Speech motor development: Integrating muscles, movements, and linguistic units

Anne Smith *

Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN, United States

Received 5 April 2006; accepted 12 June 2006

Abstract

A fundamental problem for those interested in human communication is to determine how ideas and the various units of language structure are communicated through speaking. The physiological concepts involved in the control of muscle contraction and movement are theoretically distant from the processing levels and units postulated to exist in language production models. A review of the literature on adult speakers suggests that they engage complex, parallel processes involving many units, including sentence, phrase, syllable, and phoneme levels. Infants must develop multilayered interactions among language and motor systems. This discussion describes recent studies of speech motor performance relative to varying linguistic goals during the childhood, teenage, and young adult years. Studies of the developing interactions between speech motor and language systems reveal both qualitative and quantitative differences between the developing and the mature systems. These studies provide an experimental basis for a more comprehensive theoretical account of how mappings between units of language and units of action are formed and how they function.

Learning outcomes: Readers will be able to: (1) understand the theoretical differences between models of speech motor control and models of language processing, as well as the nature of the concepts used in the two different kinds of models, (2) explain the concept of coarticulation and state why this phenomenon has confounded attempts to determine the role of linguistic units, such as syllables and phonemes, in speech production, (3) describe the development of speech motor performance skills and specify quantitative and qualitative differences between speech motor performance in children and adults, and (4) describe experimental methods that allow scientists...
to study speech and limb motor control, as well as compare units of action used to study non-speech and speech movements.

© 2006 Elsevier Inc. All rights reserved.

1. Introduction

There is a significant gap between models of language processing and production and models of speech motor control (see Smith & Goffman, 2004). Models of language processing evoke many levels of linguistic units involved in planning and perceiving speech: semantics, prosody, syntax, phonology, syllables, and phonemes. Models of speech motor control, on the other hand, sometimes invoke syllables, phonemes, or abstract gestures as the input units (Browman & Goldstein, 1986), but the primary explanatory variables are motor programs, central pattern generators, motor commands, and sensorimotor integration. As an example of the theoretical distance between these two types of models, consider Levelt's (1989) well-known model of language production. In his summary diagram of the model (see his Fig. 1.1), there are many large boxes: a conceptualizer (message generation and monitoring), a discourse model, a formulator (grammatical encoding, surface structure, and phonological encoding), a lexicon (lemmas and forms), a speech comprehension system, and finally a very small box in the bottom layer of the model, “the articulator.” The lowest level black box “articulator” of Levelt (1989) is, as he states, not this primary interest. Models of speech motor control typically portray the reverse. For example, Barlow and Farley (1989) assign language processing to a very small box in an upper level of the model. Larger boxes labeled cortical fine motor control, general motor systems, brainstem vocalization system, various anatomical motor pathways, primary motor nuclei, and the musculature make up the core components of this approach to understanding speech production.

These two theoretical perspectives have very little in common, yet we know that somehow in the brain, the concepts and units of language must ultimately be translated into motor system variables. When humans speak, we produce sentences, phrases, words, and syllables that listeners understand. In order for people to speak, the brain must generate motor commands to control activation of many different motorneuron pools (the groups of neurons lying in the brain stem or spinal cord that innervate a single muscle). These motorneuron pools include those innervating muscles of the articulators, the larynx, and the chest wall. When studying the manner in which units of language might be translated into muscle contractions and movements, it is logical to ask how scientists who study the control of other human motor behaviors solve this problem, since motor control is not limited to speech. All coordinated movement requires temporal control (activations and de-activations of muscles of the right time) and spatial control (selection of the appropriate muscles to activate and the appropriate subgroups of motor units within those muscles to produce the finely graded muscle activity necessary for speech). How is temporal and spatial control achieved in other motor behaviors? What units are used to analyze other motor behaviors?

In the following sections, I consider earlier attempts to understand the organization of non-speech and speech motor behaviors on the basis of adult studies. From this review, I suggest that clues about the bidirectional mappings between linguistic units and speech
motor processes also might be gleaned from studying the development of these linkages in children. In the final section, I consider recent studies of speech motor development in relation to hypotheses about language/motor interfaces.

