Using the Self-Select Paradigm to Delineate the Nature of Speech Motor Programming

David L. Wright,
Texas A&M University

Don A. Robin,
The University of Texas Health Science Center at San Antonio

Jooyhun Rhee,
Texas A&M University

Amber Vaculin,
Texas A&M University

Adam Jacks,
The University of Texas Health Science Center at San Antonio

Frank H. Guenther, and
Boston University

Peter T. Fox
The University of Texas Health Science Center at San Antonio

Abstract

Purpose—The authors examined the involvement of 2 speech motor programming processes identified by S. T. Klapp (1995, 2003) during the articulation of utterances differing in syllable and sequence complexity. According to S. T. Klapp, 1 process, INT, resolves the demands of the programmed unit, whereas a second process, SEQ, oversees the serial order demands of longer sequences.

Method—A modified reaction time paradigm was used to assess INT and SEQ demands. Specifically, syllable complexity was dependent on syllable structure, whereas sequence complexity involved either repeated or unique syllabi within an utterance.

Results—INT execution was slowed when articulating single syllables in the form CCCV compared to simpler CV syllables. Planning unique syllables within a multisyllabic utterance rather than repetitions of the same syllable slowed INT but not SEQ.

Conclusions—The INT speech motor programming process, important for mental syllabary access, is sensitive to changes in both syllable structure and the number of unique syllables in an utterance.

Keywords

speech; motor response; memory; motor programming

Levelt and Wheeldon (1994) described a number of processing stages that occur prior to the generation of speech. The initial preparation of the appropriate concept is followed by lexical selection, phonological encoding, phonetic encoding, and, ultimately, articulation. At the phonetic level of speech preparation, van der Merwe (1997) makes an explicit distinction between motor programming and execution stages prior to the delivery of speech. According to Nijland, Maassen, and van derMeulen (2003), speech motor programming involves the translation of phonetic plans characterizing the spatial and temporal goals of an articulatory gesture into context-dependent motor specifications that are forwarded to the articulators to be implemented during the motor execution stage. Speech motor programming is important clinically because it has been identified as a potential candidate process in a number of speech disorders such as childhood apraxia of speech (Nijland, Maassen, van der Meulen, Gabreëls, Kraaimaat, & Schreuder, 2003) and adult apraxia of speech (Ballard, Granier, & Robin, 2000; Deger & Ziegler, 2002; McNeil, Doyle, & Wambaugh, 2000; McNeil, Robin, & Schmidt, 1997). Theoretically, it has been suggested that motor programming lies at the critical, but poorly understood, interface between psycholinguistic models of phonological encoding and phonetic theories of motor control in speech (Deger & Ziegler, 2002).

Evaluation of the importance of motor programming to the advance planning of sequential actions such as those commonly observed during speaking, playing musical instruments, typing, or even handwriting has relied extensively on work addressing upper limb and finger sequencing tasks (e.g., Sternberg, Monsell, Knoll, & Wright, 1978; Verwey, 1999; Wright, Black, Immink, Brueckner, & Magnuson, 2004). However, an early, but now classic, examination of motor programming focused, at least in part, on speech production. Sternberg et al. (1978), using a simple reaction time paradigm, reported a systematic impact of sequence length on the initiation of the first utterance as well as on total execution time of a speech sequence. These data contributed to the delineation of the subprogram retrieval model, which has significantly influenced understanding of the preparation of sequential behavior. Subsequently, Klapp (1995, 1996) provided a more contemporary detailing of the motor programming process during sequence production stemming from an ongoing debate focused on the appropriate reaction time (RT) methodology to capture and describe the specific demands of motor programming. Sternberg and colleagues used a simple RT paradigm (Henry & Rogers, 1960; Sternberg et al., 1978), whereas others have advocated the use of a choice RT paradigm to examine the implementation of motor programming (Klapp, Wyatt, & Lingo, 1974; Klapp et al., 1979). Klapp (1995, 1996) reported that the predominant effect associated with motor programming when using a choice RT paradigm is from the manipulation of the duration of a single element response (e.g., Klapp et al., 1974; Klapp, McRae, & Long, 1978). For simple RT, manipulating sequence length seems to be more influential (e.g., Sternberg et. al., 1978).

To account for data gathered from different RT paradigms, Klapp (1995) suggested that motor programming consists of two independent processes. One process, labeled INT, is

responsible for organizing the internal features of the programmed units within a sequence. The complexity of the basic unit being programmed determines the length of time to complete the INT process. Thus, a programmed unit of longer duration, for example, would be associated with a greater amount of time spent on INT. A second process, called SEQ, is involved in sequencing each of the programmed units into the correct order for output. The time to complete the SEQ process depends on the number of programmed units that must be placed in order. There are a number of important assumptions regarding the INT and SEQ processes central to Klapp's model. First, the INT process can be prepared in advance (i.e., preprogrammed), whereas the SEQ process cannot be preprogrammed. In circumstances in which the INT process is not preprogrammed, the INT and SEQ processes occur in parallel with the INT process taking longer to complete than the SEQ process.

Returning, then, to earlier data using a simple RT paradigm, the INT process for both the short and long duration responses is preprogrammed because the participant is aware of the required response resulting in simple RT being uninfluenced by response duration. In the case of choice RT, the INT process must occur after the receipt of the "GO" signal. Choice RT is greater when performing a long as opposed to a short duration response reflecting the greater demands on the INT process for the former. Findings such as those of Sternberg et al. (1978), using speech responses in which sequence length impacted simple RT, can also be accounted for by Klapp's model. Specifically, despite the fact the simple RT paradigm affords the performer the opportunity to preprogram the INT process, the SEQ process cannot be conducted until the receipt of the start signal. Simple RT, then, will reflect the time to complete SEQ, which in turn is dependent on the number of elements that must be prepared. For the choice RT conditions, both the INT and SEQ processes must occur after the presentation of the imperative signal. Klapp's model assumes, in this case, that INT and SEQ occur in parallel and that the INT process takes longer to complete. Thus, in this case, it is the INT and not the SEQ process that determines choice RT.

