



Published in final edited form as:

J Exp Psychol Gen. 2014 February ; 143(1): 295–311. doi:10.1037/a0031858.

INTERFERENCE BETWEEN CONVERSATION AND A CONCURRENT VISUOMOTOR TASK

Timothy W. Boiteau,

University of South Carolina

Patrick S. Malone,

University of South Carolina

Sara A. Peters, and

Newberry College & University of South Carolina

Amit Almor

University of South Carolina

Abstract

We report two experiments during which participants conversed with either a confederate (Experiment 1) or a close friend (Experiment 2) while tracking a moving target on a computer screen. In both experiments, talking led to worse performance on the tracking task than listening. We attribute this finding to the increased cognitive demands of speech planning and monitoring. Growth Curve Analyses of task performance during the beginning and end of conversation segments revealed dynamical changes in the impact of conversation on visuomotor task performance, with increasing impact during the beginning of speaking segments and decreasing impact during the beginning of listening segments. At the end of speaking and listening segments, this pattern reversed. These changes became more pronounced with increased difficulty of the task. Together, these results show that the planning and monitoring aspects of conversation require the majority of the attentional resources that are also used for non-linguistic visuomotor tasks. The fact that similar results were obtained when conversing with both a confederate and a friend indicates that our findings apply to a wide range of conversational situations. This is the first study to show the fine-grained time course of the shifting attentional demands of conversation on a concurrently performed visuomotor task.

INTRODUCTION

One type of multitasking that has been of particular interest to cognitive psychologists involves concurrent performance of tasks that require different modalities and thus superficially seem to pose little risk of cross interference (e.g., Alm & Nilsson, 1994; Almor, 2008; Baddeley, 1986; Baddeley et al., 1986; Becic et al., 2010; D'Esposito et al., 1995; Drews, Pasupathi, & Strayer, 2008; Horrey & Wickens, 2006; Just et al., 2001; Just, Keller, & Cynkar, 2008; Kubose et al., 2006; Kunar et al., 2008; Lavie, 2005; Levy et al., 2006; Logie et al., 2004; Rakauskas, Gugerty, & Ward, 2004; Salthouse et al., 1995; Strayer, Drews, & Johnston, 2003; Strayer & Johnston, 2001; Wickens, 2008; Wickens, 1984). The simultaneous performance of a linguistic task together with a visual, spatial, and/or motor task is of special importance for researchers interested in the connection between language and other cognitive systems, and anyone required to address real-life situations that involve

such multitasking. One particularly salient example of this type of multitasking is talking on the phone while driving. While there has been much work on these topics, none of the research thus far has considered the intricate dynamics of conversation and the consequences of these dynamics on the allocation of attentional resources. Identifying and differentiating between the specific aspects of conversation that affect performance on other tasks is necessary for the understanding of both why and when language interferes. The present research thus aimed to examine the subtle dynamics of conversation and their effects on a concurrent visuomotor task.

Past research supports a cross-modal model of attention, in which both auditory language processing and visual task performance tap into a central share of resources (Kunar et al., 2008). Such an amodal model of attention can explain participants' poor performance in visuospatial and linguistic dual-task studies (e.g., Almor, 2008; Drews, Pasupathi, & Strayer 2008; Strayer, Drews, & Johnston, 2003), despite the two tasks seemingly engaging dissociable cognitive systems. Whereas previous research supports this shared resource model, it provides only some general clues as to which aspects of language, and more specifically of conversation, require these resources. For example, Kunar et al. (2008) compared participants' performance on a multiple-object-tracking task during novel word generation, speech shadowing, and no secondary task conditions. Speech shadowing did not result in any detriment to the multiple-object-tracking task while word generation did. This finding suggests that the perceptual and articulatory aspects of speech do not cause interference but that speech planning does. This likely reflects the fact that speech perception, as well as the motor aspects of speech, are largely automatic and minimally demanding, whereas composing utterances takes up a greater amount of cognitive resources. We would therefore expect that, if planning is what requires the most attentional resources, then interference on performing a simultaneous visuomotor task would be most noticeable during the stages of a conversation that involve planning, which are most likely the moments before and during speech.

Before addressing in detail the resource demands of the different stages of conversation, it is necessary to consider in more detail how cognitive resources are distributed in dual-task contexts. One specific account of resource distribution is Wickens's (2008) multidimensional resource allocation model, which can account for the majority of findings in the multitasking literature. This model defines resources according to three different dimensions: stage of processing (perception/cognition or response), code of processing (spatial or verbal), and modality (visual or auditory). The model predicts that the more dimensions shared by concurrent tasks, the worse the performance. For example, simultaneous performance on a visual-verbal task and an auditory-verbal task (which share the code of processing dimension) should be worse than performance on a visuospatial task and an auditory-verbal task (which differ along said dimension). By this account, a visuomotor task such as target tracking and a language task involving natural conversation differ on the dimensions of code of processing and modality, but are similar in the dimension of stage of processing (both involving, at the very least, perception and response). Wickens's model therefore predicts that there should be some interference between the two tasks, especially when the language task is most similar to the visuomotor task. Specifically, based on this model, we claim that speaking during a conversation is similar to target tracking in that both tasks involve planning, execution, and monitoring of planned and executed behavior (which load both the perception/cognition and response stages), whereas listening to someone speak during a conversation shares only the perception/cognition stage of processing with the visuomotor task, and this (as we will argue below) is only minimally demanding. We now expound this claim by examining the dynamics of conversation in greater detail.