2. Units of action in non-speech and speech motor behaviors

There has been a great deal of investigation of a variety of cyclic motor behaviors that are essential to human life: breathing, chewing, and walking. As can be seen in Fig. 1, human chewing involves a very clear pattern of activation of muscles of the jaw. These are electromyographic recordings or EMGs (records of the electrical activity of muscle fibers within the recording field of the electrodes), which show when the muscles are activated and de-activated. In the top three traces of the figure, the activation patterns of masseter, temporalis, and medial pterygoid (MP) muscles can be seen. These three muscles are jaw-closing muscles, and they “turn on” and “off” at the same time. The bottom trace shows the activity of the anterior belly of the digastric (ABD), a jaw opening muscle. This muscle clearly has a different pattern of activation. In fact, it demonstrates a pattern opposite from the closers: its maximum activity occurs when the jaw-closing muscles are quiet. If one were asked to choose a unit of analysis for this motor behavior, clearly one would say each chewing cycle is a unit. The pattern of activation and de-activation occurs and reoccurs with each cycle. There is a clear relationship of activation among pairs of synergistic and antagonist muscles (Moore, Smith, & Ringel, 1988).

![Fig. 1. Activity of jaw-closing (MAS, masseter; TMP, temporalis; MP, medial pterygoid) and opening muscles (ABD, anterior belly of the digastric) during chewing in a normal adult subject. [Reprinted with permission from Smith (1992).]
Mastication, metabolic breathing, and walking are hypothesized to be under the control of the activity of central pattern generators. These are neural networks lying in the brain stem or spinal cord that can generate the basic pattern of muscle activity needed for a motor behavior (see the extensive discussions of central pattern generators in the accompanying papers in this volume). Clearly, these cyclic motor behaviors offer an obvious unit of analysis, which has been very useful to investigators in understanding the neural bases of these behaviors. What about control of a different kind of motor behavior—one that might be characterized as more under “voluntary” or cortically originating control?

Investigators have studied voluntarily controlled limb movements in human and non-human primates for many years. In many of these experiments, the participants are asked to reach toward a target, and the motions of various points on the limb, joint angles, and EMGs of muscles are recorded (MacKinnon & Rothwell, 2000). In these experiments, an obvious unit of analysis has emerged: a single reaching movement. The velocity profile of a reaching movement has a stereotypic pattern, reflecting the accelerating and decelerating phases of movement. The underlying muscle activity occurs in a triphasic pattern: an initial burst in the agonist muscle, followed by an intervening burst of the
antagonist muscle, and a second burst by the agonist. Fig. 2 illustrates this pattern, which has been found in many different kinds of reaching movements performed by humans and monkeys.

 Returning to the primary question of this discussion, what characterizes units for speech movements? Is there a cyclic pattern of activity as in chewing or breathing? Are there clear EMG bursting patterns related to single movements during speech? Fig. 3 illustrates the EMG recordings from the same subject and recording session as those shown during chewing in Fig. 1. If the two figures are compared (Smith, 1992), some obvious differences can be seen in the patterns of muscle activity in speech and chewing. The more visible differences in these figures have been consistently present when investigators have recorded EMGs from many different subjects (Moore et al., 1988). First, compared to chewing, there is very little activity in the masseter or temporalis muscles during speech. This is typical. The activation level of jaw muscles for speech is much lower overall compared to chewing. For this subject during speech, the medial pterygoid and anterior belly of the digastric show slightly higher levels of activity than the masseter or temporalis. Medial pterygoid and anterior belly of the digastric tend to be the major muscles for jaw opening and closing in speech. Now examining the activity of these two muscles during chewing in Fig. 1, their clearly antagonistic behavior, as described earlier, can be seen. In speech, however, readers can see that the jaw opener muscle, ABD, and closer muscle, MP, are generally co-activated, and again, this is typical across speakers. Finally, the bottom

[Graph of EMG recordings from speech and chewing]
trace in Fig. 3 shows the position of the jaw over time as the subject speaks (the total length of the record is about 4 s; downward path of the line indicates the mandible is opening). Unlike the reaching movement considered above, single opening or closing jaw movements show no clear pattern of EMG activity or bursting underlying either opening or closing movements. Rather, continuous, graded, co-activated patterns of activity are present that reveal no obvious relationship with the ongoing phase of movement.

