Journal of Experimental Psychology:
Learning, Memory, and Cognition
2014, Vol. 40, No. 2, 376–384

0278-7393/14/$12.00 DOI: 10.1037/a0035221

Control of Task Sequences: What Is the Role of Language?

Ulrich Mayr, Killian Kleffner-Canucci, Atsushi Kikumoto, and Melissa A. Redford
University of Oregon

It is almost a truism that language aids serial-order control through self-cuing of upcoming sequential elements. We measured speech onset latencies as subjects performed hierarchically organized task sequences while "thinking aloud" each task label. Surprisingly, speech onset latencies and response times (RTs) were highly synchronized, a pattern that is not consistent with the hypothesis that speaking aids proactive retrieval of upcoming sequential elements during serial-order control. We also found that when instructed to do so, subjects were able to speak task labels prior to presentation of response-relevant stimuli and that this substantially reduced RT signatures of retrieval—however, at the cost of more sequencing errors. Thus, while proactive retrieval is possible in principle, in natural situations it seems to be prevented through a strong "gestalt-like" tendency to synchronize speech and action. We suggest that this tendency may support context updating rather than proactive control.

Keywords: executive control, sequencing, language

Language is often thought to play a pivotal role in organizing complex thought and action (Vygotsky, 1986). However, there is surprisingly little known about how exactly language might be used for this purpose. One way in which language may be helpful is that it allows one to cue upcoming demands in a proactive manner. Verbal labels may become associated with nonverbal, action-relevant codes that represent aspects such as where to direct attention or which motor actions to select (Hommel, Pratt, Colzato, & Godijn, 2001). Articulating verbal labels activates the associated codes and thus prepares their fluent application. In fact, work with the task-switching paradigm has demonstrated that verbalizing task labels reduces task-switch costs (Goschke, 2000; Kray, Eber, & Karbach, 2008). Moreover, in situations in which people need to select tasks according to an internal sequential rule, articulatory suppression produces substantial interference with performance (Baddeley, Chincotta, & Adlam, 2001), and it has been specifically argued that articulatory suppression interferes with preparatory self-cuing (Miyake, Emerson, Padilla, & Ahn, 2004).

The notion of preparatory control during complex sequential action is also consistent with standard models of serial-order control. Such models assume that complex sequences are organized hierarchically in terms of chunks of two–four sequential elements and that these chunks themselves can become recursively embedded within higher order chunks (Klapp, 1976; Rosenbaum, Kenny, & Derr, 1983). Executing such a sequence requires retrieving the currently relevant chunk into working memory, which explains why interresponse times between successive chunks are usually much larger than within-chunk interresponse times. Mayr (2009), using a paradigm that required executing complex sequences of tasks (Schneider & Logan, 2006), recently reported evidence for another type of sequence-structure effect, namely, *positional interference*. For example, when three different tasks A, B, and C are combined in the two-chunk sequence ABC-ACB, retrieval of the task C in the middle of the second chunk is impaired because of interference from that task's association with the final position of the first chunk (see also Koch, Philipp, & Gade, 2006; Schneider, 2007).[1]

Undoubtedly, most proponents of the standard model of serial-order control would assume that verbal self-cuing should help with exactly these types of retrieval-related costs. Thus, if speaking during sequencing is in fact used in a proactive manner (i.e., before the relevant action needs to be executed), then it should lead to a marked reduction or even elimination of such sequence-structure effects.

Traditionally, research on serial-order control has used paradigms in which people executed complete sequences (e.g., of key presses) from memory, which allows no control over interresponse timing and thus does not allow explicit tests of the effects of preparation and/or verbal self-cuing. This has changed with the so-called task-span procedure (Koch, 2003; Mayr, 2009; Schneider & Logan, 2006), in which subjects work through memorized

This article was published Online First November 25, 2013.

Ulrich Mayr, Killian Kleffner-Canucci, Atsushi Kikumoto, and Melissa A. Redford, Department of Psychology, University of Oregon.

This work was funded by the Beverly and Robert Lewis endowment and by National Institute on Aging Grant RO1 AG037564-01A1.

Correspondence concerning this article should be addressed to Ulrich Mayr, Department of Psychology, University of Oregon, Eugene, OR 97403. E-mail: mayr@uoregon.edu

[1] To avoid confusion, it is useful to clarify two aspects. First, what we refer to here as positional interference costs had been originally described as *lag-2 task repetition cost* (Koch et al., 2006; Schneider, 2007), which typically is interpreted as an aftereffect of task-set inhibition (Mayr & Keele, 2000). However, as reported in Mayr (2009), for the sequence grammars used both here and in the previous work, lag-2 task repetitions happen to coincide with points of high positional interference. It is mainly interference, rather than the aftereffect of inhibition, that is responsible for RT and error costs at these points in the sequence. Second, at first sight one could argue that positional interference should also occur at other points of the sequence. For example in a sequence such as ABC-ACB, not only at the second but also at the third within-chunk position, tasks need to be selected that occur at a different chunk position within the six-element sequence. However, according to the full model (Mayr, 2009), a just-executed position becomes deactivated via self-inhibition (i.e., in this case the Position 2 task) and therefore is removed as a potential source of interference on the following trial (i.e., when executing the Position 3 task).

sequences of tasks. Given that here a response can be executed only after the stimulus has been presented, the time available for preparation—-the interval between the previous response and the next stimulus, or response-stimulus interval (RSI)—-is under full experimenter control. For example, Schneider and Logan (2006) reported a study in which subjects had to work through multichunk sequences. Chunk-initiation times were slightly reduced for long RSIs. However, this preparation-related reduction was much smaller than would be expected if proactive, verbal self-cuing actually allowed loading upcoming sequential elements into working memory. Similarly, Koch (2003) had asked subjects to alternate between two tasks in runs of two trials (AA-BB) and found no evidence for reduced task-switch costs (which here are confounded with chunk-initiation times) as a result of increased RSIs. Finally, in the above-mentioned study (Mayr, 2009) large response time (RT) and error effects of both positional interference and chunk initiation were observed despite a 1,000-ms RSI, a duration that should have easily sufficed to retrieve the next sequential element into working memory.