Klapp (1995), using a timed finger sequencing task, revealed both patterns of results (i.e., the differential impact of response duration and sequence length on choice and simple RT, respectively) using procedures that allowed an a priori specification of the particular conditions under which each pattern of findings should emerge. More recently, Klapp (2003) provided additional support for his two-process motor programming account in the speech domain by conducting a series of experiments in which individuals were required to articulate pseudowords. In Experiment 1, Klapp revealed that choice but not simple RT was sensitive to the number of syllables in the word. He argued these data were congruent with the idea that the INT process was more demanding in the multisyllabic pseudoword. In this case, Klapp assumed that the pseudoword was represented and prepared as a single chunk and that the number of syllables within the chunk contributed to its internal complexity (see Klapp, Anderson, & Berrian, 1973). Klapp contended that the pseudowords in Experiment 1 were likely treated as a single chunk because the participants were encouraged to articulate words as quickly as possible. Moreover, despite the use of pseudowords, individuals were only required to articulate a few exemplars throughout the experiment, so they were assumed to be quickly learned. In a number of subsequent experiments (Experiments 2–4; Klapp, 2003), Klapp created situations in which similar pseudowords were spoken, but the syllables contained within the words were planned separately rather than as a chunk. This

was accomplished, for example, by changing the speech rate. In such cases, it was simple RT that was influenced by the number of syllables and not choice RT, which was interpreted to reflect a shift in the programmed unit from the word to the syllable level.

Unfortunately, support for Klapp's model for motor programming using simple limb (Klapp, 1995) and speech responses (Klapp, 2003) has relied almost exclusively on cross-experiment (i.e., simple and choice RT) and between subject comparisons to describe the role of the INT and SEQ processes. To overcome this problem, Immink and Wright (2001) used a self-select paradigm that allowed the temporal delays associated with the INT and SEQ processes to be assessed concurrently for the same individual within a single experimental protocol (for adaptation of this protocol used in Experiments 1 and 2, see Figure 1). In the self-select paradigm, a trial begins with the "ready" signal being presented in the center of the computer display. Shortly afterwards, the individual is provided information that pertains to a response that should be prepared. In Figure 1, a multi-element speech response, "baaabababaaa," is required, in which the middle two /ba/s were shorter in duration than the /ba/ in Positions 1 and 4. Study time (ST) is defined as the interval between the presentation of a pre-cue (i.e., "4L") informing the participant of the nature of the upcoming response and the indication (pressing the "END" key in the example provided in Figure 1) that the participant has readied the required response. This interval is assumed to capture the temporal demands of the INT motor programming process as described by Klapp (1995). Following the depression of the END key, a variable foreperiod occurs prior to the presentation of the imperative "GO" signal. Reaction time (RT) is defined as the interval between the presentation of the "GO" signal and the depression of the appropriate response key or the initiation of speech, in the present case. The RT interval is assumed to capture the temporal demands of the motor programming process labeled SEQ. In keeping with Klapp's (1995) model, Immink and Wright (2001) revealed that a manipulation of the duration of a single element response had a significant influence on ST, whereas sequence length had its primary influence on the RT.

The self-select paradigm was used in some preliminary pilot work to address motor programming during speech production. For this work, we merely tried to mirror the conditions used in previous work by Immink and Wright (2001) with finger sequencing tasks. Specifically, individuals articulated monosyllabic responses (e.g., the CV /ba/) for the duration of 150 or 450 ms and assumed that the latter would introduce relatively greater INT process demands that should be reflected in a longer mean ST. Alternatively, we compared the preparation of monosyllabic and multisyllabic responses, the latter of which involved a combination of the aforementioned short and long duration CV /ba/s. In this case, we anticipated an increase in RT due to the longer SEQ processing needed for multisyllabic responses.

As one might anticipate, the participants in this pilot work were very capable of producing the utterance with correct durations and sequential constraints. Congruent with previous findings in the speech domain (Sternberg et al., 1978), as well as with sequential limb responses, increases in sequence length slowed preparation. More specifically, as expected, there was an increase in mean RT when more syllables had to be prepared. These findings have been interpreted as support for a more extensive search of the motor buffer that must

occur before releasing a syllable for articulation and is described as part of the SEQ process in Klapp's (1995, 2003) two-process account of motor programming. Mean ST for the short and long single utterances of /ba/, however, was essentially the same. This was unanticipated and inconsistent with our initial expectations based on Klapp's model. One possibility is that the speech apparatus is unique in its planning and programming demands and, as such, will operate in a different manner than observed for other effectors. It is also possible that adjusting duration of the unit being programmed (i.e., the syllable, in this case) is quite trivial for a native speaker, and, as such, the expected influence at the level of INT was not present.

In the present experiment, we conducted a more formal assessment of speech production using the self-select paradigm, and in doing so, we adopted manipulations of syllable complexity (INT load) and sequence complexity (SEQ load) that have been used by previous speech researchers (Bohland & Guenther, 2006; Riecker et al., 2000). More specifically, the primary impetus for this approach was work by Bohland and Guenther that evaluated the cortical and subcortical neural substrates critical to organizing and articulating sequences of simple speech sounds. Importantly for the present experiment, the utterances used by Bohland and Guenther were categorized on the basis of two unique features. First, syllable complexity was characterized as being either (a) simple consisting of a single consonant and a vowel (CV) or (b) complex involving a consonant cluster (CCCV or CCV) followed by a vowel. Within the framework of Klapp (1995, 2003), preparing to articulate a CV or CCV should create different demands for the INT process and would lead to an increase in mean ST for the complex version. Second, sequence complexity was defined as either (a) a simple sequence that involved three repetitions of the same syllable or (b) a complex sequence composed of three unique syllables. Thus, rather than use the number of syllables as the factor determining the demand of SEQ, Bohland and Guenther assumed that the presence of unique elements would be a candidate feature that would increase the demands involved in articulating syllables in the correct serial order. Although this may be the case, in addition to including this feature in the present experiment, we also included the sequence length component to specifically target the SEQ process in a manner consistent with our earlier finger-sequencing work and Klapp's model.

Figure 2 provides an overview of the experimental conditions used in the experiment. We considered the crucial INT manipulation to involve the comparison of mean ST between uttering a simple CV and complex CCCV syllable. Mean ST should be greater for the more complex item. Importantly, in addition to the influence of syllable structure on ST, we also anticipated a lack of impact of this manipulation on mean RT, since we were still dealing with a single programmed unit in each of these cases. Examination of SEQ was considered in two separate ways. First, we examined the difference between mean RT when preparing a single CV or CCCV compared to the case in which either of these syllables had to be repeated consecutively. Mean RT would be greater when more than one syllable had to be spoken. Second, we focused on the feature that the work of Bohland and Guenther (2006) implicitly assumed to be important to the implementation of SEQ. That is, we compared mean RT for the CCCV condition that included different syllables (referred to as $CCCV_{unique}$ in Figure 2) and CCCV condition that has repetitions of the same syllable (referred to as $CCCV_{repetition}$ in Figure 2). In this case, if the presence of unique syllables

adds additional demands to the SEQ process beyond that incurred from planning extra syllables, then this would be manifest as a further increase in mean RT beyond that observed for the multiple repetition condition—that is, $CV_{repetition}$ or $CC(C)V_{repetition}$ (see Bohland & Guenther, 2006; Deger & Ziegler, 2002).