Most language production models posit at minimum three stages: construction of a message, which is held briefly in working memory, formulation of that message into phonetic strings, and articulation of those strings through a motor program (Bock, 1995; Levelt, 1983). In order to account for speech repairs, recent production models also include a parser and monitor component (Horton & Keysar, 1996, Postma, 2000). The parser is similar to one used in language comprehension, except that this parser is also able to parse out inner speech. Its main purpose is to reconstruct the meaning of the generated pre-acoustic phonetic strings and actual speech output so that the monitor can compare them with the intended message and certain linguistic standards. If the monitor detects an error, it sends instructions for repair to other components in the model (Postma, 2000). According to this general production model, the primary cognitive resource demands of language production occur during both the first stage of message planning, which relies on an early conceptual loop for message formation (Blackmer & Mitton, 1991), and during the fourth stage of parsing and monitoring, in which the appropriateness of the planned utterance is verified and compared with the intended message as both are held in working memory (Oomen & Postma, 2002; Horton & Keysar, 1996). According to Wickens's (2008) multidimensional resource allocation model, speech planning would involve both the perception/cognition and response stages of processing, while monitoring would be only perception/cognition (Wickens, 2002; Shallice, McLeod, & Lewis, 1985). Thus, speech production occupies both stages of processing according to Wickens's model, necessarily producing interference on any other concurrent task.

Unlike language production, which poses clear cognitive resource demands for planning and monitoring, the resource demands of language comprehension are less clear and likely depend on the complexity of the input. According to several accounts, processing simple constructions during comprehension does not necessarily require many general cognitive resources (Gibson, 1998). Although more resources may be required for the full processing of complex constructions and nonlocal dependencies during comprehension (Gibson & Pearlmutter, 1998), complex constructions that are difficult to process may be less common in normal conversation (Chafe, 1985; Sacks, Schegloff, & Jefferson, 1974), which is characterized by short exchanges and interactive turn taking (Clark, 1996; Sacks, Schegloff, & Jefferson, 1974). Moreover, even when encountering complex constructions, listeners may not engage in full processing, relying instead on quick and efficient interpretation strategies (Ferreira, Bailey, & Ferraro, 2002; Frazier & Rayner, 1982), or multiple cues (MacDonald et al., 1994). For example, according to Ferreira, Christianson, & Hollingworth's (2001) *good enough language processing*, comprehenders do not create full, resource-draining representations, but rather form quick, "good enough" interpretations. According to the constraint-based view (MacDonald, 1999; Trueswell, Tanenhaus, & Kello, 1993), interpretation during comprehension is achieved by the consideration of multiple probabilistic cues reflecting an underlying distribution determined by production constraints.

Thus, it is mainly language production that is subject to cognitive resource constraints, while language comprehension is mostly sensitive to the distributional patterns that emerge as a result of these production constraints (Gennari & Macdonald, 2009). In summary, according to multiple views, listening during conversation does not require any planning and requires only limited monitoring so as to ensure the interpretation of the input is congruent with the context; it therefore likely requires fewer cognitive resources than speaking. This supports our main claim that speaking (and especially speech planning and monitoring) during a conversation will likely cause greater interference on the performance of a concurrent visuomotor task than listening during a conversation.

Together, theories of dual-task performance and theories of language production and comprehension predict differences in the allocation of attention during different parts of a

conversation. Generally, these theories predict that production is more demanding than comprehension, which has been empirically supported (e.g., Almor, 2008; Kunar et al., 2008). More importantly, these theories lead to a specific prediction about the time course of conversational impact on the performance of an ongoing visuomotor task. In particular, these theories predict increasing cross-task interference as interlocutors are transitioning from listening to speaking and thus are progressively engaging in more planning and monitoring. However, no previous study has examined these moment to moment changes in attentional demands. Partially this may be due to the reliance on coarse analytical methods in the literature, resulting in a gap in our understanding of dual-task performance. Here we aim to remedy this situation by measuring and analyzing statistically the fine temporal changes in visuomotor task performance as a function of conversational dynamics.

The final factor that we consider is the difficulty and resource demands of the visuomotor task. When a participant in a conversation has to simultaneously perform a demanding visuomotor task, the models of resource sharing and resource demands during conversation predict that there will be fewer resources available for the conversation, which will be especially problematic during the more demanding speaking parts of conversation (Almor, 2008; Kunar et al., 2008). Israel et al. (1980) showed that increasing the difficulty of a tracking task resulted in less efficient performance on a concurrent signal detection task. Given their results, we expect that when the visuomotor task is difficult, the differences between the effects of planning, speaking, and listening on the performance of the concurrent visuomotor task will be greater than when the visuomotor task is easy.

The literature includes a number of studies that examine the effects of natural conversation on a concurrent visuomotor task, especially studies related to driving and talking on cell phones. Two meta-analyses (Caird et al., 2008; Horrey & Wickens, 2006) found that talking on a cell phone while driving negatively impacts reaction time but does not affect tracking or lane maintenance. At the same time, these meta-analyses also highlight a major drawback of a number of previous studies in how researchers have dealt with the linguistic component. Most studies contrasted a conversation condition with a no conversation condition or, at the most complex, a speaking condition with a listening condition (Kunar et al., 2008; Strayer, Drews, & Johnston, 2003). It is especially problematic when these studies use findings from simple linguistic or information-processing tasks (e.g. a word generation task; Strayer & Johnston, 2001) to make claims about full-fledged conversation, which is an interactive and collaborative joint activity (Clark, 1996). Indeed, Horrey & Wickens (2006) showed that conversation tasks were in general more detrimental to secondary task performance than information-processing tasks.