A comprehensive review of EMG activity in articulatory, laryngeal, and oral facial muscle systems is beyond the scope of the present paper. I want to demonstrate, however, the general consistency of this co-activated pattern of EMG activity across the speech subsystems. Fig. 4 provides EMG recordings from laryngeal muscles, thyroarytenoid (TA) and cricothyroid (CT), from a normal young adult during conversational speech (Smith, Denny, Shaffer, Kelly, & Hirano, 1996). The TA is the primary muscle within the vocal fold itself, and CT is an extrinsic laryngeal muscle which is critical for pitch control. Fig. 4 clearly demonstrates that these two muscles have continuous, graded activity throughout speech. The plot shows approximately 6 s of speaking and includes a breath pause at 1.3 s into the record (see silent interval in the subject’s audio record from 1.3 to 2.2 s). Note that neither CT nor TA de-activates during the breath pause, nor do they de-activate between words (as shown by short silent intervals of the audio signal in the bottom trace). These figures demonstrate that laryngeal muscles and muscles of the jaw display continuous patterns of activity during speech that are not obviously related to movement or acoustic events.

Fig. 4. Continuous activity is characteristic of muscles during speech. In this case activity of laryngeal muscles cricothyroid and thyarytenoid continue even during a breath pause at about 1.3 s into the record. [Reprinted with permission from Smith et al. (1996).]
It is apparent from the figures I have included in this paper, some of which are from my own laboratory, that my general approach to studying speech production has been from the motor control perspective. One may then question how scientists with such a perspective have advanced knowledge about the manner in which linguistic units interface with the motor system. My answer is that, although significant progress has been made in terms of understanding the basic neurophysiology of speech motor control, science has not yet been able to link the basic principles of motor system operation to linguistic units beyond the basic assertion that some unit, usually either phonemes or syllables, is involved in the planning process.

3. Examining speech production within the frame of linguistic units

An alternative approach, taken very early in the study of speech production, is to posit the existence of linguistic units and to look for evidence of their operation in either the acoustic or the physiological output of the speaker. From this point of view, theorists have considered speech production units of varying sizes, from phonetic features, to phonemes, to syllables, and to phrase-level units (see review in Kent, Adams, & Turner, 1996). There have been many studies since the 1950s looking for evidence of the operation of units in movements, muscle activity, or the acoustic output of the speaker. Regardless of the precise experimental approach, all of these studies led to the conclusion that there is no simple mapping of phonemes, syllables, or phrases to the physiological or acoustic output. This general result reflects the ubiquitous presence of coarticulation. The phenomenon of coarticulation refers to the influence of adjacent units of speech on one another. This means that when we produce a given sound or syllable, like /p/ or /pa/, the physiological events we use to produce it and the acoustic output signature of the speaker’s output are not always the same. The physiological production events (e.g., the muscle activity for lip closure) and the acoustic output signature (e.g., the formants leading into and away from the burst) change depending on the phonemes and syllables preceding and following the segment or syllable in the particular utterance being spoken. This lack of a one-to-one mapping between the linguistic units, either phonemes or syllables, and the output variables in the physiology and acoustics of speech production make it difficult to discern what the input units might be.

One could reasonably ask how far from the target segment these coarticulatory effects extend. In other words, one might hypothesize that if one could find a boundary over which coarticulation did not occur, these boundaries might provide clues indicating units of production. This is exactly the question addressed in a classic study published by Daniloff and Moll (1968) in the late 1960s. In order to produce the distinctive acoustic characteristics of the rounded vowel /u/, the lips produce an anterior motion. Earlier studies had demonstrated that this motion could start several segments before and continue for several segments after (as determined from the acoustic signal) the /u/ in the acoustic output. Daniloff and Moll (1968) used an X-ray motion picture system, which allowed them to obtain X-ray films while people spoke. They glued metal markers on the various articulators so they could follow the separate motions of the lips, tongue tip, tongue blade, and jaw. The speakers produced a series of words and phrases all containing
the vowel /u/, such as “two” and “eaten stew,” which were embedded in meaningful sentences. By looking at the onset of the lip rounding movement for /u/ relative to other articulatory events related to the sounds surrounding the /u/, they were able to determine to what extent the lip rounding movement affected the surrounding phonemes. They found that the lip protrusion movement extended over as many as four consonants in a sequence preceding the rounded vowel /u/, and that the movement could extend over syllable and word boundaries. An example of a lip rounding motion and the timing of other articulatory events from their study is shown in Fig. 5.

