adults (Nitttrouer et al., 1989). Both the analytic approach and the scope of the unit analyzed differed substantially between prior studies of coarticulation and the findings reported here. We are not claiming that segmental and syllabic effects do not play a significant role in the organization of speech production; there is substantial evidence in favor of the operation of these units. However, the present data imply that mappings across levels are more complex than envisioned in these prior studies. Studies using acoustic data assess the interplay among multiple articulators while measurement of the oral movements of single articulators reduces the complexity. Both levels of analysis illuminate the nature of the speech production process, which is complex and involves multiple mappings across levels. It is, of course, critical that the rounding trajectory represents only one component of movement that contributes to the final acoustic output. Thus, analysis of this component, if anything, underspecifies the degree of coarticulation, as it is well known that there are trading relations among articulators. The simple kinematic component recorded here—that is, the anterior–posterior lip rounding associated with the production of [u]—shows that one aspect of these mappings involves units that transcend the syllable or word.

Future Research

Additional experiments are needed to understand the source of the broad coarticulatory effects observed in the present experiment. For example, processing demands are greatly reduced in the imitative task included in this experiment. Future experiments should incorporate tasks that require increased online formulation. Especially for adults, the production of a well-practiced and frequently occurring syllable, in contrast to one that is highly infrequent, may also lead to reduced coarticulatory effects (Ernestus, Lahey, Verhees, & Baayen, 2006). Such frequency effects may differ in children, who may also have reduced familiarity with particular words. Furthermore, it may be that coarticulatory effects do not cross certain boundaries, such as those marking phrases or sentences. In the present study, one of the word pairs (i.e., *goose/geese*) included only a plural distinction. It is possible that factors beyond the segment, such as morphological status or phrase or sentence boundaries, may influence coarticulation. We acknowledge that the imitative task used here, and in other prior studies, does not allow for online sentence processing necessary in more spontaneous speech production tasks. A goal of future research will be to test whether the breadth of planning is influenced by varying task demands, particularly those incorporating increased planning, in a more natural production task.

Acknowledgments

This research was supported by the National Institute on Deafness and Other Communication Disorders Grants DC02527 and DC04826. We thank Alex Francis and Bill Saxton for their assistance with this project.

References

- Bates, E.; Bretherton, I.; Snyder, L. From first words to grammar: Individual differences and dissociable mechanisms. New York: Cambridge University Press; 1988.
- Bell-Berti F, Harris K. A temporal model of speech production. *Phonetica* 1981;38:9–20. [PubMed: 7267724]
- Benguereel A, Cowan HA. Coarticulation of upper lip protrusion in French. *Phonetica* 1974;30:41–55. [PubMed: 4851063]
- Boyce S. Coarticulatory organization for lip rounding in Turkish and English. *The Journal of the Acoustical Society of America* 1990;88:2584–2595. [PubMed: 2283431]
- Burgemeister, B.; Blum, L.; Lorge, I. Columbia Mental Maturity Scale. 3. New York: Harcourt Brace Jovanovich; 1972.
- Cholin J, Levelt WJM, Schiller NO. Effects of syllable frequency in speech production. *Cognition* 2006;99:205–235. [PubMed: 15939415]
- Daniloff R, Moll K. Coarticulation of lip rounding. *Journal of Speech and Hearing Research* 1968;11:707–721. [PubMed: 5719225]
- Demuth, K. The prosodic structure of early words. In: Morgan, JL.; Demuth, K., editors. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Erlbaum; 1996.
- Ernestus M, Lahey M, Verhees F, Baayen RH. Lexical frequency and voice assimilation. *The Journal of the Acoustical Society of America* 2006;120:1040–1051. [PubMed: 16938990]
- Ferguson CA, Farwell CB. Words and sounds in early language acquisition. *Language* 1975;51:419–439.
- Fowler CA, Saltzman E. Coordination and co-articulation in speech production. *Language and Speech* 1993;36:171–195. [PubMed: 8277807]
- Gelfer CE, Bell-Berti F, Harris KS. Determining the extent of coarticulation: Effects of experimental design. *The Journal of the Acoustical Society of America* 1989;86:2443–2445. [PubMed: 2600314]
- Gerken LA. Prosodic structure in young children's language production. *Language* 1996;72:683–712.
- Goffman L. Kinematic differentiation of prosodic categories in normal and disordered language development. *Journal of Speech, Language, and Hearing Research* 2004;47:1088–1102.
- Goffman L, Malin C. Metrical effects on speech movements in children and adults. *Journal of Speech, Language, and Hearing Research* 1999;42:1003–1115.