One previous study, Almor (2008), overcame some of these problems by investigating the effects of an exchange-based language task on vision-based attention tasks. This study found that participants' performance on a tracking task was worse when they were speaking than when they were listening, and was worst in the moments when they were preparing to speak. Almor attributed these results to the greater attentional demands of speaking, and particularly of speech planning, relative to listening. While Almor's results are compatible with our general claims in that they are indicative of attentional differences between comprehension and production, a few design issues prevent us from being able to generalize these results to natural conversation. First, the nature of the language task used by Almor was artificial. Participants listened to prerecorded narratives and responded to specific questions about what they heard, thus preventing a lot of the overlap and interactivity inherent to natural conversation (Schegloff, 2000). Second, they spent much more of their time listening than speaking. When they did speak, there was no specific feedback from their interlocutor, preventing participants from being able to ask questions, alter the topic of conversation, or more generally participate in the conversation as active contributors (Clark,

1996). Third, the lack of feedback meant participants were not pressed to monitor their production, as errors in their speech would go unnoticed. They may also have spent more time planning what to say before speaking (as opposed to composing as they spoke), as they were answering specific questions and as there was no interlocutor whose patience would be tried by a delay in response (Arnold, Kahn, & Pancani, 2012). Fourth, because the narratives used by Almor were instructions involving visuospatial information (how to build an igloo, how to navigate by the sun or stars, how to set up a fish tank, and how to build a compost pile), it may be that his results reflect the overlap in content between these narratives and the secondary visuomotor task (Kubose et al., 2006). Fifth and most importantly, Almor (2008) did not examine the fine temporal dynamics of the impact of conversation on the concurrent visual task, which is our main focus here.

The present study is designed to overcome these issues by examining how an unscripted natural conversation with a remote partner interferes with performing a visuomotor task. Most importantly, we record and analyze dual-task performance on a fine-grained timescale, using a technique novel to the dual-task literature (described in detail below). By doing so, we obtain a clearer picture of how attention is distributed during conversation in a multitasking scenario. Because we use unscripted conversations, we add two additional categories (overlapping speech, i.e., when participants and their conversation partners were speaking simultaneously; and pauses between speaking segments) to the three (listening, speaking, and preparing to speak) into which the linguistic data were segmented in Almor (2008). As previous research has found, both listening and talking can lead to worse performance on a secondary visuomotor task (Kubose et al., 2006), so it is probable that doing both simultaneously will lead to even poorer performance. Pause segments may reflect difficulty in speech planning and monitoring intended utterances and should thus lead to more interference on the concurrent task than listening (Ramanarayanan et al., 2009).

In our experiments, participants only engaged in remote conversation, simplifying our conversations along the code of processing (verbal only, allowing us to ignore body language and other visual paralinguistic cues) and modality (auditory only) dimensions of Wickens's model of attention (Wickens, 2008). Our main concern then was to develop specific hypotheses about how attention during conversation varies along the dimension of stage of processing. Listening involves only the perception/cognition dimension. Note that some degree of utterance planning can and does occur during listening portions of the conversation, but it would be inefficient for interlocutors to assign too much attention to planning while the conversation partner is speaking, as the subsequent response could change based on the as-yet-unheard linguistic input. Also, intensive speech planning could interfere with comprehension enough to result in the need for utterances to be repeated. Thus, it makes more sense for an interlocutor (especially one performing a cognitive-resources-consuming secondary task) to postpone speech planning until near the end of the conversation partner's utterance. Following this logic, we expect to see participants performing worse at the end of segments of listening than at the beginning. Our fine-grained temporal analyses will be able to test this hypothesis directly. Speech planning will occupy considerably more resources than listening, as speakers will be holding in memory those topics that the conversation partner mentioned and that will need to be addressed in the upcoming response. In addition, there will be the propositions of the response held in memory before undergoing syntactic and phonological translation. Thus, speech planning actually consists of elements of perception/cognition and response stages of processing, but the degree to which each is involved will naturally depend on the type of response and the linguistic input from the conversation partner. During speech, propositions are translated from memory into linguistically viable utterances, and speech planning continues throughout until near the end of an utterance. Also, as explained above, speakers need to monitor what they say as it is being produced (Postma, 2000; Levelt, 1983). Based on this model we

therefore predict that as one moves from listening to planning to speaking, additional cognitive resources will be required and thus there will be increasing interference on a concurrent secondary task. Overlapping speech may lead to further detriments on the concurrent task, as speakers could be attempting to simultaneously parse their own and their partners' speech. And, as we have mentioned, pauses in speech likely reflect difficulty in planning and monitoring the appropriateness of planned speech, leading to worse performance on the visuomotor task. In the work we report here, we had participants converse with a confederate (Experiment 1) or a friend (Experiment 2) as they performed a target tracking visuomotor task.

EXPERIMENT 1

The goal of Experiment 1 (E1) was to test whether the findings of Almor (2008), which used an exchange-based language task, could generalize to natural conversation. Specifically, we wanted to test whether conversation interferes with the performance of a concurrent tracking task and identify the conversational stages where this effect is the most detrimental. We also wanted to test whether this interference is affected by the difficulty of said visuomotor task. Consequently, this experiment employed a dual ball-tracking and language task design similar to Experiment 2 in Almor (2008), with two main differences: first, the linguistic task involved participants conversing with a confederate, instead of listening and responding to pre-recorded narratives; second, the ball-tracking task was altered, such that the speed of the ball was constrained into slow, medium, or fast 1 minute blocks; whereas in Almor's study, the ball speed randomly alternated across the entire range. This allowed us to test directly the effect of target moving speed on performance. Our main interest in this study was in what happens to a participant's tracking performance during listening and speaking segments, the brief moments between listening and talking, moments in which the participant and his/her partner overlap, and pauses in speech during the slow, medium and fast tracking task conditions. We expect that performance on the task will be best during listening, worse during talking, preparing to speak, and pauses, and that overlapping speech segments will show an increased detriment to the visuomotor task relative to talking.