One can conclude from the classic study of Daniloff and Moll (1968), as well as other investigations preceding and following it, that phonemes, syllables, and words are co-produced during connected speech. Is it possible, then, that the signals that drive muscle activity in speech are integrated over multiple levels of units simultaneously (Smith & Goffman, 2004)? In other words, in adult speakers, might there be a complex mapping of linguistic units operating at many levels, and no single unit (e.g., phoneme or syllable) can serve as the link between language formulation and speech production? On the basis of evidence from adult speakers, Smith and Goffman (2004) have proposed that this is the case. Adults are typically error-free, highly consistent speakers. They produce highly consistent speech movements across both longer and shorter units of time, from the phrase, to the word, to the phoneme levels. Adults are adaptive to perturbation; we can speak while eating. Science has learned a great deal about the adult speaker, and it is postulated that a complex mapping already exists from the language formulation networks to the speech production system. Indeed, adults may have stored speech motor commands for phrases, words, and syllables.

However, infants are clearly not born with these highly complicated, many layered mappings between language formulation networks and speech motor control systems. Furthermore, it takes years for these adult systems to develop. In a recent large-scale,
cross-sectional study of 240 children and adults, aged from 4 years to young adulthood (18–21 years), my colleagues and I discovered that speech motor development follows a very protracted time course (Smith & Zelaznik, 2004; Walsh & Smith, 2002). In terms of oral motor coordination patterns, there is still a significant increase in consistency after age 14 years. Furthermore, in terms of achieving adult-like speech rates, children are continuing to increase their rates in the late teenage years. Interestingly, these studies reveal a plateau in speech motor development from ages 7 to 12 years. Contrary to our prediction, teenage girls did not reach mature levels of speech motor coordination before boys. My colleagues and I have hypothesized that this protracted developmental course for speech motor control reflects the continuing, growing interaction of the speech motor system with the developing language systems of the brain (Smith & Zelaznik, 2004).

Thus, a reasonable strategy to adopt to learn about the operations of units that map between language and motor systems is to study the development of the speech motor system over its many years of dynamic changes. It is often suggested that the units of speech production for children are different than those used by adults—that children’s speech might operate with larger, less specified units (Kent et al., 1996). My colleagues and I have undertaken a series of studies to examine speech motor performance in typically developing children and in children with developmental speech and language disorders. The remainder of this paper focuses on that work.

4. Clues from the developing system

Infants do not start life with language and motor mappings in place; many years of learning must occur. As suggested in the DIVA model (Callan, Kent, Guenther, & Vorperian, 2000; Guenther, this volume), the speaker must develop a set of maps that include language, motor, and auditory networks. My group has not yet attempted to work with auditory targets, but we have examined speech motor output in relation to varying linguistic goals. In a series of papers, we have looked at children and adults as phrase, word, syllable, and phoneme level production goals are changed.

Our general method is to record movements of the lips and jaw with an optical movement tracking system. Fig. 6 illustrates the light emitting markers that are attached to the lips and jaw of a young subject. Markers are also attached to specially modified goggles and to the forehead so that we can track the motion of the head and correct lip and jaw movements for head motion artifact. Participants are free to move during the recording session, as long as the light emitting markers stay in view of the cameras. The system works extremely well in children as young as 3–4 years of age and has a 3-D accuracy estimated to be 0.1 mm. Children are seated in front of cameras that track the motion of the light emitting markers. They are cued either auditorily or visually to produce various linguistic stimuli. Obviously in pre-literate children, we must use auditorily presented stimuli. Typically in our experiments, we obtain 10–15 repetitions of the target word or phrase.