- Goffman L, Smith A. Development and phonetic differentiation of speech movement patterns. *Journal of Experimental Psychology: Human Perception and Performance* 1999;25:649–660. [PubMed: 10385982]
- Goodell EW, Studdert-Kennedy M. Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: A longitudinal study. *Journal of Speech and Hearing Research* 1993;36:707–727. [PubMed: 8377484]
- Henke, W. Unpublished doctoral dissertation. Massachusetts Institute of Technology; 1966. Dynamic articulatory models of speech production using computer simulation.
- Jusczyk, PW. *The discovery of spoken language*. Cambridge, MA: MIT Press; 1997.
- Katz WF, Kripke C, Tallal P. Anticipatory co-articulation in the speech of adults and young children: Acoustic, perceptual, and video data. *Journal of Speech and Hearing Research* 1991;34:1222–1232. [PubMed: 1787704]
- Kent RD, Forner LL. Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics* 1980;8:157–168.
- Levelt, WJM. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press; 1989.
- Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 1999;22:1–75. [PubMed: 11301520]
- Levelt WJM, Wheeldon L. Do speakers have access to a mental syllabary? *Cognition* 1994;50:239–269. [PubMed: 8039363]
- MacNeilage PF. Motor control of serial ordering of speech. *Psychological Review* 1970;77:182–196. [PubMed: 5454132]
- MacNeilage, PF.; Davis, BL. Acquisition of speech production: Frames then content. In: Jeannerod, M., editor. *Attention and performance XII: Motor representation and control*. Hillsdale, NJ: Erlbaum; 1990. p. 453-476.
- MacNeilage PF, Davis BL. On the origin of internal structure of word forms. *Science* 2000 April 21;288:527–531. [PubMed: 10775113]
- The MathWorks, Inc. *MATLAB: Higher performance numeric computation and visualization software*. Natick, MA: Author; 2001. [Computer software]
- Nittrouer S. The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech and Hearing Research* 1993;36:959–972. [PubMed: 8246484]
- Nittrouer S, Studdert-Kennedy M, McGowan RS. The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research* 1989;32:120–132. [PubMed: 2704187]
- Perkell JS, Matthies ML. Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within-and cross-subject variability. *The Journal of the Acoustical Society of America* 1992;91:2911–2925. [PubMed: 1629484]
- Recasens D. An EMA study of VCV coarticulatory direction. *The Journal of the Acoustical Society of America* 2002;111:2828–2841. [PubMed: 12083217]
- Reynell, J.; Gruber, C. *Reynell Developmental Language Scales–U.S. Edition*. Los Angeles: Western Psychological Services; 1990.
- Riely RR, Smith A. Speech movements do not scale by orofacial structure size. *Journal of Applied Physiology* 2003;94:2119–2126. [PubMed: 12576407]
- Robbins J, Klee T. Clinical assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders* 1987;52:271–277. [PubMed: 3455449]
- Sereno JA, Baum SR, Marean GC, Lieberman P. Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children. *The Journal of the Acoustical Society of America* 1987;81:512–519. [PubMed: 3558969]
- Sharkey SG, Folkins JW. Variability of lip and jaw movements in children and adults: Implications for the development of speech motor control. *Journal of Speech and Hearing Research* 1985;28:8–15. [PubMed: 3982001]
- Smith A, Goffman L. Stability and patterning of movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research* 1998;41:18–30.