Methods

Participants—Twenty-four University of South Carolina undergraduates, recruited through undergraduate psychology classes, participated in this study for extra credit after signing a consent form. Data from 3 participants were unusable due to equipment malfunction or experimenter errors. Of the participants used, 14 were female and 7 were male. Mean participant age was 19.62. The confederate was a female undergraduate (age 27 yrs.), who was aware of the hypothesis that in general conversation leads to worse performance than a control no conversation condition but was unaware of the specific hypotheses about talking and speech planning being worse than listening.

Procedure

Linguistic Task: In this experiment, participants conversed with the confederate for 15 minutes, communicating via microphones. The participant was seated at the center of a surround sound speaker system through which the confederate's voice was broadcast. The dyad was instructed to talk about whatever interested them. The confederate was given a sheet with a list of topics, which she could use if they ran out of things to discuss during the experiment. The conversations that were recorded included a wide range of topics, including classes, mutual friends, entertainment, etc., almost all of which were included in the topic sheet. The confederate was instructed to keep the participant engaged and talking.

In determining whether conversations were actually natural or typical, we relied on the judgment of the participants and the confederate, who rated the conversation as a whole and their partner's performance as a converser in a survey given at the end of the experiment. In the survey, both participants and the confederate rated the naturalness of the conversation and the degree to which the topics of the conversation were typical of normal conversations. They were also asked how distracted they and the conversation partner were during the experiment. All responses were made on a 7 point Likert-type scale. We do not include comparisons of post-experiment survey results in our E1 Results section, saving them for a between experiment comparison in Experiment 2.

Visuomotor Task: Participants were instructed to perform a smooth pursuit task while conversing with the confederate. In this visuomotor task, participants used the computer mouse to track a moving ball (with a radius of 12 pixels) on a computer monitor with a square 15" screen and a resolution of 800 × 600. The ball's movement was controlled by sinusoid functions, resulting in smooth motion with periodic changes in speed and direction. The ball moved at three different speeds (Low = 96.15 pixels/second, Med = 293.65 pixels/second, and High = 484.39 pixels/second), changing speed every minute pseudorandomly so that the participant encountered each speed block five times during the experimental stage. The computer upon which this task was run recorded the distance between the cursor and the moving target as well as the target's speed at 12ms intervals. The computer also emitted a periodic beep, unheard by the participants but recorded by the experimenter's computer and later used to synchronize the motor task with the conversation. Participants started the experiment with a 3 minute practice, during which the ball cycled through one iteration of each speed level. This was followed by 3 more minutes of performing the task without conversation, during which the cycles of speed repeated. These latter 3 minutes were used as a control baseline in analyses mentioned below. Following the period of no conversation, there were 15 minutes of concurrent conversation and visuomotor task.

Segmentation of Conversation: The recorded conversational data were segmented after the experiment into six distinct categories: Control (the 3 minute block of visuomotor task before the conversation), Listen (segments of the confederate speaking), Talk (segments of the participant speaking), Overlap (segments of both parties speaking simultaneously), Prepare (silences following a Listen segment and preceding a Talk segment), and Pause (silences longer than half a second following a Talk segment and preceding another Talk segment). While this system of classification may not capture the full range of possible conversational interactions, it does capture the main categories that are likely relevant for the distribution of attention, our primary concern in this study. During the Talk segments, the participant is producing and composing with no feedback from the confederate, increasing demands on the conceptual loop and error detection monitors. While they are listening, they are comprehending and (potentially) composing a response. The gaps of silence between Listen and Talk segments should (as they did in Almor, 2008) show the effect of the cognitive demands of composing a suitable utterance. Similarly, gaps of silence between Talk segments might also reflect the demands of speech planning. Overlaps should account for dramatic shifts in attention, where participants are both producing and comprehending.

Growth Curve Analyses (GCA): GCAs are a statistical technique used to analyze changes in performance across time. They consist of fitting to the data different multilevel regression models with terms representing time and choosing an optimal model based on fit and parsimony (Long, 2012). The terms that are included in the chosen model and their coefficients indicate which factors are important predictors of the data. Growth curve models typically include terms that represent the effect of time in the first level and terms

that express the effect of conditions on the intercept and time in the second level. Identifying the terms in the second level that are important for fitting the data allowed us to assess which factors in our experiment affected changes in tracking performance across time. For this set of analyses, we focus only on Listen, Prepare, and Talk because our hypotheses about dynamic shifts in attention during conversation apply most clearly to these segments. Also, we chose to look at only the beginning and end of segments, because our theoretical predictions apply most clearly to the beginning and end of segments and because of the large variability in segment durations. Specifically, we predicted that performance will deteriorate towards the end of Listen segments as participants are beginning to plan their own utterances, will further worsen during Prepare segments, and will decline even further as participants start speaking, because at that time they are likely planning the rest of what they will say in addition to monitoring their speech. In contrast, the end of speaking segments should be characterized by improving performance as only self-monitoring is required and planning is no longer necessary. By the same logic, the end of listening segments should be characterized by decreasing performance as participants are beginning to plan what they will say next as the conversation partner is wrapping up their turn. We further predicted that this pattern will become more pronounced with increased tracking task difficulty (Israel et al., 1980; Wickens, 2002). To accommodate possible nonlinear changes in tracking performance, we tested models that included first order (linear), second order (quadratic), third order (cubic) and fourth order (quartic) time effects in our models.