In our earliest experiment in this series (Smith & Goffman, 1998), we studied 4-year-old, 7-year-old, and adult speakers producing a simple phrase, “buy Bobby a puppy.” An innovative aspect of our approach was that we analyzed the articulator motion for the entire
phrase, rather than extracting single movements to make amplitude, velocity, or timing measures (Smith, Goffman, Zelaznik, Ying, & McGillem, 1995). We used a computer algorithm to reliably extract the articulator motion for the entire sentence for the 10 trials. We then normalized the 10 motion trajectories relative to time and amplitude. This procedure simply put each of the 10 productions of the utterance on a common, relative time base (0–100%), so that we were able to see how well the signals “lined up” or converged when they were all plotted on the same scale. Fig. 7 provides an example of a 4-year-old’s, 7-year-old’s, and adult’s records. The original, non-normalized data are shown in the top panel, and the normalized data in the middle panel. The bottom panel shows the standard deviation of the records as we move through relative time. In order to capture the variability in the set of 10 movement trajectories for the utterance, we computed a variability index, which in early studies we called the spatiotemporal index. Fig. 7 illustrates that the 4-year-old has a spatiotemporal index that is much higher than that of the young adult, and the 7-year-old has an intermediate value. This is a typical finding, and variability in articulator motion continues to decrease, as noted above, throughout the late childhood and teenage years. Generally based on studies like the above, we have found that compared to young adults, children are slower and much more variable in articulatory patterns until 14–16 years of age.

Using the basic methods described above, my colleagues and I extended our studies of phrase level performance in children and adults by examining the effects of increased...
linguistic demands. In these studies, we have taken simple phrases like “buy Bobby a puppy” and embedded them in longer and more complex sentences (Kleinow & Smith, 2006; Maner, Smith, & Grayson, 2000; Sadagopan & Smith, in preparation). When one examines only the target phrase portion of the utterance, children show higher movement variability when the phrase is embedded in a longer and more complex sentence compared with when it is spoken in isolation. For children, the variability of a phrase (e.g., “buy Bobby a puppy”) is higher in the embedded conditions as late as age 14 years (Sadagopan & Smith, in preparation). Again, these findings underline the protracted development of the speech motor system and point to interactions of the complexity of language goals and the nature of movement production.

Another relevant finding from such studies is the fact that adults shorten the duration of the phrase when it is embedded in a longer sentence, compared with when they speak the phrase in isolation. Very young children do not do this. The shortening of the phrase begins to appear between 7 and 9 years of age, and is well developed by 12 years of age. Thus, there appear to be “unit-like” properties of the phrase in adult speech, such that they modify its duration when the whole phrase becomes a subunit in a larger utterance. Children seem to treat the phrase differently until their speech systems mature. Future studies will explore differences between how adults and children “chunk” speech in their planning and execution processes.

As indicated above, my group’s strategy has been to examine motor output in relation to varying linguistic goals at different levels. Early in our work, we hypothesized that children
might have more primitive movement patterns that were linked more generally to phonetic goals (Goffman & Smith, 1999). In other words, given the long time course of speech motor development and the possibility that children might have less elaborated links between motor commands and specific language structures, we tested the hypothesis that adults, but not young children, would show a high degree of phonetic specificity in their speech motor output. We asked 4-year olds, 7-year olds, and young adults to produce the phrase, “Bob saw man again,” and varied the target word to start with the consonants [m, p, b, f, v]. In this study, we examined only the closing and opening movement of the lips into and away from the target bilabial consonant. Examining the data visually, it was apparent that adults had very distinctive oral movement patterns depending on the consonant target. Children, as expected, were much more variable in their close–open movement sequences, and it was difficult to visually detect whether they were showing distinctive patterns related to the varying phonetic targets.