- Smith, A.; Goffman, L. Interaction of motor and language factors in the development of speech production. In: Maasen, B.; Kent, R.; Peters, H.; van Lieshout, P.; Hulstijn, W., editors. *Speech motor control in normal and disordered speech*. Oxford, England: Oxford University Press; 2004. p. 225-252.
- Smith A, Goffman L, Zelaznik H, Ying S, McGillem C. Spatiotemporal stability and patterning of speech movement sequences. *Experimental Brain Research* 1995;104:493–501.
- Smith A, Johnson M, McGillem C, Goffman L. On the assessment of stability and patterning of speech movements. *Journal of Speech, Language, and Hearing Research* 2000;43:277–286.
- Smith A, Zelaznik H. The development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology* 2004;45:22–33. [PubMed: 15229873]
- Smith BL. Variability of lip and jaw movements in the speech of children and adults. *Phonetica* 1995;52:307–316. [PubMed: 7568398]
- Smith BL, Gartenberg TE. Initial observations concerning developmental characteristics of labial mandibular kinematics. *The Journal of the Acoustical Society of America* 1984;75:1599–1605. [PubMed: 6736423]
- Snow D. Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech, Language, and Hearing Research* 1994;37:831–840.
- Soderstrom M, Seidl A, Kemler Nelson D, Jusczyk PW. The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language* 2003;49:249–267.
- Vihman, MM. *Phonological development: The origins of language in the child*. Cambridge, MA: Blackwell Publishers; 1996.
- Walsh B, Smith A. Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes. *Journal of Speech, Language, and Hearing Research* 2002;45:1119–1133.
- Whalen D. Effects of vocalic formant transitions and vowel quality on the English s-sh boundary. *The Journal of the Acoustical Society of America* 1981;69:275–282. [PubMed: 7217525]