Given that there can be little doubt that language is critical during complex sequencing demands (e.g., Baddeley et al., 2001) and given the generally assumed role of language in allowing to plan and precue upcoming behavior, the results summarized here clearly present a puzzle. What is missing is research that directly assesses and experimentally manipulates the role of speech during serial-order control.

The Current Research

To examine the interplay between language and sequential control, we used the same paradigm as in Mayr (2009). As basic task elements, subjects worked with three different response rules (i.e., tasks) that defined the relationship between a given stimulus (i.e., a circle in one of four corners of a square; see Figure 1) and the required response (i.e., one of four keys arranged in terms of a 2 × 2 matrix). For example, in the case of the "horizontal" rule, response locations were determined by translating stimulus locations horizontally (e.g., for a circle in the upper-left corner of the square, the upper-right response key had to be pressed). The two other rules required "vertical" and "diagonal" translations between stimulus and response locations. Thus, on a given trial, subjects needed both the stimulus and the response rule to execute a correct response. For each block of trials the relevant rules were specified through a different sequence that always consisted of two three-element chunks (see Figure 1). These sequences had to be memorized at the beginning of each block and then "cycled through" continuously until the block ended. We expected particularly high demands of retrieving upcoming sequential elements at two points: (a) the beginning of each chunk and (b) when positional interference is high (i.e., the first position in a chunk at which an element occurs that also occurs at a different chunk position within the same sequence).

We tested three groups of subjects. The first group was asked to verbalize task labels as they worked through memorized sequences of tasks without further constraints (i.e., the free-speaking group). The second group was urged to speak each trial's task label prior to the presentation of the stimulus. The third group served as a nonspeaking control condition to ensure that explicit verbalization instructions did not change overall task performance. In order to examine the relative timing of speaking and motor responses, we digitally recorded subjects' verbalizations.

With this design, we could test two critical predictions derived from the standard idea that language supports proactive retrieval of

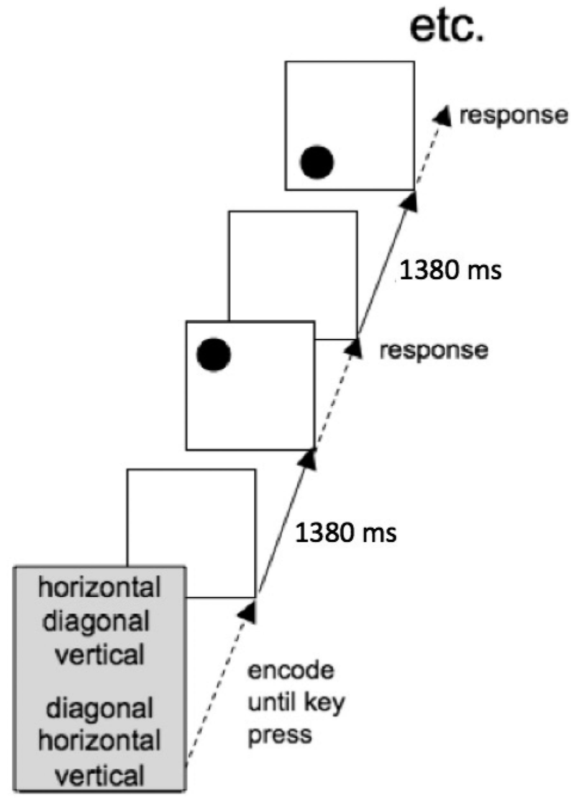

Figure 1. Sequence of events in the task-span procedure. At the beginning of each block, subjects memorized the sequence of response rules at leisure and then initiated the series of stimulus presentations to which the sequence had to be applied repeatedly (i.e., seven times per block). In the Early Speaking condition, subjects were asked to speak the task label during the 1,380-ms response-stimulus interval.

upcoming sequential elements (i.e., in this case the stimulus–response, or S-R, rules specified by the memorized sequence). First, we examined to what degree people naturally use speech in a proactive manner (in the Free Speaking group), or are at least are able to do so when explicitly prompted (in the Early Speaking group). In case language is used proactively, onset latencies of spoken labels should fall predominantly within the relatively long response-stimulus interval of 1,380 ms. Second, if proactive use of language actually does enable self-cuing, one should find that when subjects speak early, sequence-structure effects (i.e., chunk-initiation or positional interference costs) are reduced or even eliminated for manual RTs but should still be fully reflected in the pattern of speech-onset latencies.[2]

[2] We do not rule out the possibility that proactive verbalization affects the actual implementation of a given task set. Such effects would be reflected in a general speed-up of RTs, even in the absence of high sequencing demands. However, because our interest here is in serial-order control, our design is geared toward identifying sequence-structure effects (i.e., chunk initiation and positional interference) rather than such more general effects.

Method

Subjects

Forty-eight students of the University of Oregon (29 women, ages = 18–24 years) participated in this experiment in exchange for course credits.