In order to compare performance at the beginning and end of segments, we looked at the 40 samples (480 ms) at the beginning and end of Listen and Talk segments that were longer than 1 second. Because many Prepare segments were relatively short, we only looked at the 40 samples at the end of these segments and included them in a separate set of analyses of segment ends. Our reason for only looking at 480ms windows was that segments were quite variable in duration. We wanted to include as many segments in the analysis as possible, but could not include segments shorter than our cutoff point, preferring instead to have an equal number of measurements for each point in time. Overall, we carried out two sets of growth curve analyses to test for differences between conditions on performance. In the first set, we looked at differences in tracking performance in the first vs. last 480 ms of Talk and Listen segments under different target velocity conditions. In the second set of analyses, we looked at the differences in tracking performance in the last 480 ms of Prepare vs. Talk vs. Listen segments under different target velocity conditions. Both sets of analyses followed the same procedure. We first fit the data with a base random coefficients model that included the time terms for all orders but no terms representing any of the conditions (Model 1 in Table 1). We then created a conditional intercept model by adding an intercept interaction term of all the conditions (Model 2 in Table 1) and a linear model by adding to the intercept model an interaction term of all the conditions with linear time change (Model 3 in Table 1). We followed a similar procedure to create quadratic, cubic and quartic models (Models 4, 5 and 6 respectively in Table 1). The effects of intercept, linear time, and quadratic time were added as random factors varying by subject to all models. We then used both an AIC evidence based method (Forster & Sober, 1994, 2011; Long, 2012) and maximum likelihood estimates to compare the models in order to determine the necessary time order of the model. When the two methods yielded different results, we examined the data to see whether the more complex model provided any information useful for our goal and if not, chose the simpler model identified by either method (Long, 2012).

Following the choice of time order for the model, we tried to simplify the interaction terms in order to identify the simplest model of that time order that would still fit the data. We always started by attempting to replace a highest order interaction with a series of interactions of one order lower. We report for both sets of analyses the selection criteria for the models we compared and include graphs that show the fitted data as well as the

predicted values based on the chosen model. Because the interpretation of the coefficient estimates of these models is complicated, we follow common procedure in discussing the results according to the terms included in the chosen model and the graphs (Long, 2012). The estimated coefficients of the final chosen models are listed in the online supplementary material.

Analyses were carried out using the lme4 package (Bates, Maechler, & Bolker, 2011) in R (v.2.14.1, (R Development Core Team, 2011)). To eliminate possible multicollinearity issues, the time variables were represented by orthogonal, mean centered polynomials. Note that due to the centering of the time variables, intercept coefficients represent predicted values at the middle time point in the analyzed time range (240 ms from the start of the time window).

Results

In order to test whether the participants and confederate made comparable contributions to the conversations, we compared the average number of Listen and Talk segments produced per conversation and found no significant difference, Listen $M = 225.38$, $SD = 19.98$, Talk $M = 211.29$, $SD = 31.36$, $t(20) = 1.77$, $p = .09$, $\eta_p^2 = .14$. We analyzed the duration of Listen and Talk segments and found that participants spent more time talking than they did listening, Listen $M = 1.61$ s, $SD = .19$, Talk $M = 2.14$ s, $SD = .39$, $t(20) = 5.97$, $p < .001$, $\eta_p^2 = .64$. Next, we divided the Overlap category into instances of participants interrupting and being interrupted according to the following parameters: participant interrupting was defined as an Overlap segment following Listen and followed by Talk, and partner interrupting was defined as an Overlap segment following Talk and followed by Listen. On average, the confederate interrupted the participants more than twice as often as the participants interrupted her, confederate interruptions, $M = 27$, $SD = 11$, participant interruptions, $M = 11.5$, $SD = 5.46$, $t(20) = 6.98$, $p < .001$, $\eta_p^2 = .71$. Thus, overall, these analyses show that participants were making approximately the same number of contributions to the conversation as the confederate, but they took up more time speaking and were interrupting less often than the confederate.

We next aimed to examine whether participants' conversational fluency was affected by the difficulty of the tracking task by analyzing changes in speech rate and number of disfluencies according to the target speed during Talk segments. Speech rate was calculated by dividing number of words per segment by the segment duration. We ran a repeated measures ANOVA with the factor Speed (Low, Med, High) and found a significant effect, Low $M = 4.17$ (words/second), $SD = .52$, Med $M = 3.96$, $SD = .44$, High $M = 3.99$, $SD = .38$, $F(2, 40) = 3.34$, $p = .05$, $\eta_p^2 = .14$. Pairwise comparisons revealed that this effect was driven by a higher speech rate during Low than Med tracking task conditions, $p = .03$, $\eta_p^2 = .22$, while there was no significant difference in speech rate between Med and High conditions, $p = .63$, $\eta_p^2 = .01$. However, the Low/Med comparison did not survive correction for multiple comparisons using Rom's (1990) method at $\alpha = .05$ and should therefore be interpreted with caution (with 2 contrasts Rom's method corrects α to .025). Although disfluencies come in many forms, fillers *uh* and *um* were of particular interest to us, as previous research has shown that they not only represent planning difficulty but also the communication of these difficulties to conversation partners (Bortfield et al., 2001; Clark, 1994). We thus examined number of *uh* and *um* fillers per Talk segment and divided this figure by the word count of that same segment. We ran a repeated measures ANOVA with the factor Speed (Low, Med, High) but failed to find a significant effect, Low $M = .04$ (disfluencies/word), $SD = .02$, Med $M = .05$, $SD = .04$, High $M = .04$, $SD = .03$, $F(2, 40) = .6$, $p = .55$, $\eta_p^2 = .03$. Overall, these analyses show that tracking task difficulty had a minor effect on speech rate and none whatsoever on overt disfluency production.