In order to quantitatively assess potential differences between the children and adults, my colleagues and I used a statistical pattern recognition algorithm to determine whether the children’s close–open movement trajectories would sort into the five phonetic target categories as neatly as did those of the adults. Surprisingly, the children’s oral close–open movement sequences were distinctive in relation to the specific phoneme in the target word. Their waveforms were sorted by our statistical algorithm as precisely as those of the adults. In other words, children’s oral movements were reliably different for each of the five target words (e.g., “ban” versus “pan”). These results suggest that even at age four, phonetically specific mappings are beginning to develop, and they do not support the notion of a generalized primitive gesture that that is employed across consonant classes. While the phonetic specificity of their oral close open-gestures was not as obvious as it is in adults, and their movements were much more variable, children as young as 4 years were making distinctive movement patterns depending on the specific phonetic goal. My group hypothesized that the underlying EMG activity differs depending on the specific phoneme being produced in the utterance. The suggestion that differences in movement variability reflect underlying differences in variability of EMG patterns is supported by an earlier study from our laboratory (Wohlert & Smith, 2002). Orofacial EMG activations for young children show consistently higher variability, which correlated well with the movement variability observed for the various age groups. We have not yet tested the hypothesis, however, that the EMG activations for young children would show phonetically specific patterns.

Given the discussion of coarticulation in sections of this paper above, a logical question is whether children would show adult-like coarticulatory effects. If 4-year olds are already making phonetically specific movements, do these movements begin and end at the same relative time in the utterance compared with adults? Earlier studies of young children have examined this issue primarily by using acoustic data and assessing effects, over a limited time frame, of one or two adjacent syllables (Nittouer, Student-Kennedy, & McGowan, 1989). From earlier studies in adults, such as that previously described (Daniloff & Moll, 1968), we learned that coarticulatory movements can span preceeding syllable and word boundaries. Therefore, we designed an experiment that would allow a potentially maximum spread of coarticulation both before and after the target segment. We chose to examine lip rounding because it is simple to record lip movement in young children, as we
have in many earlier experiments. We tested eight individuals in each of three groups: normally developing 5-year olds, 5-year olds with specific language impairment (SLI), and young adults. Each participant produced 10–15 repetitions of the sentence, ”Mom has the ______ in the box,” produced with each of the following rounded and unrounded word pairs: goose/geese, boot/beet, moon/man. Appropriate play routines accompanied presentation of the target sentences for all groups.

There were no rounded segments present in any other portion of the sentence (”Mom has the goose in the box” and “Mom has the geese in the box” differ only by the single phoneme, which is either rounded or not rounded). Because there are no other rounded segments in the utterance, we can assume that the anterior movement of the lips (for goose, boot, or moon) or lack of it (for geese, beet, man) is incorporated into the plan for the utterance in relation to that single segment. Therefore, these stimuli allow for the potential occurrence of a very broad rounding gesture, which could in fact “spread” across the entire utterance. To enable us to determine if that potentially very broad spread occurred, we examined the movement pattern of the lips for the entire sentence, rather than for just the syllables before or after the target phoneme.

Fig. 8 shows hypothetical lip rounding motion patterns, which could be termed “broad” as shown in the top panel and “narrow” as shown in the bottom panel. The heavier dot–dash line in both panels represents motion of the lips for the entire sentence containing the rounded vowel target word (for “Mom has the goose in the box;” the maximum upward motion of the line would represent the maximum anterior motion of the lip for the rounded vowel). The lighter, solid line represents hypothetical motion of the lip for the entire

---

**Fig. 8.** Hypothetical displacement data for the upper lip. The upper panel illustrates the case in which the rounding movement of the lip is very broad, occurring over the entire duration of the sentence. The lower illustration shows a much narrower rounding movement that affects only the middle portion of the lip movement trajectory for the sentence (Mom has the goose in the box, dot–dash line; Mom has the geese in the box, solid line).
sentence containing the unrounded target word (“Mom has the geese in the box”). In the top panel, we can see the hypothetical, very broad coarticulation, such that lip motion is different throughout the entire utterance depending on whether the vowel is /u/ or /i/. In the bottom panel, we see the hypothetical case in which coarticulatory effects are narrower, occupying only about the middle one fourth of the record.

Results revealed that both groups of children, those with specific language impairment and those who were typically developing, showed the broad pattern of coarticulation as depicted in the top panel. There were differences in lip protrusion throughout the movement for the entire sentence. Again, as in earlier studies, the movement patterns of the children were much more variable than those of adults, and those of the children with SLI were equally variable to those of their typically developing peers. Despite this variability, the broad coarticulatory movements were apparent. Adults also produced broad coarticulatory movements. Their movement patterns were much more consistent, highly organized, and similar across speakers. Our statistical analysis of the amount of relative time during the whole sentence that was composed of the rounding gesture revealed that all three groups were similar with the time from onset to offset of the rounding motion, occupying 50–60% of the entire duration for the sentence.