Method

Participants

Twelve undergraduate students at Texas A&M University participated in this experiment in order to fulfill a class requirement. All participants were native English speakers and self-reported being free of neurological disease. Prior to participation in the experiment, all individuals read and signed an informed consent. None of the participants in this experiment participated in previously described pilot work.

Apparatus

All experimental procedures were administered and data acquisition was completed using an IBM-PC compatible microcomputer with a standard XGA color monitor in combination with a set of stereo amplified speakers, and an Audio-Technica ATM75 condenser microphone. All phases of the experiment were programmed using LabVIEW 6 (National Instruments; Austin, TX).

Tasks

The tasks required the participants to articulate either a monosyllabic or multisyllabic utterance. Two separate monosyllabic utterances—one with a CV (simple) structure (/ta/) and one with the CC(C)V (complex) structure (/stra/)—were used. Each utterance was paired with an utterance-specific cue that allowed us to pre-cue the required utterance. In this case, /ta/was labeled "1T," and the /stra/ was associated with "1S." Two multisyllabic utterances were also used that involved repetition of the aforementioned monosyllabic responses. One labeled "4T" involved repeating /ta/ four times, whereas "4S" consisted of repeating /stra/ the same number of times. Finally, two additional multisyllabic utterances were used that included the syllable /ru/ to force the participant to encounter more unique syllables within a sequence. One was labeled "4Y" and consisted of the sequence "ta-ru-stra-ta," and the other was associated with "4X" and involved the sequence "ta-stra-ru-ta." Participants were not provided explicit temporal requirement for the production of the syllables used in this experiment. Nonetheless, we expected the participants to exhibit temporal structures similar to those of the models used at the beginning of the experiment and as feedback following each trial.

Procedure

All participants were provided both written and verbal instructions of the sequence of events that occurred within each trial of the experiment as well as specific information regarding the requirement for each speech task. Participants were explicitly informed of the general course of events in the self-select paradigm used by Wright and colleagues (Immink & Wright, 2001; Kimbrough, Wright, & Shea, 2001; Wright et al., 2004; see Figure 1). Prior to any practice, each individual was presented with a series of notecards with the labels (e.g.,

"4X") printed on the front and were required to immediately articulate the associated utterance. Each participant worked through the notecards until they completed three consecutive trials without an error in associating the label with the correct utterance.

Each trial was preceded by the presentation of four squares oriented horizontally across the computer display. The word "READY" appearing in the left-most box indicated to the participant that a trial was about to begin. After a variable interval between 500 and 2000 ms, one of the symbols (1S, 1T, 4S, 4T, 4X, 4Y) associated with the speech tasks appeared in the next box. The participants were instructed to use this information to prepare the identified speech task as "best they could" prior to the presentation of the imperative "GO" signal. Each individual was also informed that they should take as long as they felt necessary to completely organize the required speech response and, when ready, they should "left-click" the mouse. The temporal interval between the presentation of the task symbol and the mouse click was defined as study time (ST). Another variable interval between 800 and 1,200 ms was initiated, after which the imperative "GO" signal appeared in the third box. Participants were instructed to articulate the required speech task (indicated by the task symbol previously presented) as soon as possible after the presentation of the "GO" signal. The temporal duration between the presentation of the "GO" signal and the initiation of the first syllable was defined as reaction time (RT).

Prior to training, all participants listened to four presentations of each of the to-be-practiced speech tasks. On completion of the presentation of the speech tasks, individuals were provided with a chance to ask questions regarding the speech tasks and/or the general procedures that were outlined in the instructions. All participants were administered four blocks of 24 trials for a total of 92 speech responses. Each block of 24 trials included a random presentation of four trials for each of the six specific tasks described in the Tasks section. Each participant was encouraged prior to the beginning of each block to use the ST interval to "best" prepare the upcoming speech task so that a quick and accurate response could be made following the presentation of the imperative "GO" signal. To facilitate the production of each speech task, 1 s after the speech task was complete, the appropriate model for that task was replayed. This was the same model sound that was provided to the participant prior to the beginning of practice. The participant was encouraged to compare his or her response with the modeled behavior with the intent of matching the responses.

Data Analysis

Each speech response was recorded using an Audio-Technica ATM75 condenser microphone that was fitted to the participant's head with a padded, adjustable headset prior to participation such that a constant mouth-to-microphone distance of about 5 cm on the left side was established. Raw speech signals were later transformed into smooth and rectified speech wave envelopes (see Figure 1). Speech signals were rectified at the mean amplitude of the signal, and all the negative amplitude values were converted into positive values. Rectified speech signals were passed through a Butterworth low pass filter. Three dependent variables were calculated on the basis of these data. First, recall that RT was defined earlier as the latency between the presentation of the imperative "GO" signal and the initiation of the first syllable in a speech task. The determination of speech initiation was based on a

previous approach used by Deger and Ziegler (2002), in which the beginning and end of a syllable was defined as 30% of the min-to-max threshold of the SPL contour. Although the approach to identify the beginning and end of a syllable was similar in the present study, speech initiation or termination was operationally defined as 15% of the rising or falling SPL contour, respectively. Thus, RT was the onset latency between presentation of the "GO" signal and 15% of the rising SPL contour for the first syllable. Syllable duration was determined by taking the temporal delay between 15% of the rising and falling SPL contour for each syllable contained in the speech task. Finally, for all four-syllable tasks, intersyllable duration was computed as the time between 15% of the falling SPL contour and 15% of the rising SPL contour of the next syllable (see Figure 1).

Results

Performance of Speech Tasks

Because there was no pre-determined temporal requirement for the speech tasks in this experiment, we calculated and formally analyzed the mean and variability of the total utterance duration for each of the speech tasks. These data are presented in Table 1. Formal analyses of these data consisted of a 6 (task: 1T, 1S, 4T, 4S, 4X, 4Y) × 4 (block: 1–4) analysis of variance (ANOVA) with repeated measure on both factors. The analysis of mean utterance duration revealed a significant main effect of task, $F_{(5, 55)}$[1] = 782.93, $p < .01$, n_p^2 = .99. The main effect of block, $F_{(3, 33)} = 0.50$, $p = .68$, and the Task × Block interaction, $F_{(15, 165)} = 0.61$, $p = .86$, were not significant. As would be expected, the main effect of task resulted from a shorter mean utterance duration for the 1T (M= 299 ms, standard error of the mean [SEM] = 6 ms) and 1S (M = 311 ms, SEM = 9 ms) responses compared with 4T (M = 1,788 ms, SEM = 28 ms), 4S (M = 1,882ms, SEM=28ms), 4X(M= 1,827ms, SEM=31ms), and 4Y (M = 1,848 ms, SEM = 27 ms). It is important to note that post hoc tests indicated that the mean total utterance duration for the monosyllabic responses did not differ reliably. This was also true for the multisyllabic responses.