Tasks, Stimuli, and Procedure

Subjects performed sequences of the spatial rules task (Mayr, 2009). Stimulus presentation occurred on a 17-inch Macintosh monitor. The stimulus display contained a frame in the form of a square, with side lengths of 8 cm (9.1°). On each trial, a circle with a diameter of 1 cm appeared in one of the four corners of the frame (see Figure 1). Responses were entered using four keys on the standard Macintosh keyboard with the same spatial arrangement as the four circle locations (*1, 2, 4,* and *5*) on the numerical keypad. Subjects were instructed to rest the index finger of their preferred hand in the middle between the four keys and to move the finger to the correct key. After pressing the key, subjects were instructed to bring the index finger back to the center position. Correct keys were cospecified by the circle location and the currently relevant rule. Specifically, the rule indicated the spatial translation necessary to get from the stimulus position to the response position. For the "horizontal rule," response locations were determined by "moving" the stimulus location horizontally either from left to right or from right to left (e.g., upper-left stimulus leads to upper-right response); for the "vertical" rule, response locations were determined by moving the stimulus location vertically either from top to bottom or bottom to top (e.g., upper-left stimulus leads to lower left response); and for the "diagonal" rule the stimulus locations had to be moved diagonally to derive the response location (e.g., upper-left stimulus leads to lower right response).

The rule to be applied to the stimulus for a given trial was specified through that trial's position in the task sequence. Sequences could have three different abstract "grammars," which not only ensured that each task occurred equally often on each position, but also that points of high positional interference were counterbalanced across sequence positions (in bold): 1 = **A**BC-**B**AC, 2 = A**B**C-A**C**B, or 3 = AB**A**-CB**C**. Note, that the three different grammars are constructed from the same basic sequence, just shifted in position. In total, there were 12 different specific sequences that could be implemented by combining three abstract sequences with three different rules in all possible manners, one for each of 12 sequence blocks (i.e., counterbalancing rules across sequence positions). Blocks were 42 trials long so that each six-element sequence was repeated seven times within a block. Prior to each block, the relevant six-rule sequence of task labels was presented on the screen above the empty stimulus frame. To facilitate organization of the sequence into two chunks of three trials each, the two chunks were presented with a spatial separation (for further details, see Mayr, 2009). With a press of the space bar, subjects could initiate the actual block. At this point, the sequence cue disappeared, and after 1,000 ms the first stimulus was presented until the correct response was entered; all following stimuli were presented in the same manner, with a constant response-stimulus interval (RSI) of 1,380 ms. Subjects moved through the memorized sequence element by element, starting over after the sixth element. In case of an error, the stimulus remained on the screen and the sequence of six task labels reappeared with the currently relevant task label highlighted. This allowed subjects to realign themselves with the sequence. The sequence information disappeared and the block resumed only after subjects entered the correct response.

Subjects were randomly selected into one of the three groups: the No Speaking, the Free Speaking, or the Early Speaking group. No Speaking subjects received standard instructions without mentioning the option to explicitly label the to-be-performed tasks. Subjects in the Free Speaking group also received standard instructions but in addition were asked to overtly name each task label (e.g., "vertical") while going through the task sequence. For most subjects this seemed to be a relatively natural activity, and few required additional reminders during the experiment. Subjects in the Early Speaking group were instructed to speak the task label during each trial's prestimulus interval. Subjects in this group were provided an opportunity to familiarize themselves with speaking in a proactive manner in the practice blocks, and when they showed difficulties adhering to the instruction they were reminded between blocks to speak early.

Data Treatment

Aside from standard response times and accuracy, we also recorded spoken output digitally from the subjects in the two speaking groups. We used the phonetic analysis software Praat (Boermsa & Weenink, 2009) to represent the intensity contour of the spoken output as a function of time, which allowed us to manually trace the onset and offset of each utterance. We also coded speech errors. Thus, we were able to characterize each trial in terms of RT, response accuracy, onset and offset latency of spoken task labels, and accuracy of task labels.

Results and Discussion

We excluded RTs longer than 4,000 ms after stimulus onset (i.e., .5% of all RTs), error trials, and trials following errors, as well as the first six trials of each block (i.e., the first cycle with a new sequence). Data from one subject in the No Speaking group were corrupted and had to be dismissed. Figure 2 shows RTs for the three groups as a function of the six different sequence positions (i.e., averaging across the three different sequence grammars) and positional interference. Note that, for the positional interference factor, for each sequence position, one abstract sequence grammar that produces high positional interference can be contrasted with the average of two grammars that produce low positional interference (i.e., for the first position, **A**BC-**B**AC vs. A**B**C-A**C**B and AB**A**-CB**C**). For the Free Speaking and the Early Speaking groups, we also show the average onset and offset of utterances (note that offset latencies carried no relevant information over and above onset latencies and therefore are not reported here). In analyzing the results, we first compare the No Speaking group with the Free Speaking group and then the two speaking groups with each other.

No Speaking Versus Free Speaking

As is obvious from Figure 2, RTs for both of these groups showed strong chunk position and positional interference effects,