In order to analyze smooth pursuit performance, we calculated the mean distance to target in each of the conversation segments, which we submitted to a 3×6 repeated measures ANOVA with the factors Speed (Low, Med, High) and Conversation (Control, Listen, Overlap, Pause, Prepare, Talk). The analysis found a significant main effect of Speed, $F(2, 40) = 467.91, p < .001, \eta_p^2 = .96$, and a significant main effect of Conversation, $F(5, 100) = 8.35, p < .001, \eta_p^2 = .29$, but the interaction failed to reach significance, $F(10, 200) = 1.33, p = .22, \eta_p^2 = .06$ (see Figure 1). A series of planned pairwise comparisons of mean distance in the different conversational conditions revealed that distances between the target and cursor were smaller during Listen compared to Talk, $t(20) = 5.68, p < .001, \eta_p^2 = .59$, during Prepare compared to Talk, $t(20) = 5.03, p < .001, \eta_p^2 = .42$, and during Overlap compared to Talk, $t(20) = 2.5, p = .02, \eta_p^2 = .25$. However, Pause and Talk did not differ significantly, $t(20) = 1.5, p = .15, \eta_p^2 = .02$ (see Table 2 for means). These results hold following correction for multiple comparisons using Rom's (1990) method at $\alpha = .05$, except for the Overlap vs. Talk contrast, which was only marginally significant (with 4 contrasts Rom's method corrects α to .0127) and should thus be interpreted with caution.

In order to compare our results to Almor's (2008) main effect of Conversation, we used the *MSes* reported by Almor to calculate the effect size from that study ($\eta_p^2 = .06$) and perform a power analysis using G*Power (Buchner et al., 2012). This analysis indicated that 12 participants are required in order to achieve .80 power ($\alpha = .05$) in our experiment to detect the main effect of conversation observed by Almor. The much higher power in our study than Almor's reflects the higher number of observations per participant in our study ($M = 211, SD = 31$), than in Almor's (exactly 47 observations per participant). Because Almor did not report effect sizes for individual comparisons, we could not perform power analyses for individual contrasts.

GCA was used to compare models of distance to target with different predictors of change over 40 samples spaced 12ms apart within segments. In our first set of analyses (Analysis 1a) the predictors included whether the time window was at the beginning or end of segments, how fast the target was moving (Low, Medium, or High), and whether it was a Talk segment or a Listen segment. Table 3a shows the results of model comparison according to the AIC weight of evidence criterion (Long, 2012). As can be seen in the table, the quadratic model was the most plausible as it had probability close to 1 among the compared models. Table 3b shows the results of model comparison according to log likelihood criteria, which also found that the quadratic model provided a better fit than the linear model, but none of the higher order models improved this fit. Therefore, the full quadratic model was chosen as the best fitting model, indicating that all the predictors interacted to affect the change and rate of change in distance to target. Figure 2 shows the model fitted to the observed data. The full parameter estimates of the model can be found in Online Supplement 1. As is indicated by the inclusion of the interaction effects in the chosen model, performance differed between the different conditions. The graph shows there was little difference between the Listen and Talk conditions at the beginning and end of conversation segments across all speeds. During all Begin segments there was a divergence, characterized by gradual improvement in tracking task performance during Listen and gradual deterioration in performance during Talk. During End segments this effect reversed, such that tracking task performance during Listen deteriorated and during Talk it improved. Although this trend was found during all speed conditions, it appeared to become more exaggerated as the tracking task became more difficult. The presence of the quadratic interaction in the chosen model seems to reflect the nonlinear change in performance at the end of Listen segments in the High speed condition. This suggests that participants rapidly modulate resource allocation at the end of Listen segments as they are beginning to plan their next contribution to the conversation, indicating that participants may have been anticipating the ends of the confederate's utterances.

Our second set of analyses (Analysis 1b) included the Prepare segments in addition to the Listen and Talk segments but only included time windows at the end of segments. Therefore, in these analyses, the predictors included how fast the target was moving (Low, Medium, or High), and conversational segment (Talk, Listen, or Prepare). Table 4a shows the results of model comparison according to the AIC weight of evidence criterion (Long, 2012). As can be seen in the table, the quadratic model was the most plausible as it had a probability of .95 among the compared models. Because the quadratic model was the most probable one and the more complex models did not add any new information about the simpler linear and quadratic terms, we chose the simpler quadratic model for these data. Figure 3 shows the model fitted to the observed data. The full parameter estimates of the model can be found in Online Supplement 2. As shown by model and graph, at the end of segments, Prepare was characterized by a gradual deterioration in tracking task performance. Moreover, performance during Prepare segments generally transitioned from the level seen at the end of Listen segments to the level seen at the beginning of Talk segments. Additionally, speed had an interesting effect on performance: during Low speed conditions, Prepare matched closely with Listen; during Med, Prepare was somewhere between Listen and Talk; and during High, Prepare patterned closely with Talk.

Discussion

The difference in tracking task performance between Listen and Talk replicates Almor (2008) and shows that the finding from that paper can be generalized to freeform conversation and is not just a characteristic of the linguistic task Almor used. The GCAs add to this finding that the demands of conversation change dynamically in line with our hypotheses about the increases in speech planning and monitoring. Performance at the beginning of Talk segments deteriorates over time but recovers as the segment nears its end. The reverse is true for Listen segments. These results can be ascribed to shifts in allocation of resources to and from planning and monitoring speech. Indeed, the fact that performance worsens during the beginning of Talk segments indicates that it is the combined effect of monitoring and speech planning that is most detrimental to concurrent performance on the tracking task. The fact that performance deteriorates towards the end of Listen segments indicates that participants begin to plan their utterances before Listen segments end. According to Clark (1996), interlocutors aiming to facilitate smooth transitions between turns, gauge the timing of the end of their partners' utterances, thus effectively reducing the amount of time spent during Prepare segments.