Analysis of the variability in total utterance duration revealed a main effect of task, $F_{(5, 258)} = 6.70$, $p < .01$, $n_p^2 = .11$. Variability in the total utterance duration was lower for the 1T (M = 34 ms, SEM = 5 ms) and 1S (M = 40 ms, SEM = 6 ms) responses compared with 4T (M = 95 ms, SEM = 13 ms), 4S (M = 95 ms, SEM = 10 ms), 4X (M = 106 ms, SEM = 21 ms), and 4Y (M = 108 ms, SEM = 15 ms). Again, post hoc tests also indicated that the variability in total utterance duration for the monosyllabic responses did not differ reliably, which was also the case when comparing the multisyllabic responses. The main effect of Block, $F_{(3, 258)} = 1.34$, $p = .26$, and the Task × Block interaction, $F_{(15, 258)} = 1.28$, $p = .21$, were not significant.

Impact of syllable complexity on STand RT—Table 2 displays mean ST and RT as a function of each speech task as a function of training block. The 2 (complexity: simple, complex) × 4 (block: 1–4) ANOVA with repeated measures on both factors was conducted on mean ST and RT. For ST, this analysis revealed significant main effects of complexity,

---

[1]For all analyses, we assumed no violations of sphericity.

$F_{(1, 88)} = 4.06$, $p < .05$, $n_p^2 = .04$, and block, $F_{(3, 88)} = 13.05$, $p < .01$, $n_p^2 = .31$. Mean ST was lower for the simple monosyllabic response (M= 863 ms, SEM = 58 ms) compared with the complex monosyllabic response (M= 1,058 ms, SEM= 99ms). In addition, mean ST was significantly higher for Block 1 (M = 1,465 ms, SEM = 167 ms) compared with that observed for Block 2 (M= 922 ms, SEM = 72 ms). Mean ST for Block 2 was, in turn, higher than the mean ST for Block 3 (M = 760 ms, SEM= 45 ms). However, mean ST for Blocks 3 and 4 (M= 1,019 ms, SEM = 88 ms) did not differ reliably. The Complexity × Block interaction for mean ST, $F_{(3, 88)} = 1.13$, $p = .34$, was not reliable. The analysis of RT failed to reveal significant main effects of complexity, $F_{(1, 88)} = 2.50$, $p = .12$, and block, $F_{(3, 88)} = 0.69$, $p = .56$, as well as the Complexity × Block interaction, $F_{(3, 88)} = 0.47$, $p = .70$.

Impact of the sequence complexity from increasing the number of syllables on ST and RT—The 2 (complexity: single, multi) × 4 (block: 1–4) ANOVA with repeated measures on both factors was conducted on mean St and RT. For ST, the main effect of complexity, $F_{(1, 11)} = 10.02$, $p < .01$, $n_p^2 = .99$, and block, $F_{(3, 33)} = 19.07$, $p < .01$, $n_p^2 = .07$, were reliable. The Complexity × Block interaction, $F_{(3, 33)} = 2.68$, $p = .06$, $n_p^2 = .09$, was marginally significant. The complexity main effect revealed a larger mean ST for the multisyllable utterance (M= 1,039ms, SEM = 70 ms) compared with the monosyllabic response (M = 958 ms, SEM = 68 ms). The block main effect was a function of mean ST being greater for Block 1 (M = 1,452 ms, SEM = 124 ms) compared with Block 2 (M = 1,026 ms, SEM = 71 ms), which in turn was higher than the mean ST for Block 3 (M = 824 ms, SEM = 51 ms). Mean ST for Block 3 did not differ reliably from Block 4 (M = 693 ms, SEM = 45 ms). The Complexity × Block interaction revealed that the difference in mean ST between themulti- and monosyllabic responses was restricted to Blocks 2 and 3. These data are displayed in Figure 3.

For RT, the main effect of complexity, $F_{(1, 88)} = 2.87$, $p = .09$, $n_p^2 = .03$, was marginally significant. The mean RT for the monosyllabic responses (M = 664 ms, SEM = 23 ms) was lower than the RT reported for the multisyllabic response that included syllable repeats (M = 725 ms, SEM = 28 ms). The main effect of block, $F_{(3, 88)} = 1.67$, $p = .17$, and the Complexity × Block interaction, $F_{(3, 88)} = 0.23$, $p = .87$, failed significance.

Impact of the sequence complexity from preparing unique syllables on ST and RT—Using data for the multisyllabic responses only, a 2 (sequence: unique, repetition) × 4 (block: 1–4) ANOVA with repeated measures on both factors was conducted on mean ST and RT. For ST, this analysis revealed reliable main effects of sequence, $F_{(1, 11)} = 18.0$, $p < .01$, $n_p^2 = .62$, and block, $F_{(3, 33)} = 51.81$, $p < .01$, $n_p^2 = .82$, as well as a significant Sequence × Block interaction, $F_{(3, 33)} = 5.19$, $p < .01$, $n_p^2 = .32$. Mean ST for each sequence as a function of block is displayed in Figure 4. The Sequence × Block interaction indicated that mean ST was higher when unique syllables were involved as opposed to merely repeating the same syllable for all blocks. In addition, mean ST was reliably reduced across all blocks for the utterances that included unique syllables. However, in the case of sequences involving repetition of the same syllable, mean ST only differed reliably between the first and last trial blocks. For RT, the main effect of block, $F_{(3, 88)} = 2.27$, $p = .09$, $n_p^2 = .07$, was marginally significant. Mean RT was gradually reduced from Block 1 (M= 817

ms, SEM=39ms) through Block 4 (M = 677 ms, SEM = 45 ms). The main effect of sequence, $F(1, 88) = 0.34$, $p = .56$, and the Sequence × Block interaction, $F(3, 88) = 0.21$, $p = .89$, were not significant.