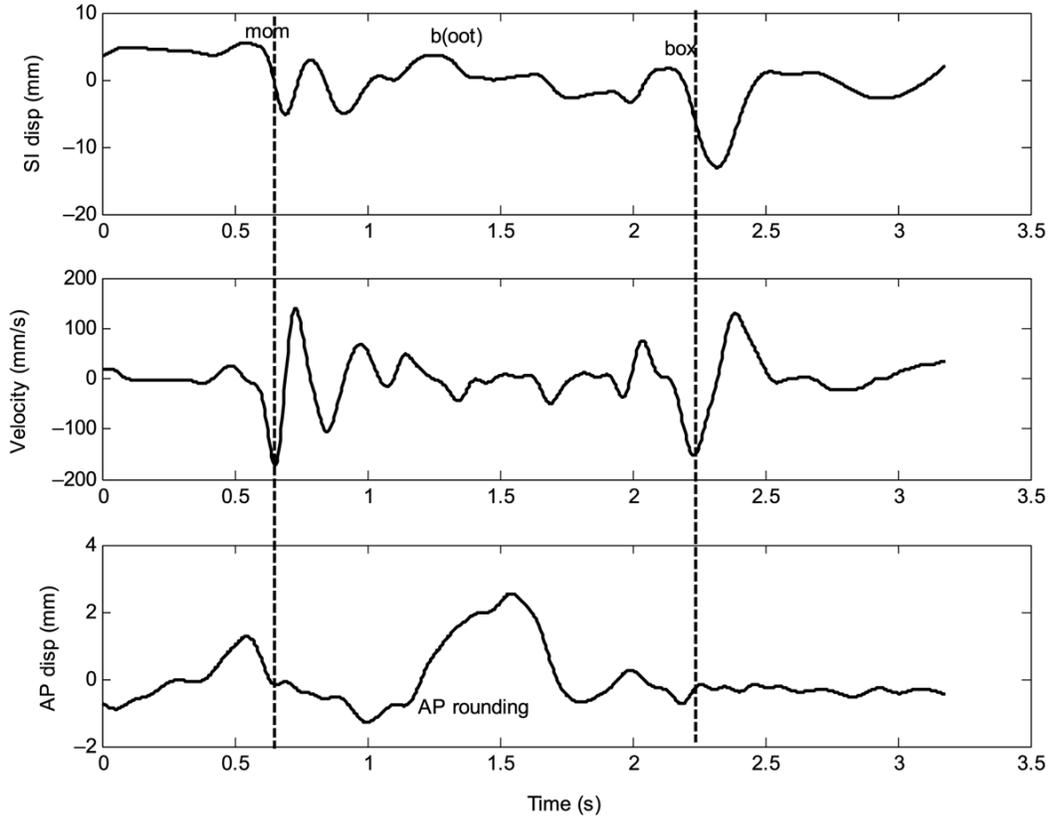


Figure 1. Example of extracted movement sequence from a child producing the target sentence “Mom has the *boot* in the box.” The top panel shows the superior–inferior (SI) displacement record associated with lower lip + jaw movement, and the middle panel depicts the derived velocity record. These two panels illustrate how movements for the utterance were extracted, beginning with the peak velocity in the opening movement for *mom* and ending with the peak velocity for the opening movement in *box*. The bottom panel shows the time-locked record for upper lip anterior–posterior (AP) movement (i.e., lip rounding). The dotted lines at the onset and offset of the sentence show the portion of the upper lip AP movement used in all analyses.

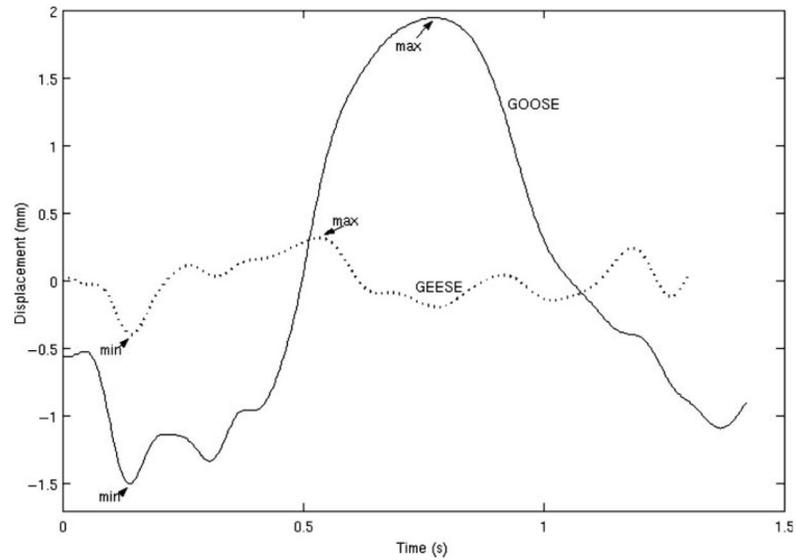


Figure 2.

Method for assessing the degree of lip protrusion. The solid line shows an utterance containing the word *goose*, with minimum (min) and maximum (max) values identified. The dotted line shows an example of the sentence containing the word *geese*. The lip protrusion measure was determined by subtracting the minimum from the maximum in the upper lip AP displacement.