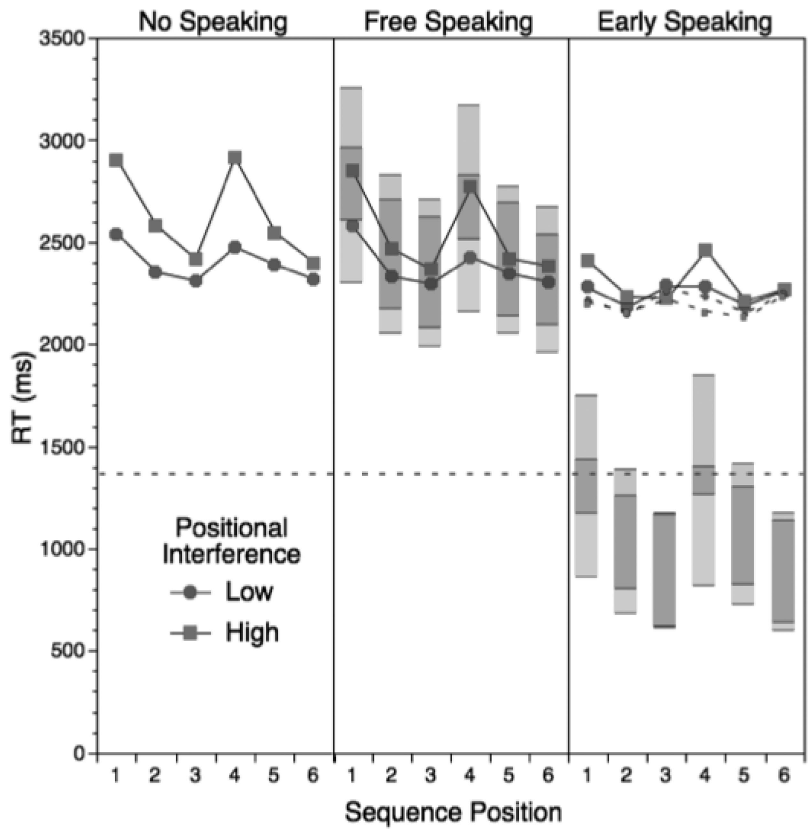

Figure 2. Response times (RTs) since the preceding response as a function of the six sequence positions (Positions 1–3 = Chunk 1, Positions 4–6 = Chunk 2) and low versus high positional interference for all conditions. The horizontal dashed line indicates the stimulus onset. The horizontal end points of the shaded rectangles represent average speech onsets and offsets for the speaking conditions. The small circles and squares with dashed lines in the Early Speaking condition represent RTs in trials for which the speech onset occurred prior to stimulus onset.

which according to Mayr (2009) represent costs of retrieving the current task rules from long-term memory (LTM). We committed RTs to an analysis of variance with the between-subjects factor Speaking/No Speaking, the within-subject factor Chunk 1/Chunk 2, low/high positional interference, and two orthogonal contrasts, the first comparing within-chunk Position 1 with Positions 2 and 3, and the second comparing within-chunk Position 2 with 3. For the sequence structure factors, we found that high positional interference trials produced larger RTs than did low positional interference trials, $F(1, 29) = 107.43$, $MSE = 64{,}614.23$, $p < .001$; that Position 1 RTs were reliably longer than RTs for the remaining two chunk positions, $F(1, 29) = 126.74$, $MSE = 52{,}670.01$, $p < .001$; that there also was a much smaller difference between chunk Positions 2 and 3, $F(1, 29) = 13.70$, $MSE = 27{,}133.21$, $p < .01$, and $MSE = 36{,}998.16$, $p < .001$, respectively; and that the positional interference cost was particularly large for within-chunk Position 1, $F(1, 29) = 36.26$, $MSE = 33{,}817.22$, $p < .01$. None of the effects involving the Speaking/No Speaking factor were reliable (all Fs < 1.18, $p > .28$).

Errors are presented in the left two panels of Figure 3. For the Speaking group, one can distinguish errors where subjects spoke the incorrect rule label (speech errors) and errors where the label was correct but the response incorrect (response errors). For total errors, we again found considerably higher error rates for high than for low positional interference trials, $F(1, 29) = 35.52$, $MSE = 62.77$, $p < .001$. While there was no main effect for the Position 1/Positions 2–3 contrast, the positional interference effect was larger for Position 1 than for Positions 2 and 3, $F(1, 29) = 6.10$, $MSE = 25.82$, $p < .05$, and this effect was further increased for the first chunk, $F(1, 29) = 5.61$, $MSE = 15.39$, $p < .05$. In addition, there were more errors for Position 2 than for Position 3, $F(1, 29) = 12.16$, $MSE = 66.50$, $p < .01$. Again, however, these sequence-structure effects were not modulated by the Speaking/No Speaking factor (Fs > 1.5, $p > .2$). As is apparent from the distinction between the speaking and response errors, the majority of the total errors were due to speaking errors, and in particular the sequence structure effects were reflected in speaking errors, suggesting that these errors were due to incorrect retrieval of the relevant task rule. Overall, we replicated in both RTs and errors the basic pattern of sequence-structure effects from Mayr (2009) and found no evidence that the speaking demands altered serial-order processing in a theoretically significant manner. This is consistent with our intuition that even in the No Speaking condition, subjects verbalize (mostly) internally, and therefore the speaking instruc-

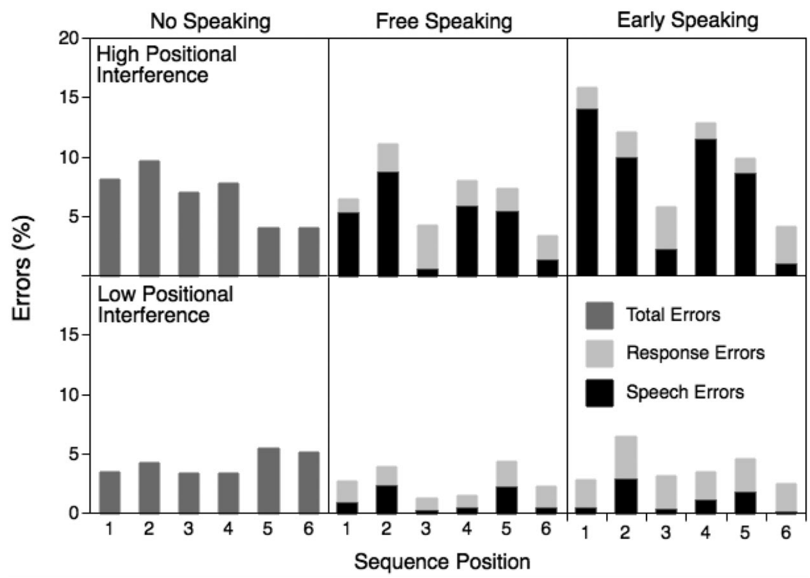

Figure 3. Errors as a function of sequence position and high versus low positional interference for all conditions. For the two speaking conditions, we distinguished between errors that result from speaking the incorrect task label and response errors that occur despite speaking the correct label.

tion presents no qualitative change beyond what most subjects would be doing anyway.