These results indicate that, when the brain generates a motor command for the lip muscles for sentences that differ by a single phoneme (“goose” versus “geese”), the lip muscle command for the duration of the entire sentence is different depending upon the single phoneme. This is true for young children as well as adults. Thus, it appears that, even at 5 years of age, the command for the entire sentence is modified by the change in a single segment. Even in young children, there appears to be no simple mapping of units. Thus, at this young age, phrase or sentence level units appear to be part of the mapping between language processing and movement output. These results, along with our earlier study of phoneme level movement specificity, suggest that by 4–5 years of age, children are establishing multiunit language motor mappings.

My discussion thus far has centered on cross-sectional studies in which my colleagues and I examined differences in groups of young and more mature groups of speakers. In the general motor control literature, it is well established that short-term changes in motor performance can occur with practice in both children and adults (Newell, Liu, & Mayer-Kress, 2001). As children mature, they are learning new words, presumably involving new mappings to movement, and linking them to auditory and linguistic neural networks. Our studies, as well as those from many other laboratories, have consistently shown that young children’s articulatory patterns are much more variable compared to those of adults. We have suggested that this variability is beneficial because it is adaptive. If the underlying neural systems must learn new modes of behavior (new words), this variability may be a sign of more flexible organization, which would be more adaptive to learning new behaviors.

We designed a study in which the speech motor performance of 9–10-year-old children and young adults in a novel word learning task “could be examined” (Walsh, Smith, & Weber-Fox, in press). The participants in this study heard randomized lists of five novel non-words (for example, “mabfaisheib”), and produced the word in response to the auditory stimulus. Just as in our earlier studies, we obtained 10–15 repetitions of each of
the novel non-words. We used the same kind of movement variability index that we employed in our earlier studies, and we computed a movement variability index for the early versus the late trials. In other words, for each subject we recorded 10 trials, and we computed separate movement variability indices for the first 5 and the last 5 trials. We only included trials in the analysis in which the subject produced the non-word fluently and without errors.

We hypothesized that the children would show a learning effect, such that the later trials would be more consistent than the earlier trials. The data clearly supported this hypothesis. The young children showed a learning effect, such that their movement variability was lower on the later trials and their word durations were shorter. Thus, the 9 and 10-year olds were systematically changing their motor plans and becoming more consistent and faster during the course of the 20-min experimental session. Adults showed much less overall movement variability than the children, and they showed no change from early to late trials. The adult participants had never heard these words before, yet they produced very consistent patterns of movement from the 1st to the 10th production. My colleagues and I interpreted these results to suggest that adults have highly stable coordinative synergies (collectives of muscles linked in coordinative control) that they can employ even when producing novel words. Here we observed a clear, qualitative difference between the performance of children and adults. This study has interesting implications for neural plasticity in the speech motor system in children and adults and for the learning of second languages.

5. Conclusion

Our ideas regarding language and motor interactions have evolved with the challenges of these new findings. First, we expected that young children would show dramatic differences in their speech motor performance compared to adults. The data strongly supported this assertion. Children, even into their teenage years, produce much noisier, less reliable, slower speech movements compared with young adults (Walsh & Smith, 2002). Children are quantitatively less consistent in their movement output compared to adults.

In terms of language and motor interfaces, we are also interested in qualitative differences. Distinctiveness in the nature of the performance, for example, may suggest that children use a different organizational unit or a different planning strategy. My colleagues and I have been surprised that some of the qualitative differences we expected to find have not been supported. For example, our studies suggest that 4 and 5-year-old children organize speech motor commands specifically around phonetic goals (Goffman & Smith, 1999), phrase level goals (Maner et al., 2000), and sentence level goals (Goffman, Smith, Heisler, & Ho, in preparation). Thus, we conclude that bidirectional linkages between language and motor systems are occurring at multiple levels already in preschoolers. We have, though, observed a number of qualitative differences in the performance of children and adults. There are relative timing differences in sentence production for young children and adults, with a tendency for young children to treat embedded phrases with less “unit-like” status. As we noted above, it appears that children
may “chunk” speech subunits differently in the planning the production process. Children are also qualitatively different from adults in terms of short-term motor learning in the production of novel non-words.