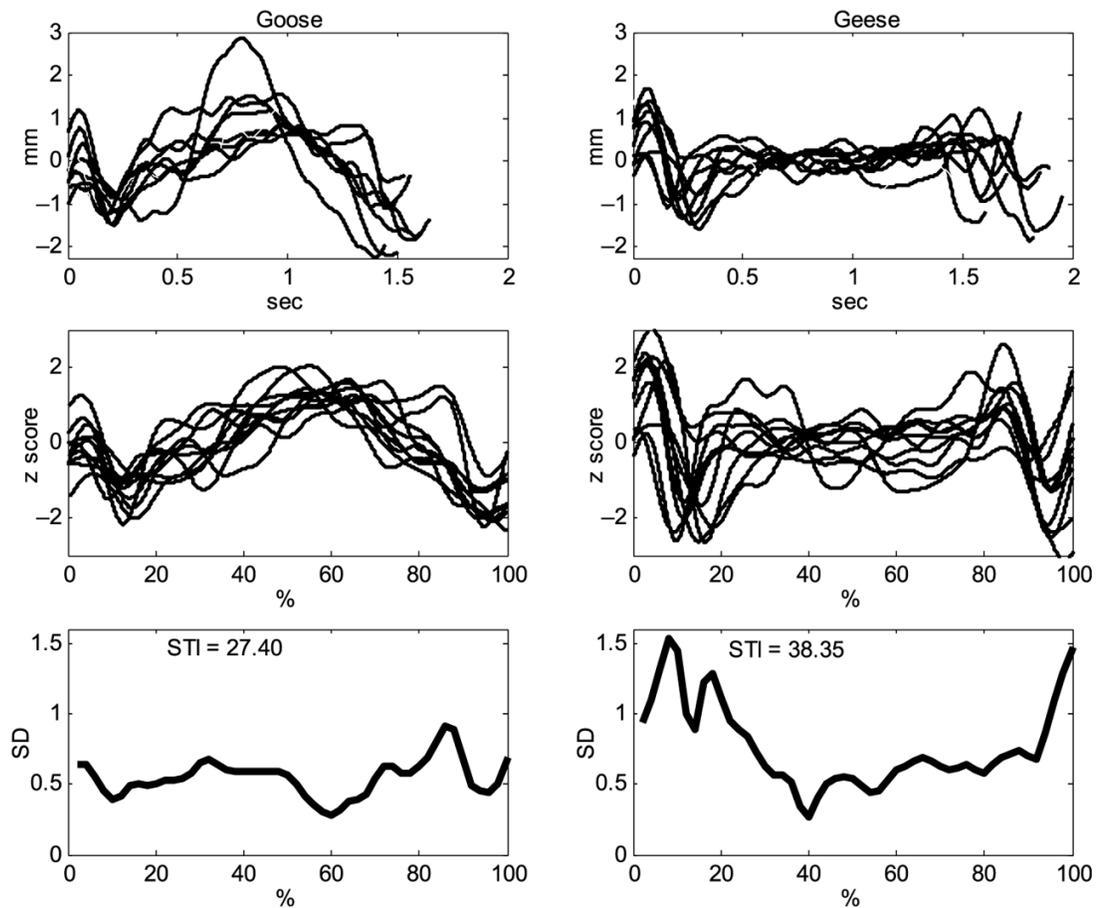


Figure 3.

Example of the spatiotemporal index for child productions of sentences containing the words *goose* and *geese*. These upper lip AP records were collected from the entire utterance, as shown in Figure 1. The top panel illustrates the original, non-normalized records; the middle panel illustrates the same records, now time- and amplitude-normalized; and the bottom panel illustrates the spatiotemporal index (STI).

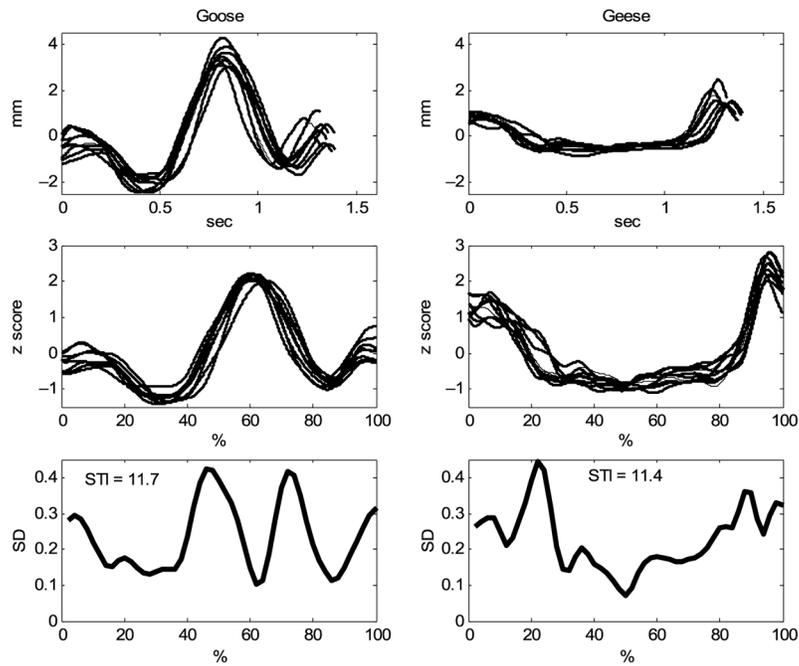


Figure 4.