Next, we turn to the timing of the spoken task labels in the Free Speaking condition. If speech in the context of serial-order control is used in a proactive manner, then one should expect onset latencies to precede RTs substantially—-ideally they should occur prior to stimulus presentation. However, as Figure 2 shows, according to the average onsets and offsets of the spoken task label, subjects began each utterance just about 200 ms before, and finished about 300 ms after, the average manual response time in each condition. This also implies that the sequence-structure effects were fully represented in the speech onset/offset times. Accordingly, we found highly significant positional interference costs, $F(1, 15) = 23.79$, $MSE = 1{,}136{,}240.29$, $p < .001$; Position 1 versus Positions 2/3 costs, $F(1, 15) = 74.35$, $MSE = 53{,}851.34$, $p < .001$; and a further increase of the positional interference costs for Position 1, $F(1, 15) = 10.68$, $MSE = 40{,}301.41$, $p < .01$.

One can also look at within-subject cross-correlations between speaking and responding. Figure 4 presents the relevant scatterplots of speech-onset times as a function of RTs for each subject. In order not to confound actual synchronization with condition effects that affect both RTs and speech-onset times (see Figure 2), we residualized both variables with regard to all condition effects and then added these residuals to the individual subject's overall means. A proactive use of speech would imply a substantial amount of data points that occur to the left of the diagonal line, which indicates perfect synchronization between speaking and manual responses. As is evident, overall subjects exhibited a strong synchronization tendency. In fact, most subjects very rarely spoke in the interval prior to stimulus onset (i.e., for 13 of the 16 subjects, prestimulus speech onsets occurred on less than 3% of trials) and even for the remaining three subjects the proportion of early speaking trials never exceeded 34% (average = 5.4%, SD = 1.1%). Thus, for a majority of subjects, synchronization between speaking and manually responding was the dominant mode of integrating the two streams of processing.

Free Speaking Versus Early Speaking

Next we turn to the contrast between the Free Speaking and the Early Speaking group. The first question one can address here is whether subjects were able to desynchronize speaking and manually responding when instructed to do so. The answer is clearly "yes": The average speaking onsets occurred during the prestimulus interval (see right panel of Figure 2). In addition, Figure 5 shows the individual-specific cross-correlations of speech-onset times and RTs (again both variables residualized with regard to condition effects). As is obvious, subjects responded during the prestimulus phase on a majority of trials (93.4%, SD = 3.84%). Nevertheless, at least for some subjects there also seemed to be an occasional tendency to revert to a synchronization pattern, as indicated by the "diagonal tail" in the data pattern.

The next question one can ask is to what degree early speaking in fact reduces the sequence-structure effects, as would be expected if speaking actually allows subjects to retrieve the upcoming task demands prior to stimulus onset. As Figure 2 shows, not only were RTs generally shorter, but in particular the sequence-structure effects (i.e., positional interference and chunk-initiation costs) were reduced compared to the Free Speaking group. For statistical analyses, we compared manual RTs from the Early Speaking group with the Free Speaking group. Overall, the Early Speaking group responded faster than did the Free Speaking group, $F(1, 30) = 5.71$, $MSE = 513{,}587.20$, $p < .03$. The group effect was particularly large for high positional interference trials, $F(1, 30) = 8.54$, $MSE = 31{,}800.07$, $p < .01$. Also, it was larger for chunk Positions 1 versus 2 or 3, $F(1, 30) = 18.71$, $MSE = 31{,}854.08$, $p < .01$, and for chunk Position 2 versus 3, $F(1, 30) = 8.74$, $MSE = 15{,}175.10$, $p < .01$. Speaking-induced preparation