Discussion

Speech motor programming was examined using a recent model forwarded by Klapp (1995, 1996, 2003) as a theoretical backdrop that was derived primarily for sequential behaviors not involving the speech apparatus. As opposed to using separate simple and choice RT methods commonly adopted to address such issues (e.g., see Deger & Ziegler, 2002; Klapp, 2003), we incorporated the self-select paradigm. This approach has been used to simultaneously differentiate key motor programming processes described by Klapp (i.e., INT and SEQ). Individuals articulated utterances that differed with respect to the number of syllables and syllable complexity. Prior to considering the speech motor programming per se, it is important to note that the particular speech responses used in this experiment were performed in the manner that one might have anticipated. That is, monosyllabic responses were of shorter duration and had less variability than the multisyllabic responses. In the latter case, whether multiple syllables were repetitions or included different syllables, mean duration and variability tended to be very similar.

Moving, then, to the planning of these utterances, we examined whether the structure of the syllable influenced the temporal delays in preparing monosyllabic utterances. This question focused on the INT process of Klapp's model because it was assumed that the syllable was the fundamental unit being programmed when preparing these responses (see Bohland & Guenther, 2006; Levelt & Wheeldon, 1994). Concentrating on the syllable seemed appropriate, given that a number of previous researchers have considered the syllable important as a fundamental unit of speech production (see Bohland & Guenther, 2006; Cholin, Levelt, & Schiller, 2006; Ferrand & Segui, 1998; Guenther, Ghosh, & Tourville, 2006; Sevald, Dell, & Cole, 1995; Ziegler & Maassen, 2004).Moreover, as was the case in Bohland and Guenther (2006), the methods used for stimulus presentations and instructions were based on the syllable that likely contributed to the participants, considering the syllable as the key organizational unit.

In this experiment, we evaluated the influence of the CV structure for planning the constituent syllables (Bohland & Guenther, 2006). Specifically, we used /ta/ and /stra/ as exemplars for simple and complex syllables based on previous complexity delineations (Bohland & Guenther, 2006; Dogil et al., 2002; Riecker et al., 2000). For example, Dogil et al. (2002) indicated that both syllable types comply with appropriate phonotactic rules of English but differ systematically in phonetic gestural complexity in that /ta/ is a model CV syllable, whereas /stra/ has a complex onset. The results indicated that complexity of the syllable did indeed exert the predicted effect on the INT process manifested as an increase in mean ST for the more complex monosyllabic /stra/ than the simpler /ta/. With respect to Klapp's model in the context of the self-select paradigm, the influence of syllable complexity was not present for RT. This profile is congruent with the claim that features of speech motor programming that exert an influence on INT, such as the CV structure of the syllable in this case, can be implemented prior to the imperative signal, or pre-programmed.

We assume that this finding results from the greater demand associated with loading a more intricate syllabic gestural score from the mental syllabary into the motor buffer when more complex syllables are part of the upcoming utterance.

Increasing the number of syllables in the utterance resulted in a concomitant increase in mean RT. This finding was consistent with our data in our preliminary pilot work as well as in the extensive literature on the sequence length effect for both speech (Klapp, 2003; Sternberg et al., 1978) and nonspeech sequential behaviors (Verwey, 1999). The effect of sequence length has been attributed to the SEQ process in Klapp's account, which is responsible for delivering the programmed units to the output apparatus in the correct serial order. Thus, for the multisyllabic responses articulated in the present experiment, whether utterances consisted of repetitions of the same syllable or a group of different syllables, separate gestural scores for each syllable were read from the motor buffer to the articulatory apparatus.

The present work also addressed Bohland and Guenther's (2006) assumption that an important consideration when preparing multisyllabic responses is the number of unique syllables in the utterance; specifically, repeating the same syllable three times would impose less demand than producing an utterance consisting of a number of unique syllables in sequence. Deger and Ziegler (2002), however, revealed the typical sequence length effect for simple RT when articulating syllable chains of varying length and varying segmental content but failed to report any impact on simple RT and/or increased intersyllable intervals for speech sequences that involved alternating syllables. Accepting, for the moment, that some cost is incurred when preparing utterances made up of different syllables, we assumed that part of this demand would be housed at the level of the SEQ process because such articulations would presumably require particular attention to the serial order at output. This would be particularly true if the utterance was organized as separate items, which appears to have been the case, given the emergence of the aforementioned sequence length effect for RT in both the CCCV$_{unique}$ and CCCV$_{repetition}$ conditions. We considered this issue by evaluating whether articulating sequences of syllable repetition or unique syllables influenced SEQ. In the self-select paradigm, we expected that an impact of preparation of unique syllables on SEQ would be manifest as an additional increase in mean RT, beyond that incurred by programming more syllables, for the CCCV$_{unique}$ compared with the CCCV$_{repetition}$ conditions. This was not the case. Mean RT for each of the four syllable conditions was similar and greater than the mean RT for the monosyllabic responses. Thus, although preparing more syllables taxes the SEQ process, the process was not sensitive to the nature of the items.

Although organizing different syllables within an utterance imparted no influence on SEQ, this was not the case for INT. There was a sizable impact of preparing different syllables on this process reflected in an increase for mean ST. In the case of unique syllables being prepared, it is not difficult to envision that the INT process must be separately employed for each unique item that must be activated within the mental syllabary. Thus, it is not the number of syllables per se, as is the case for SEQ, that is important for INT but, rather, the number of times that the INT process must be independently called upon to activate information to support loading of the motor buffer (see Immink & Wright, 2001). These data

imply the presence of unique syllables in speech sequence tax processes involved in loading of the motor buffer as opposed to its emptying (i.e., SEQ).

Interestingly, the reported impact of unique syllable preparation for the demand on the INT process in the present data also sheds some light on why Deger and Ziegler (2002) failed to reveal an influence of alternating syllables beyond mere repetitions for normal speakers on simple RT. Recall that the use of a simple RT paradigm is particularly suited for assessing the impact of the SEQ process in Klapp's model, a process that, according to the present data, is not involved in accounting for the preparation of unique syllables during programming. Had Deger and Ziegler used a choice RT paradigm, they may have observed the anticipated effect of having to prepare more syllables in the alternating conditions because this RT protocol would be more sensitive to the INT process (Klapp, 1995). It is this latter process that, according to our data, would be most impeded by the additional demands incurred from preparing more unique syllables for an upcoming utterance.