We failed to replicate Almor's (2008) finding that preparing to speak was more detrimental to the concurrent visuomotor task than speaking. This difference could be attributed to the extemporaneous nature of talking during natural conversation, mainly that planning is pushed closer to the moment of (and onto) the utterance. This possibility was supported in the GCAs where we saw a rising slope in tracking task detriment during Prepare segments. Indeed, the ends of Prepare segments were nearly indistinguishable from performance during Talk segments. Another factor that may have contributed to speech preparation being more difficult in Almor's study than here is that in Almor's study participants only responded to questions about the narratives. Thus, unlike in the more natural conversational task we employed here, Almor's linguistic task made it impossible for participants to plan their response while listening because they did not know what they will be asked about. A final difference between Almor's study and ours is that the narratives in Almor's study were of a highly visuospatial nature whereas here they were unconstrained and seldom included visuospatial information. Thus, prepare segments in our study reflect the brief pauses between a wide variety of exchanges in typical conversation, not just questions about specific visuospatial narratives.

Interestingly, pauses between speaking (segments of complete silence) were comparable to talking segments in terms of performance on the tracking task. This is likely because participants were planning what to say next and monitoring the appropriateness of what they were about to say. It is also possible that these breaks in speech were a result of detecting an unsuitable utterance during monitoring, and that participants were essentially pausing in order to revise the utterance.

The GCAs also revealed that the target velocity manipulation had a more interesting effect than was detected in the ANOVAs of segment means. Specifically, the GCAs revealed that target velocity exaggerated differences between conversational segments. This too is in line with our claims: when the target is moving slowly, fewer resources are needed to monitor performance. As speed increases, more and more resources are required. Thus, during Low speed conditions, participants were able to allocate sufficient resources to the tracking task, maintaining good performance of both conversation and visuomotor task. During Medium and High speed conditions, more resources are required to maintain good performance. However, resources are limited due to conversational demands, resulting in more noticeable differences between Listen and Talk conditions. This increase in resource demands, as mentioned above, leads to a subtle reduction in speech rate. More generally, the results of the GCA analyses highlight the importance of looking at the dynamics of conversation at a fine-grained temporal resolution for the understanding of the performance of dual tasks involving natural language. We can also see in our data that performance changes within segments.

One of our goals in conducting E1 was to test whether findings from Almor (2008) could be generalized to natural conversation. The conversation task used in this experiment replicated Almor's finding of a difference between Talk and Listen segments but not the difference between Listen and Prepare. The GCA allowed for a fine-grained analysis of the effects of the dynamics of conversation, further extending Almor's results.

While these results extend previous research to unconstrained conversation, differences in Talk and Listen segment durations and in the number of interruptions made by participants and the confederate raise some concern that conversation with a confederate may be different than normal conversation, or that these results could somehow reflect idiosyncrasies of conversing with our particular confederate. We thus address this issue in Experiment 2 by changing the conversation partner to friends of the participants.

EXPERIMENT 2

The only change to the design between this experiment and E1 was having participants converse with a friend they had known before the experiment, as opposed to a confederate. By changing this aspect of the design we will be able to more confidently make claims about the allocation of attention during natural conversation. We expect that our findings from E1 will effectively be replicated in Experiment 2 (E2). Using friends (a different conversation partner for each subject) will also allow us to compare our conversational fluency measurements with interlocutors not performing a secondary tracking task.

Methods

Participants—Forty (20 pairs) of University of South Carolina undergraduates participated in this study for extra credit after signing a consent form. Participants were recruited in friend dyads from the psychology participant pool. Data from 4 pairs were unusable due to equipment malfunction. Of the participants used, 26 were female and 10 were male. Dyad make up consisted of 8 all-female groups and 8 mixed-gender groups. Mean participant age was 18.97.

Procedure—Every aspect of this study was identical to that of E1, the only difference being that the conversational partner was a friend instead of the confederate.

Results

All questions on the post-experiment survey were on a 7 point Likert-type scale with 1 being not at all *x* and 7 being very *x*. The post-experiment surveys revealed that both the partners and participants felt their conversations had been fairly natural, partners' $M = 6.5$, $SD = .52$, participants' $M = 6.69$, $SD = .46$, $t(30) = 1.1$, $p = .3$, $d = .38$. The surveys also revealed that partners and participants agreed on how typical the conversations were of most conversations they would have together, partners' $M = 6.19$, $SD = .75$, participants' $M = 6.44$, $SD = .89$, $t(30) = .86$, $p = .4$, $d = .30$. Partners also reported that the participants were somewhat distracted, $M = 2.5$, $SD = 1.37$, while participants reported that the partners were not distracted during the conversation, $M = 1.13$, $SD = .34$, $t(30) = 3.91$, $p = .001$, $d = 1.38$. Partners and participants disagreed on how distracted the participants were during the conversation, partners' $M = 2.5$, $SD = 1.37$, participants' $M = 3.56$, $SD = 1.55$, $t(30) = 2.06$, $p = .05$, $d = .73$. We compared the results of the post-experiment surveys from E2 with those collected in E1 to see if the presence of the confederate in lieu of a friend had resulted in overall differences in conversation quality. Unpaired t-tests revealed that participants rated conversations with friends as more natural than conversations with a confederate, E1 $M = 5.91$, $SD = 1.14$, E2 $M = 6.69$, $SD = .48$, $t(35) = 2.58$, $p = .01$, $d = .9$. They also rated conversations with friends as more typical than conversations with a confederate, E1 $M = 5.62$, $SD = 1.36$, E2 $M = 6.44$, $SD = .89$, $t(35) = 2.09$, $p = .04$, $d = .71$. However, participants from both experiments rated themselves as being equally distracted, E1 $M = 3.71$, $SD = 1.23$, E2 $M = 3.56$, $SD = 1.55$, $t(35) = .33$, $p = .74$, $d = .11$.