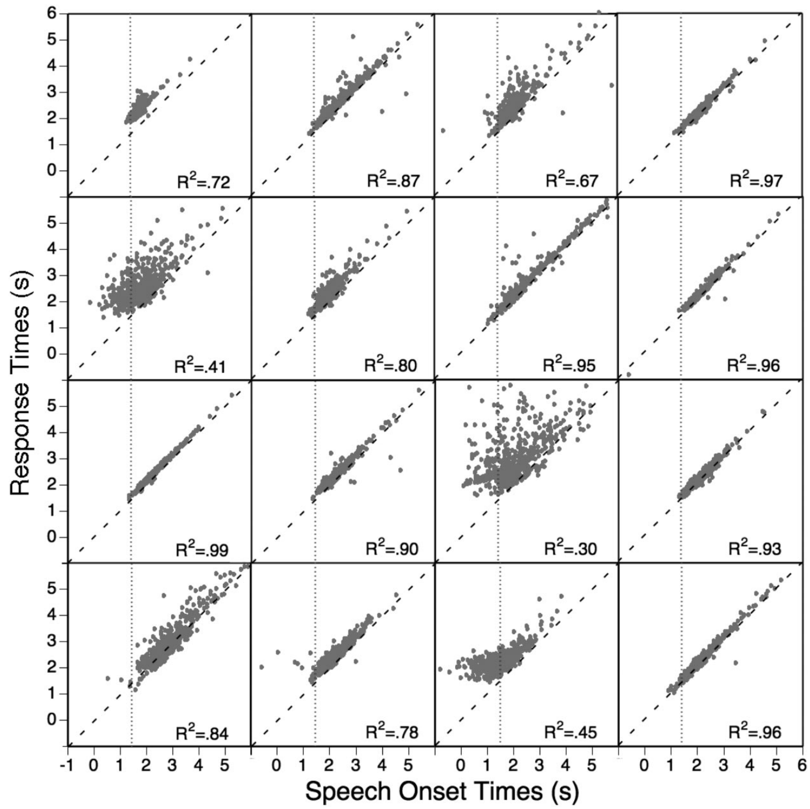

Figure 4. Individual subject's response times (RTs) as a function of speech-onset RTs (both residualized with regard to all condition effects) and linear fit statistics (R^2) in the Free Speaking condition.

did not completely eliminate the sequence-structure effects. While the positional interference effect in the Early Speaking group was no longer reliable, $F(1, 15) = 2.50$, $MSE = 43{,}297.66$, $p > .1$, the Position 1 slowing was, $F(1, 15) = 18.86$, $MSE = 26{,}755.32$, $p < .01$, as was the interaction between the positional interference factor and the contrast between Position 1 and Positions 2/3, $F(1, 15) = 8.74$, $MSE = 81{,}757.49$, $p < .05$.

Despite the instruction to speak early, there was a portion of trials (i.e., 6.6%) in which subjects failed to speak prior to stimulus onset. Therefore, we also show in Figure 2 the RTs after excluding these trials (small circles and dashed lines). As is apparent, for the remaining "true" early speaking trials, sequence-structure effects are essentially eliminated.

The flip side of the prediction that early speaking eliminates sequence-structure effects in manual RTs is that the effects of sequence-structure effects should be fully present in the speech-onset latencies. As Figure 2 shows, the sequence-structure effects in speech-onset times were in fact very similar to those in the Free Speaking group. There were no significant interactions between the group factor and the critical sequence-structure effects, all $Fs(1, 30) < 1.0$. The only exception was a Group × Position 1 versus Positions 2/3 interaction, $F(1, 30) = 7.04$, $MSE = 6{,}948.88$, $p < .05$, due to the fact that for the Free Speaking group the Position 1 slowing was larger for the first chunk and for the Early Speaking group it was larger for the second chunk. Irrespective of this small, unexpected effect, the overall pattern is clearly consistent with the standard model: The sequence structure effects in the speech-onset latencies suggest that speaking implies retrieval of the upcoming task rule into working memory, and the reduction of these effects for the manual RTs suggests that preloading of task rules does in fact produce preparatory benefits. In light of these results it seems even more surprising that when subjects use speech in an unconstrained manner, they do not utilize these preparatory benefits.

An important hint why this may be the case comes from the analysis of errors. As Figure 3 (right two panels) shows, subjects in the Early Speaking condition committed more total errors than did subjects in the Free Speaking condition, $F(1, 30) = 3.97$, $MSE = 113.25$, $p = .056$. Furthermore, the error costs of early speaking were particularly prominent for trials with high positional interference and chunk beginnings. The interaction between the Position 1 versus Positions 2/3 contrast and group was reliable, $F(1, 30) = 4.86$, $MSE = 40.12$, $p < .05$, and this effect was further modulated by the positional interference factor, $F(1, 30) = 34.24$, $MSE = 42.95$, $p < .01$.

Before further interpreting the error costs during free speaking, it is important to rule out the possibility that these result from a speed–error trade-off. If this were the case, then one should find that within individuals, trials for which speaking occurred early were particularly error-prone. To examine this possibility, we used a generalized mixed regression model with a logistic linking function to predict error probability as a function of speech onset times in the Early Speaking group. We included speech onset times, and also first versus second chunk, chunk beginnings, and positional interference plus all interactions between these vari-

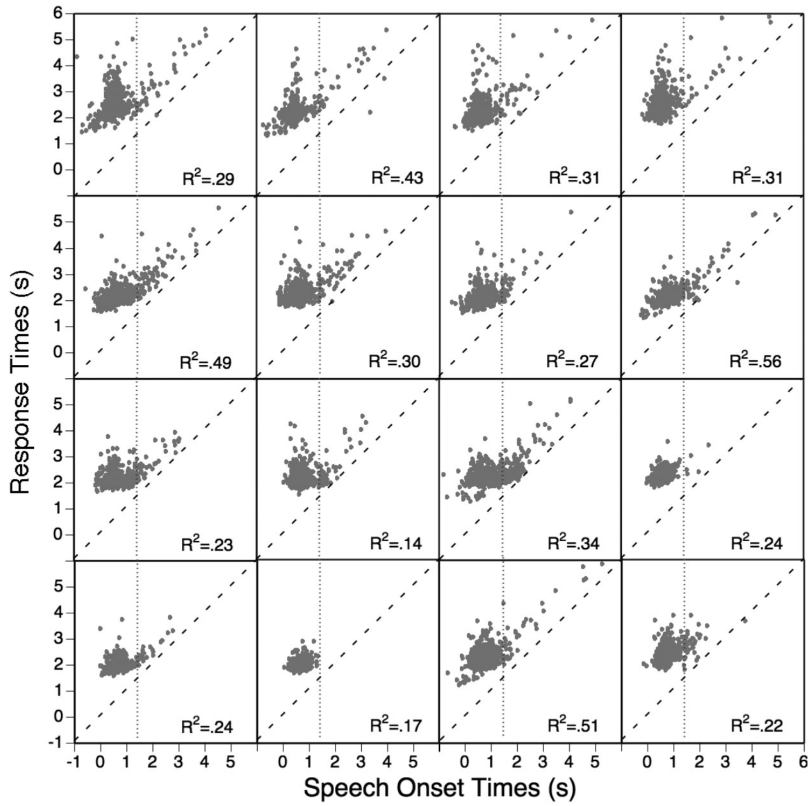

Figure 5. Individual subject's response times (RTs) as a function of speech onset RTs (both residualized with regard to all condition effects) and linear fit statistics (R^2) in the Early Speaking condition.