While the previously described account for longer mean ST associated with preparation of unique syllables is the most parsimonious, with the theoretical account providing the impetus for this work, there is an alternative explanation that at this point cannot be completely ruled out. Specifically, it is possible that the nature of the symbolic coding (i.e., 4X and 4Y) that was linked with the utterances involving multiple unique syllables was less compatible than the relationship between the codes used for utterances involving merely repetitions of the same syllable (e.g., 4T for "ta-ta-ta-ta"). Resolution of compatibility issues has traditionally been considered to influence response selection, a process captured within the ST interval of the self-select paradigm. This being the case, if the utterances involving unique syllables was indeed instigated by a less compatible stimulus–response (S–R) pairing, a response selection as opposed to response programming process would be responsible for the greater mean ST reported when readying unique syllables as opposed to repeating syllable utterances. Although we cannot discount this possibility, we should point out that individuals were trained on the S–R mappings prior to practice to a level in which the participant consistently linked the appropriate stimulus with the correct speech response without error. Moreover, no participant indicated that this S–R mapping was difficult to use. [2]

Knowledge regarding the unfolding of speech production has relied heavily on behavioral evidence similar to that presented in the current work. However, more recent approaches have attempted to understand the neural underpinnings of speech production using functional imaging techniques. This has led to some early speculation as to the nature of a minimal speech network (Guenther, 2003; Guenther et al., 2006; Riecker et al., 2005). Some of this work has particular relevance for the present work because it potentially offers some insight into the neural implementation of the processes responsible for the present behavioral data. For example, Shuster and Lemieux (2005) used fMRI to reveal that the left parietal cortex was related to the performance of words differing in sequence length. This was true whether words were covertly or overtly articulated, suggesting a contribution to speech preparation. Shuster and Lemieux offered a number of possible functional

---

[2]We thank an anonymous reviewer for this interpretation of this particular finding.

preparatory roles for left parietal activity that might be related to handling the demands for utterances of greater length. Two of these accounts, in particular, are quite intriguing given the current delineation of the SEQ process in Klapp's model of motor programming. First, they suggested that there might be a heightened demand for motor attention for articulations of greater length (Rushworth, Johansen-Berg, Goebel, & Devlin, 2003). This proposal is consistent with SEQ being involved in the brief maintenance of the articulatory score, in the case of speech, in a "motor buffer" as opposed to working memory, the latter of which might be more appropriately associated with increased activation inBA46(dorsolateral prefrontal cortex [DLPFC]; see Magnuson, Robin, & Wright, 2008). Second, Shuster and Lemieux considered Guenther's (2003) proposal that the left inferior parietal region contributes to the matching of feedback from the actual speech production system with an internal representation of the feedback that is expected for the prepared utterance. This involves the use of a "forward model" that predicts sensations during speech production. Forward models have previously been included as part of the SEQ process during the execution of simple key pressing sequences (Magnuson, Wright, & Verwey, 2004; Magnuson et al., 2008). Although these authors considered that such a forward model would be important to supervise the emptying of the motor buffer in the correct order, there is nothing to preclude the alternative role, suggested by Shuster and Lemieux, that this aspect of SEQ might be readying the system to evaluate the output.

Although Shuster and Lemieux (2005) focused primarily on the sequence length effect, which for the purpose of the present discussion targets the SEQ process, some of our own work focused on a manipulation of the complexity of the syllable while fixing sequence length, thus providing us an opportunity to speculate as to the neural regions important for INT (Bohland & Guenther, 2006). In a manner similar to Shuster and Lemieux, Bohland and Guenther evaluated deviations in activation patterns and regions beyond those observed for the basic speech network. In this case, they chose to use a GO–NOGO paradigm to determine the basic speech network, which has some advantages over the covert–overt paradigm used in numerous previous studies. Only a few neural regions appeared to be directly sensitive to the structure of the syllable, the complexity of which was manipulated in the same manner as in the present study (CV vs. CCCV). These included the supplementary motor area, inferior frontal gyrus, and anterior insula bilaterally as well as the right cerebellar cortex (Lobule VI). Interestingly, Riecker et al. (2000) also found activation of the same portion of the cerebellum when /stra/ was repeated compared with repetition of /ta/, the same exemplars used in the present study. Taken together, then, these data provide some preliminary evaluation of the neural implementation of the processes central to Klapp's portrayal of motor programming as it pertains to the domain of speech motor control.

In summary, the INT process in speech motor programming can be seen as the means by which the mental syllabary (see Levelt, 1992; Levelt & Wheeldon, 1994) is accessed. We suggest that this process involves loading the motor buffer with the requisite gestural scores that, when packaged together, will entail the appropriate articulatory score to be forwarded to the speech apparatus via SEQ. The present data implicate both syllable structure as well as unique syllables within a sequence as candidate sources for delays associated with the unfolding of the INT process. Syllable structure will influence the "loading" delay for a

syllable, and the number of unique syllables influences the number of times the loading process must occur.

Acknowledgments

References

Ballard KJ, Granier JP, Robin DA. Understanding the nature of apraxia of speech: Theory, analysis, and treatment. Aphasiology. 2000; 14:969–995.

Bohland JW, Guenther FH. An fMRI investigation of syllable sequence production. NeuroImage. 2006; 32:821–841. [PubMed: 16730195]

Cholin J, Levelt WJM, Schiller NO. Effects of syllable frequency in speech production. Cognition. 2006; 99:205–235. [PubMed: 15939415]

Deger K, Ziegler W. Speech motor programming in apraxia of speech. Journal of Phonetics. 2002; 30:321–335.

Dogil G, Ackermann H, Grodd W, Haider H, Kamp H, Mayer J, et al. The speaking brain: A tutorial introduction to fMRI experiments in the production of speech, prosody, and syntax. Journal of Neurolinguistics. 2002; 15:59–90.

Ferrand L, Segui J. The syllable's role in speech production: Are syllables chunks, schemas, or both? Psychonomic Bulletin & Review. 1998; 5:253–258.

Guenther, FH. Neural control of speech movements. In: Schiller, NO.; Meyer, AS., editors. Phonetics and phonology in language comprehension and production: Differences and similarities. Mouton de Gruyter; New York: 2003. p. 209-239.

Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain and Language. 2006; 96:280–301. [PubMed: 16040108]

Henry FM, Rogers DE. Increased response latency for complicated movements and a "memory drum" theory of neuromotor reaction. ResearchQuarterly. 1960; 31:448–458.

Immink MA, Wright DL. Motor programming during practice conditions high and low in contextual interference. Journal of Experimental Psychology: Human Perception and Performance. 2001; 27:423–437. [PubMed: 11318057]

Kimbrough SK, Wright DL, Shea CH. Reducing the saliency of intentional stimuli results in greater contextual-dependent performance. Memory. 2001; 9:133–143. [PubMed: 11338938]

Klapp ST. Motor response programming during simple and choice reaction time: The role of practice. Journal of Experimental Psychology: Human Perception and Performance. 1995; 21:1015–1027.