As we discuss qualitative and quantitative differences between children and adults, it is important to note that this review has focused on children 4 years and older. The language motor interface must also be explored in infants as they begin to babble and in toddlers as they begin to produce words. Other labs are beginning this methodologically challenging work (Green, Moore, Higashikawa, & Steeve 2000). It seems reasonable to hypothesize that in earlier development, syllable and word level units are dominant, and studies of coarticulation in babbling would be informative. Fig. 9 summarizes a hypothetical developmental course for various units that might be used by the brain to “translate” between language and motor systems. Again, auditory target space is included, because based on the effects of hearing impairment on speech development, we know that auditory mappings are essential for speech motor learning.

Finally, I must return to my starting point, which was the gap between models of language processing/production and models of speech motor control. It is clear that studies of the development of language production force us to bridge the gap between the two types of models. Children must develop the complex, multilayered mappings that adults use with such apparent ease. The experimental data have revealed that very precise details of the motor commands that drive muscle activity are sculpted by details of the linguistic units being produced. By 4–5 years, this multilevel sculpting appears to be taking place. In an earlier paper (Smith & Goffman, 2004), my colleague and I argued that not only do linguistic goals shape motor commands, but preferences and features of the motor system shape linguistic processes, as well. The influences and linkages between language and motor systems, we suggested, are bidirectional, rather than top–down from language to

![Fig. 9. An conceptual model for the change in mappings of linguistic units to movement and auditory space over the course of development.](image)
motor networks. This argument is supported by the fact that children with specific language impairment show speech motor performance delays and differences and in their ability to modulate the speech motor system output to achieve specific linguistic goals, such as prosodic targets (Goffman, 1999). Similarly, speakers with chronic developmental stuttering show differences in the language organization of their brains during reading tasks (Weber-Fox, 2001).

The field of speech–language pathology needs a comprehensive, theoretical account of how the human organism communicates ideas through linguistic structures and ultimately through muscle contraction and movement in order for professionals to provide the best diagnosis and treatment of childhood speech and language disorders. The recent advances described in this paper are meant to point the way toward a more enlightened theory by exploring the development of these linkages across the lifespan.

Acknowledgements

The work from our laboratories reviewed in this paper has been supported by the National Institutes of Health, National Institute on Deafness and Other Communicative Disorders, Grants DC00559, DC02527, and DC04826. I am especially indebted to Lisa Goffman, who has been a wonderful source of information and inspiration for our work on development. Finally, thanks to Janna Berlin and the graduate students who have helped to collect and analyze the data over the years.

Appendix A. Continuing education

1. The primary explanatory variables involved in models of motor speech control include:
   (a) sensorimotor integration and central pattern generators;
   (b) phonology and syntax;
   (c) syllables, prosody, and phonemes;
   (d) all of the above.
2. Linguistic units can refer to:
   (a) syllables;
   (b) phonemes;
   (c) phrases;
   (d) all of the above.
3. How does the effect of coarticulation influence an experiment’s ability to determine the role of linguistic units in speech production?
   (a) Coarticulation makes this determination easier because the phonemic change in the acoustic signal signifies the end of a unit.
   (b) Coarticulation does not affect this determination.
   (c) Coarticulation confounds this determination, as it results in a lack of consistent mappings between linguistic units and physiologic/acoustic output signals.
   (d) Coarticulation confounds this determination by creating a 1:1 mapping between linguistic units and physiologic/acoustic output signals.
4. From a motor speech perspective, adult speakers are different from child speakers in that adults:
(a) have a larger vocabulary;
(b) are more consistent in their productions;
(c) may have stored speech motor commands for phrases, words, and syllables;
(d) (b) and (c).

5. In terms of developing oral motor coordination patterns, there is still a significant increase in consistency after age:
(a) 12;
(b) 14;
(c) 18;
(d) 21.

Answers: 1 (a); 2 (d); 3 (c); 4 (d); 5 (b).

References