ables, both in terms of fixed effect predictors and in terms of main effects only, also as subject-specific random effects. Inconsistent with the speed–accuracy trade-off interpretation, we found that error probability significantly increased rather than decreased as a function of within-individual speech onset times ($b = .0009$, $SE = .0002$, $p < .001$), whereas none of the interactions between speech onset and any of the design variables was reliable. On average, subjects spoke 241 ms later on error trials than on nonerror trials. In addition, we also checked whether individuals who tended to speak early were also those who made more errors, both in general and specifically at chunk beginnings or when positional interference was high. However, none of these correlations came even close to significance (all ps $> .5$).

Given that a simple speed–accuracy trade-off explanation can be ruled out, the pattern of error effects suggests that desynchronized, early speaking may help in preparing the upcoming sequential element; however, it may also increase the overall sequencing demands. Specifically, it is plausible that speaking task labels concurrently with manual action aids the process of updating the current position in the sequence. Consistent with this interpretation, error effects were particularly strong for chunk beginnings or when positional interference was high, and the bulk of these effects was reflected in sequencing/speech errors rather than in response errors (see Figure 3).

If it is in fact the case that speaking early makes it harder to update the current sequence position, then it should become more difficult to retrieve a task when, for the preceding position in the sequence, updating would have been particularly relevant but, due to early speaking, less efficient. It is known that positional interference induces high updating demands (Mayr, 2009). Thus, one can expect postponed speaking onsets on trials that follow high positional interference trials, but only in the Early Speaking condition. We analyzed speech-onset latencies in trials that either followed positional interference trials or not. We limited this analysis to trials that were not themselves high in terms of positional interference to account for the fact that while low positional interference trials could be followed by high positional interference trials, there were never two high positional interference trials in a row. We found that in the Free Speaking group, speech onsets were actually 34 ms faster after high positional interference trials, $t(15) = 1.89$, $p < .08$, but 68 ms slower in the Early Speaking condition, $t(15) = 1.94$, $p = .01$. This interaction was highly reliable, $F(1, 30) = 12.12$, $MSE = 3{,}532.54$, $p < .01$. Positional interference on the previous trial affected only speech onsets and not regular RTs, $F(1, 30) < .2$. Combined, these results are consistent with the idea that early speaking interferes with sequence updating and therefore slows down retrieval of sequential elements at points at which updating should be particularly important.

Conclusion

The existing evidence on the role of speech during serial-order control had posed a puzzle: On the one hand, speech clearly plays an important role when complex sequences need to be executed (e.g., Baddeley et al., 2001). On the other hand, there has been little definitive evidence that people actually use speech in a proactive manner—as the standard model of serial-order control might suggest. The current results solve this puzzle. Through slight nudging (i.e., in the Early Speaking group), we were able to entice people to use verbal control in a proactive manner. As a consequence, demands of retrieving sequential elements, while clearly affecting speech-onset times, influenced actual response times to a much lesser degree than when people were not asked to verbalize the sequence. The fact that proactive verbalization is sufficient to reduce or even eliminate retrieval effects from the RT pattern is consistent with the standard model of how verbalization should aid proactive control (Rosenbaum et al., 1983). However, to our knowledge this has never been empirically demonstrated.

While the results from the Early Speaking group indicate the people can use speech proactively in principle, the results from the Free Speaking group show that people's natural way of using language along with complex serial action is to synchronize speaking and acting. In light of the obvious effectiveness of proactive verbalization, it is particularly surprising that in a more natural, free speaking situation people do not seem to use speech proactively. Rather, they seem to use a "just in time" mode of retrieving upcoming task demands, which allows them to successfully perform the task and to synchronize speech and action, albeit at the cost of increased RTs when retrieval demands are high.

The finding that synchronization dominates the speaking–acting relationship explains why in past work on serial task control, the opportunity to prepare seemed to have so little effect (Koch, 2003; Mayr, 2009; Schneider & Logan, 2006). It would be interesting to explore whether a synchronization tendency is also present in the absence of explicit sequencing demands. Specifically, the notorious residual task-switch cost (Kiesel et al., 2010; Monsell, 2003) that is found when tasks are cued in a trial-by-trial manner may at least in part be due to subjects' tendency to delay proactive retrieval until the response is imminent.

The fact that people tend to synchronize multiple streams of action that in principle could be handled independently is not a novel result (Krampe, Mayr, & Kliegl, 2005). In dual-task research on the psychological refractory period, for example, response grouping is a common phenomenon and an important challenge to the interpretation of standard dual-task effects (Ulrich & Miller, 2008). It is less clear, however, why people tend to synchronize different streams of behavior. The highly regular pattern seen in the within-subject scatterplots between speaking and acting suggests that a rather strong, Gestalt-like force is at work here (Klapp & Jagacinski, 2011). Perceptual gestalt laws allow the visual system to arrive at parsimonious descriptions of raw sensory input. In the action domain, a Gestalt-like tendency to synchronize different outputs that are related to the same subgoal would allow a parsimonious representation of complex, multicomponent action events that may overall facilitate control of behavior. In the context of dual-task research, there is actually some evidence that grouped responses can be executed more swiftly than independent responses (Miller & Ulrich, 2008). In the context of serial-order control, verbalization should perhaps be better viewed as self-labeling rather than self-cuing: Naming ongoing actions *as they occur* creates temporally distinct memory traces of each sequential event, and this in turn may help updating one's current position in a complex action/event sequence.