Klapp, ST. Reaction time analysis of central motor control. In: Zelaznik, HN., editor. Advances in motor learning and control. Human Kinetics; Champaign, IL: 1996. p. 13-34.

Klapp ST. Reaction time analysis of two types of motor preparation for speech articulation: Action as a sequence of chunks. Journal of Motor Behavior. 2003; 35:135–150. [PubMed: 12711585]

Klapp S, Abbott J, Coffman K, Greim D, Snider R, Young F. Simple and choice reaction-time methods in the study of motor programming. Journal of Motor Behavior. 1979; 11:91–101. [PubMed: 15189802]

Klapp ST, Anderson WG, Berrian RW. Implicit speech in reading reconsidered. Journal of Experimental Psychology. 1973; 100:368–374.

Klapp ST, McRae LE, Long W. Response programming vs. alternative interpretations of the "ditdah" reaction time effect. Bulletin of the Psychonomic Society. 1978; 11:5–6.

Klapp ST, Wyatt EP, Lingo WM. Response programming in simple and choice reactions. Journal of Motor Behavior. 1974; 6:263–271. [PubMed: 23961840]

Levelt WJM. Accessing words in speech production: Stages, processes, and representations. Cognition. 1992; 42:1–22. [PubMed: 1582153]

Levelt WJM, Wheeldon L. Do speakers have access to a mental syllabary? Cognition. 1994; 50:239–269. [PubMed: 8039363]

Magnuson CE, Robin DA, Wright DL. Sequencing multiple elements of the same duration: Detailing the INT process in Klapp's two-process account of motor programming. Journal of Motor Behavior. 2008; 40:532–544. [PubMed: 18980906]

Magnuson CE, Wright DL, Verwey W. Changes in the incidental context impacts search but not loading of the motor buffer. The Quarterly Journal of Experimental Psychology. 2004; 57A:935–951. [PubMed: 15204123]

McNeil, MR.; Doyle, PJ.; Wambaugh, J. Apraxia of speech: A treatable disorder of motor planning and programming. In: Nadeau, SE.; Rothi, L. J. Gonzales; Crosson, B., editors. Aphasia and language: Theory to practice. Guilford; New York: 2000. p. 221-266.

McNeil, MR.; Robin, DA.; Schmidt, RA. Apraxia of speech: Definition, differentiation, and treatment. In: McNeil, MR., editor. Clinical management of sensorimotor speech disorders. Thieme; New York: 1997. p. 311-344.

Nijland L, Maassen B, van der Meulen S. Evidenceof motor programming deficits in children diagnosed with DAS. Journal of Speech, Language, and Hearing Research. 2003; 46:437–450.

Nijland L, Maassen B, van der Meulen S, Gabreëls F, Kraaimaat FW, Schreuder R. Planning of syllables in children with developmental apraxia of speech. Clinical Linguistics & Phonetics. 2003; 17:1–24. [PubMed: 12737052]

Riecker A, Ackermann H, Wildgruber D, Meyer J, Dogil G, Haider H, Grodd W. Articulatory/phonetic sequencing at the level of the anterior perisylvian cortex: A functional magnetic resonance imaging (fMRI) study. Brain and Language. 2000; 75:259–276. [PubMed: 11049668]

Riecker A, Mathiak K, Wildgruber D, Erb M, Hertlich I, Grodd W, Ackermann H. fMRI reveals two distinct cerebral networks subserving speech motor control. Neurology. 2005; 64:700–706. [PubMed: 15728295]

Rushworth MFS, Johansen-Berg H, Goebel SM, Devlin JT. The left parietal and premotor cortices: Motor attention and selection. NeuroImage. 2003; 20:89–100.

Sevald CA, Dell GS, Cole JS. Syllable structure in speech production: Are syllables chunks or schemas? Journal of Memory and Language. 1995; 34:807–820.

Shuster LI, Lemieux SK. An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. Brain and Language. 2005; 93:20–31. [PubMed: 15766765]

Sternberg, S.; Monsell, S.; Knoll, RR.; Wright, CE. The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In: Stelmach, GE., editor. Information processing in motor control and learning. Academic Press; New York: 1978. p. 117-152.

van der Merwe, A. A theoretical framework for the characterization of pathological speech sensorimotor control. In: McNeil, MR., editor. Clinical management of sensorimotor speech disorders. Thieme; New York: 1997. p. 1-25.

Verwey WB. Evidence for amultistagemodel of practice in a sequential movement task. Journal of Experimental Psychology: Human Perception and Performance. 1999; 25:1693–1708.

Wright DL, Black CB, Immink MA, Brueckner S, Magnuson C. Long-term motor programming improvements occur via concatenating movement sequences during random but not blocked practice. Journal of Motor Behavior. 2004; 36:39–50. [PubMed: 14766487]

Ziegler, W.; Maassen, B. The role of the syllable in disorders of spoken language production. In: Maassen, B.; Kent, R.; Peters, H.; Lieshout, P.; Hulstijn, V., editors. Speech motor control in normal and disordered speech. Oxford University Press; Oxford, United Kingdom: 2004. p. 415-445.
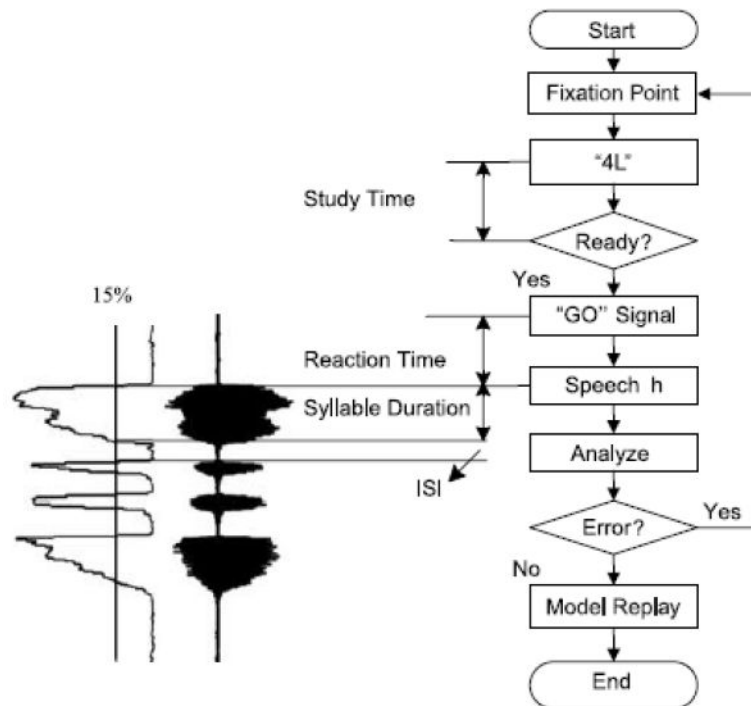
Figure 1.