Example of the spatiotemporal index for adult productions of sentences containing the words *goose* and *geese*. These upper lip AP records were collected from the entire utterance, as shown in Figure 1. The top panel illustrates the original, non-normalized records; the middle panel illustrates the same records, now time- and amplitude-normalized; and the bottom panel illustrates the STI.

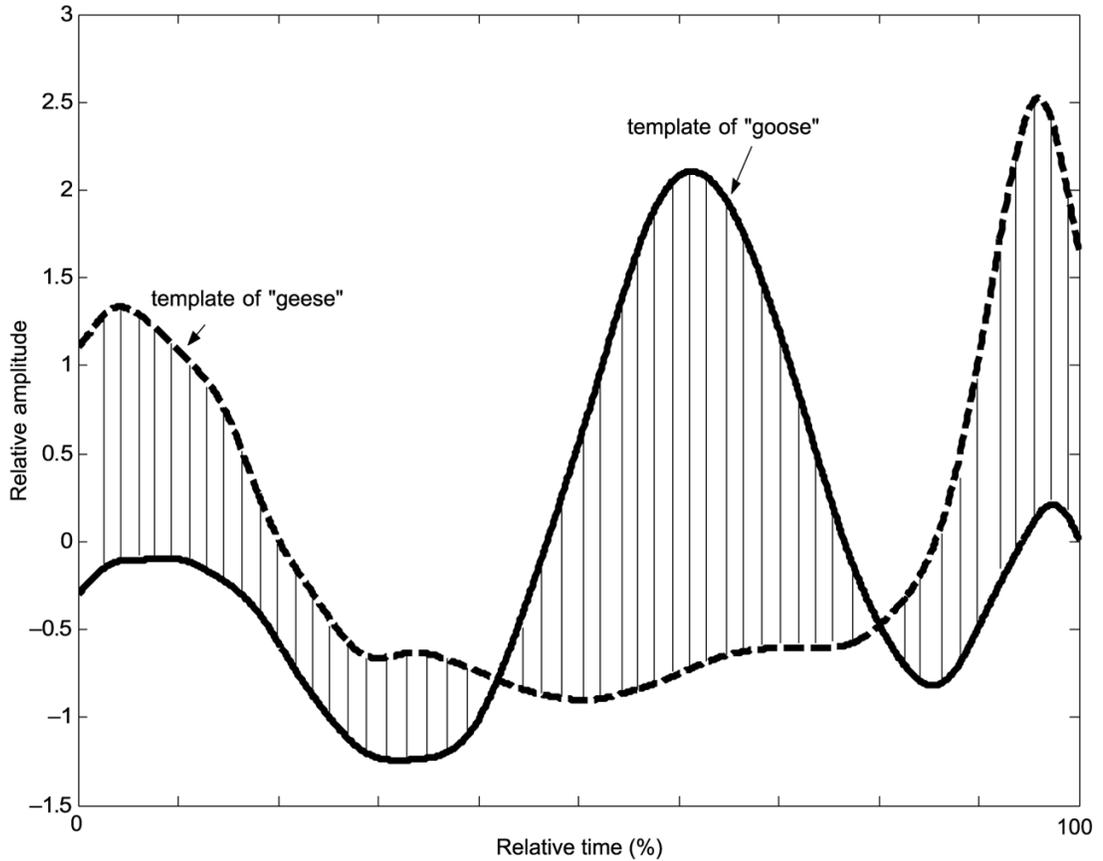


Figure 5.

Method for assessing how total normalized area of difference between rounded and unrounded conditions was calculated. A normalized average template of the upper lip AP movement for 10 productions of sentences containing the word *goose* is shown overlaying the average template for the upper lip AP movement for 10 sentences containing the word *geese*. The vertical lines are illustrative of the areas included in the integrative sum. A point-by-point difference score was obtained in the regions denoted.