Consistent with the idea that synchronized speaking produces more robust sequential representations, we found that subjects in the Early Speaking group made more sequence errors, in particular when serial-order control demands were high (i.e., due to positional interference and chunk boundaries). We also found that in the Early Speaking condition, people took longer to speak the task label *after* points of high positional interference—where efficient sequence updating would have been particularly critical but, due to the desynchronized speaking, did not happen.

Our conclusion that synchronized speaking and acting may help sequence updating is consistent with results reported by Bryck and Mayr (2005). These authors had revisited previous claims that preventing subjects from speaking (through articulatory suppression) during task switching selectively affects the process of switching from one task to the next, presumably because it interferes with retrieving upcoming task demands (Miyake et al., 2004). However, inconsistent with the retrieval/switching hypothesis and consistent with the sequencing/updating idea, they found that articulatory suppression affected switch and no-switch trials alike, but only when sequencing demands were high (for related results, see also Saeki & Saito, 2009).

Combined, our results suggest that in the absence of additional external constraints, people synchronize speaking and acting during sequential control. We also showed that proactive, desynchronized verbal control is possible and does aid fluid performance. However, such proactive control comes at the cost of more sequencing errors, presumably because of less robust updating of the current position in the sequence. Thus, speaking during serial-order control can serve two different functions—preparation of the next sequential element or updating of the current sequence position. Emphasis on either one may come at the expense of the other, requiring people to carefully calibrate how language is used to control complex action.

References

Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General, 130,* 641–657. doi:10.1037/0096-3445.130.4.641

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 5.1. 44) [Computer program]. Retrieved from http://www.fon.hum.uva.nl/praat/

Bryck, R. L., & Mayr, U. (2005). On the role of verbalization during task set selection: Switching or serial order control? *Memory & Cognition, 33,* 611–623. doi:10.3758/BF03195328

Goschke, T. (2000). Intentional reconfiguration and involuntary persistence in task set switching. In S. Monsell & J. Driver (Eds.), *Attention and Performance XVIII: Control of cognitive processes* (Vol. 18, pp. 331–355). Cambridge, MA: MIT Press.

Hommel, B., Pratt, J., Colzato, L., & Godijn, R. (2001). Symbolic control of visual attention. *Psychological Science, 12,* 360–365. doi:10.1111/1467-9280.00367

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin, 136,* 849–874. doi:10.1037/a0019842

Klapp, S. T. (1976). Short-term memory as a response preparation state. *Memory & Cognition, 4,* 721–729. doi:10.3758/BF03213240

Klapp, S. T., & Jagacinski, R. J. (2011). Gestalt principles in the control of motor action. *Psychological Bulletin, 137,* 443–462. doi:10.1037/a0022361

Koch, I. (2003). The role of external cues for endogenous advance reconfiguration in task switching. *Psychonomic Bulletin & Review, 10,* 488–492. doi:10.3758/BF03196511

Koch, I., Philipp, A., & Gade, M. (2006). Chunking in task sequences modulates task inhibition. *Psychological Science, 17,* 346–350.

Krampe, R. T., Mayr, U., & Kliegl, R. (2005). Timing, sequencing, and executive control in repetitive movement production. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 379–397. doi:10.1037/0096-1523.31.3.379

Kray, J., Eber, J., & Karbach, J. (2008). Verbal self–instructions in task switching: A compensatory tool for action–control deficits in childhood and old age? *Developmental Science, 11,* 223–236. doi:10.1111/j.1467-7687.2008.00673.x

Mayr, U. (2009). Sticky plans: Inhibition and binding during serial-task control. *Cognitive Psychology, 59,* 123–153. doi:10.1016/j.cogpsych.2009.02.004

Mayr, U., & Keele, S. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology: General, 129,* 4–26.

Miller, J., & Ulrich, R. (2008). Bimanual response grouping in dual-task paradigms. *Quarterly Journal of Experimental Psychology, 61,* 999–1019. doi:10.1080/17470210701434540

Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. C. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica, 115,* 123–142. doi:10.1016/j.actpsy.2003.12.004

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7,* 134–140. doi:10.1016/S1364-6613(03)00028-7

Rosenbaum, D. A., Kenny, S. B., & Derr, M. A. (1983). Hierarchical control of rapid movement sequences. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 86–102. doi:10.1037/0096-1523.9.1.86

Saeki, E., & Saito, S. (2009). Verbal representation in task order control: An examination with transition and task cues in random task switching. *Memory & Cognition, 37,* 1040–1050. doi:10.3758/MC.37.7.1040

Schneider, D. W. (2007). Task-set inhibition in chunked task sequences. *Psychonomic Bulletin & Review, 14,* 970–976.

Schneider, D. W., & Logan, G. D. (2006). Hierarchical control of cognitive processes: Switching tasks in sequences. *Journal of Experimental Psychology: General, 135,* 623–640. doi:10.1037/0096-3445.135.4.623

Ulrich, R., & Miller, J. (2008). Response grouping in the psychological refractory period (PRP) paradigm: Models and contamination effects. *Cognitive Psychology, 57,* 75–121. doi:10.1016/j.cogpsych.2007.06.004

Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: MIT Press.

Received April 1, 2013
Revision received September 5, 2013
Accepted September 6, 2013 ■