In the case of the self-select paradigm, a trial begins with fixation stimulus presented in the center of the computer display. Shortly after, the individual is provided information pertaining to which speech response is required. In this example, the participant should prepare the response "baaabababaaa," with the first and last /ba/ being 450 ms in duration and the second and third elements being 150 ms. Study time is defined as the interval between the presentation of a precue (i.e., "4L") and the participant pressing the "END" key. This period is assumed to capture processes specific to INT, which should then be reflected in the latency of this interval. Following the depression of the "END" key, a variable foreperiod occurs prior to the presentation of the "GO" signal. Reaction time (RT) is defined as the interval between the presentation of the "GO" signal and the initiation of the speech task. In keeping with the approach adopted by Deger and Ziegler (2002), we used the SPL contour as a means of identifying the beginning and end of the speech segments (see text for details). The RT interval is assumed to capture the temporal demands of SEQ. Feedback in the form of the model was presented at the end of a trial. ISI = intersyllable interval.
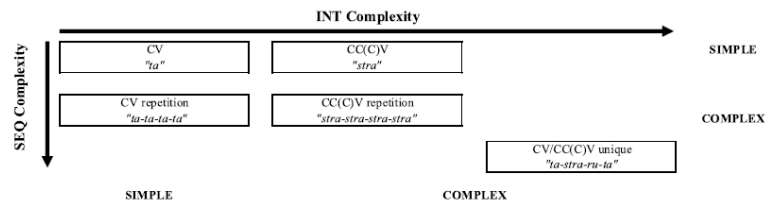
Figure 2.
Five experimental conditions were established that differentiated speech responses on the basis of INT complexity (horizontal dimension) and SEQ complexity (vertical dimension). At the INT level, simple syllables were composed of a single consonant and a vowel (CV), whereas the complex syllable began with a consonant cluster (CCCV) by a vowel. At the SEQ level, complexity was increased by (a) adding repetitions of the same syllable ($CV_{repetition}$ or $CCCV_{repetition}$) and (b) including transitions between unique syllables (CV/$CCCV_{unique}$) during an utterance.
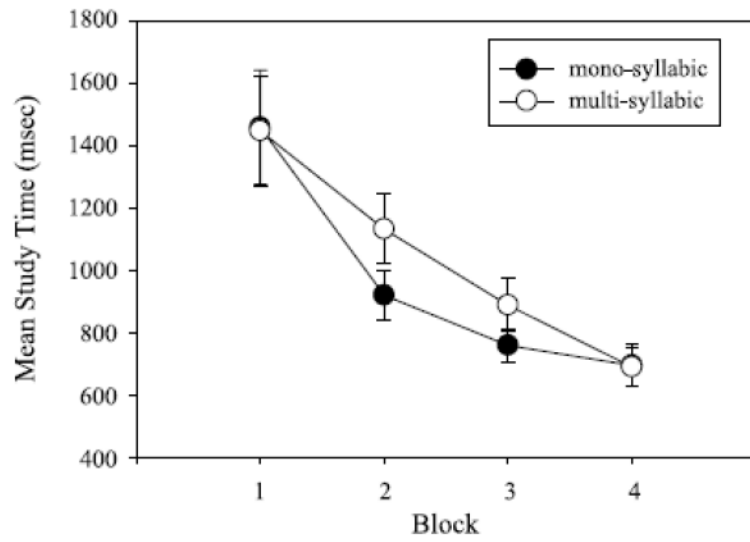
Figure 3.
Mean study time (in ms) when articulating mono- and multisyllabic responses as a function of trial block. Error bars indicate standard errors.
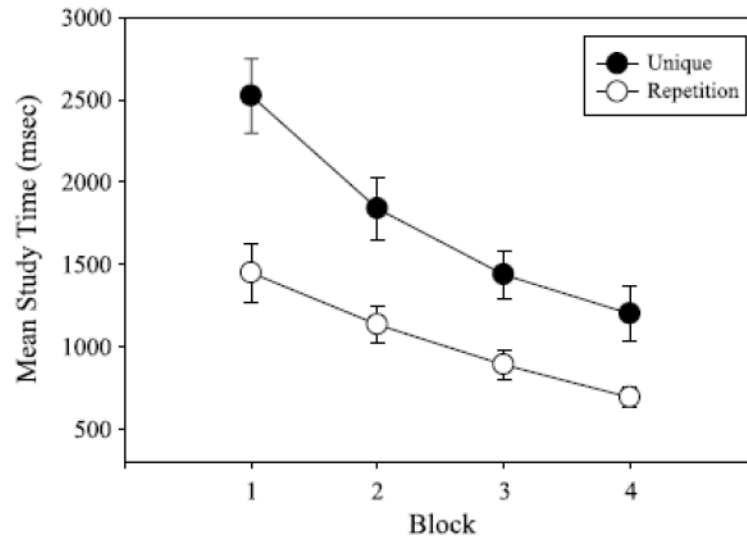
Figure 4.
Mean study time (in ms) when articulating multisyllabic responses that involve either unique syllables or repeated syllables as a function of block. Error bars indicate standard errors.

Table 1

Mean and variability (in ms) for utterance duration for each task as a function of trial block.

							Task							
	1T		**1S**		**4T**		**4S**		**4X**		**4Y**	
Block	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
1	292	36	311	38	1,785	94	1925	173	1,837	125	1,856	103
2	316	26	317	33	1,793	114	1855	81	1,833	94	1,843	115
3	296	50	300	36	1,766	95	1870	55	1,800	54	1,860	104
4	293	25	318	53	1,806	75	1878	69	1,838	150	1,833	111

Table 2

Mean study time (ST; in ms) and reaction time (RT; in ms) for each task as a function of trial block.

Block	1T		1S		4T		4S		4X		4Y	
	ST	RT	ST	RT	ST	RT	ST	RT	ST	RT	ST	RT
1	1,224	649	1,706	744	1,296	747	1,607	829	2,330	823	2,690	868
2	844	635	1,000	714	1,061	749	1,203	777	1,707	790	1,968	734
3	764	658	756	643	931	693	836	715	1,317	683	1,551	675
4	620	592	771	661	710	655	671	626	1,147	652	1,255	761

(Task)