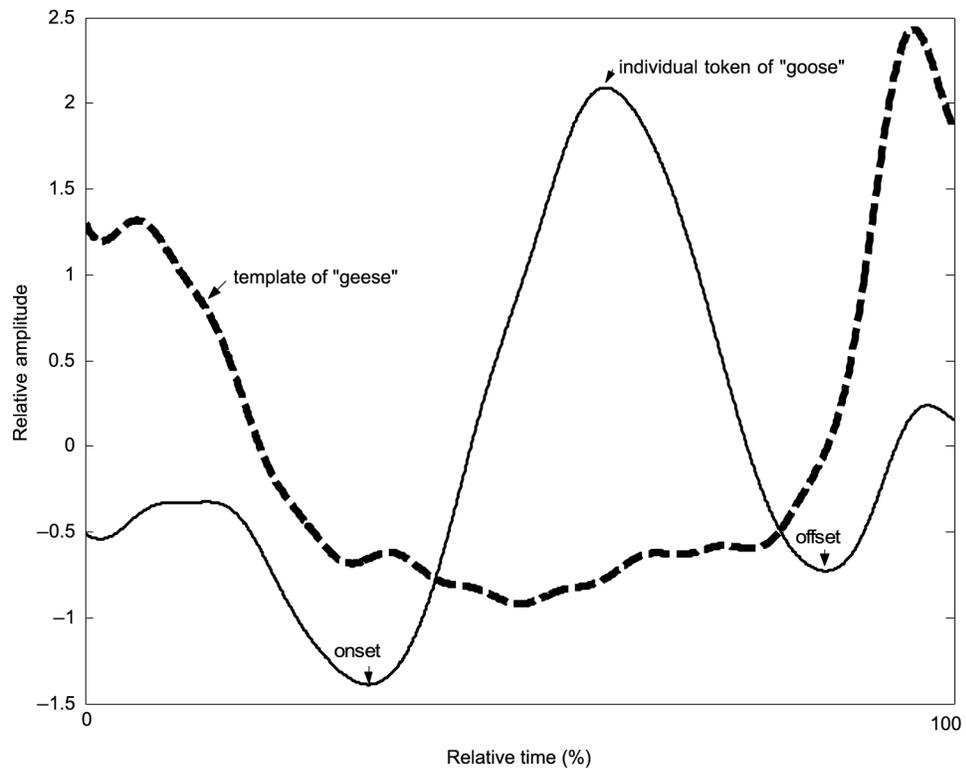


Figure 6. Method for assessing how the temporal extent of the lip rounding effect was determined. The normalized template from the control condition served as a comparison for assessing when the rounding movement began and ended. Each individual sentence containing a rounded vowel (thin solid line; in this example, *goose*) was assessed in comparison with the template for the sentence containing the unrounded vowel (thick dashed line; in this example, *geese*).

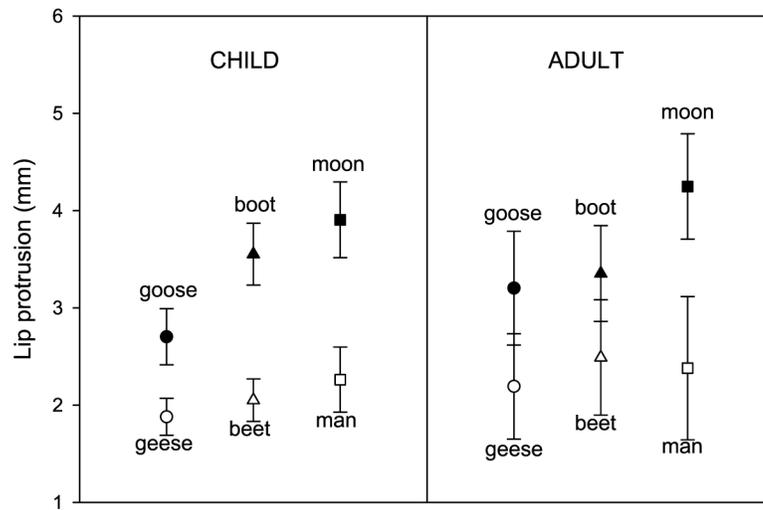


Figure 7. Lip protrusion displacement values (in mm) of upper lip AP movements associated with rounded (filled shapes) and unrounded (open shapes) targets. Error bars represent standard errors.

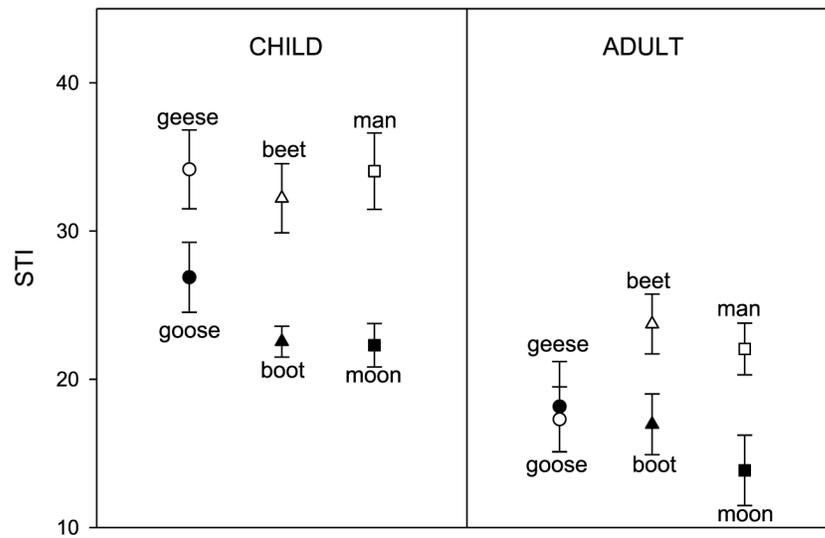


Figure 8. STI values of upper lip AP movements associated with rounded (filled shapes) and unrounded (open shapes) targets. Error bars represent standard errors.

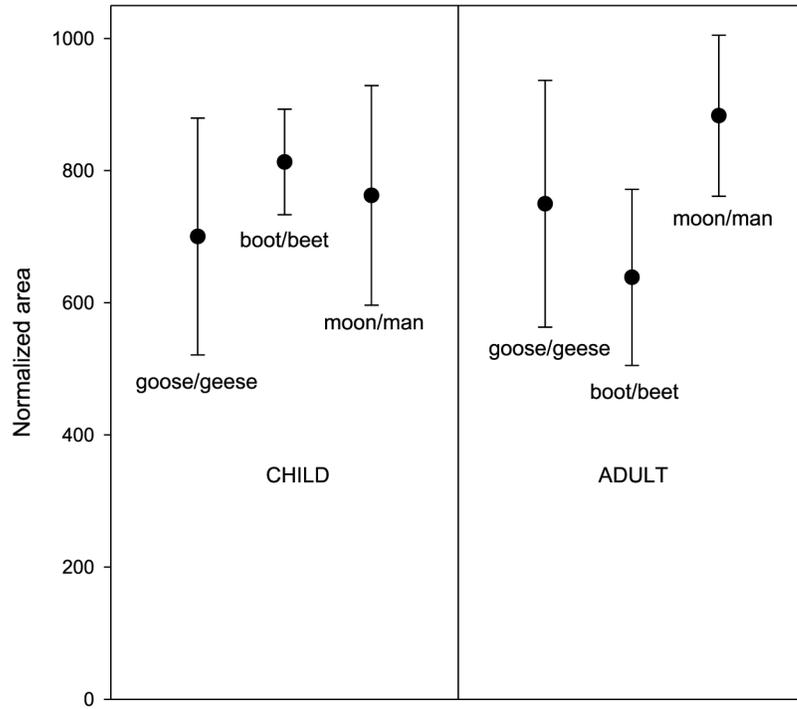


Figure 9. Total normalized area of the difference between rounded and unrounded conditions. Error bars represent standard errors.