In order to determine behavioral differences between participants and their partners, we compared the average number of Listen and Talk segments produced per conversation and found no significant difference, Listen $M = 225.4$, $SD = 51.59$, Talk $M = 213$, $SD = 47.2$, $t(15) = .14$, $p = .89$, $\eta_p^2 = 0$. Next we analyzed the duration of Listen and Talk segments and found no significant effect, Listen $M = 1.93s$, $SD = .47$, Talk $M = 1.96s$, $SD = .51$, $t(15) = .17$, $p = .87$, $\eta_p^2 = 0$. As before, we divided the Overlap category into instances of participants interrupting and being interrupted (using the same criteria as described in E1). We found that partners interrupted participants more often than they were interrupted by participants, partner interruptions $M = 45.38$, $SD = 18.52$, participant interruptions $M = 21.63$, $SD = 12.98$, $t(15) = 7.48$, $p < .001$, $\eta_p^2 = .79$. Thus, overall, these analyses show that participants were making approximately the same number of contributions to the conversation as the partner, but they were interrupting less often than their friends.

Next, we analyzed speech rate and disfluency production in several different ways. First, we compared the speech rates of participants and their friends, finding no significant difference, participant speech rate $M = 4.06$ (words/second), $SD = .32$, partner speech rate $M = 3.99$, $SD = .33$, $t(30) = .58$, $p = .56$, $d = .21$. Then we compared speech rate across different speeds. A repeated measures ANOVA with factor Speed (Low, Med, High) was not significant, Low $M = 4.03$, $SD = .34$, Med $M = 4.09$, $SD = .33$, High $M = 4.05$, $SD = .44$, $F(2, 30) = .21$, $p = .82$, $\eta_p^2 = .01$. Finally, we compared participant speech rate across experiments, running a 2×3 mixed repeated measures ANOVA with between-subjects factor Partner (Confederate vs. Friend) and within-subjects factor Speed (Low, Med, High). There were no main effects of Speed, $F(2, 70) = .87$, $p = .43$, $\eta_p^2 = .02$, or Partner, $F(1, 35) = .03$, $p = .85$, $\eta_p^2 = .001$, and there was no interaction, $F(2, 70) = 2.3$, $p = .11$, $\eta_p^2 = .06$. For disfluency production we looked first at how participants fared in comparison with their friends, finding an almost significant effect, participant $M = .03$ (disfluencies/word), $SD = .02$, partner $M = .02$, $SD = .01$, $t(30) = 2$, $p = .06$, $d = .7$. Next, we looked at disfluency production during Talk

segments across different speeds, running a repeated measures ANOVA with factor Speed (Low, Med, High). This was not significant, Low $M = .03$, $SD = .02$, Med $M = .04$, $SD = .04$, High $M = .04$, $SD = .03$, $F(2,30) = .65$, $p = .53$, $\eta_p^2 = .04$. Finally, we compared disfluency production across experiments, running a 2×3 mixed repeated measures ANOVA with between-subjects factor Partner (Confederate vs. Friend) and within-subjects factor Speed (Low, Med, High). There were no main effects of Speed, $F(2, 70) = 1.24$, $p = .3$, $\eta_p^2 = .03$, or Partner, $F(1, 35) = 1.15$, $p = .29$, $\eta_p^2 = .03$, and there was no interaction, $F(2, 70) = .01$, $p = .99$, $\eta_p^2 = 0$. To summarize the above analyses, the tracking task or interaction with a friend vs. the confederate did not have an effect on either speech rate or disfluency production.

In order to analyze smooth pursuit performance, we calculated the mean distance to target in each of the conversational segments and ran a 3×6 repeated measures ANOVA with factors Speed (Low, Med, High) and Conversation (Control, Listen, Overlap, Pause, Prepare, Talk) (see Figure 4). This analysis found the expected significant main effect of Speed, $F(2,30) = 332.72$, $p < .001$, $\eta_p^2 = .96$, no main effect of Conversation, $F(5,75) = .906$, $p = .48$, $\eta_p^2 = .06$, and no interaction, $F(10,150) = 1.43$, $p = .17$, $\eta_p^2 = .09$. Although there was no main effect for Conversation, we did have specific hypotheses about differences between conversational segments and thus ran planned pairwise comparisons to test these predictions. We found a reliable difference between Talk and Listen, such that participants were performing better on the visuomotor task when listening than when speaking, $t(15) = 4.63$, $p < .001$, $\eta_p^2 = .61$. This result remained significant following correction for multiple comparisons using Rom's (1990) method at $\alpha = .05$. Following the results of E1, we posited that Prepare would be better than Talk, and while this turned out to be the case, $t(15) = 2.63$, $p = .02$, $\eta_p^2 = .32$, this result was not significant following correction for multiple comparisons using Rom's method at $\alpha = .05$ and should thus be treated with caution (with 4 contrasts Rom's method corrects α to .0127). We had also hypothesized that Overlap would be worse than Talk but instead found that Talk was worse than Overlap, $t(15) = 2.51$, $p = .02$, $\eta_p^2 = .31$, although this result too was no longer significant when corrected for multiple comparisons. Finally, we had hypothesized that Pause would be no different than Talk, and this proved to be the case, $t(15) = .21$, $p = .83$, $\eta_p^2 = .01$ (see Table 2 for means).

To make sure our failure to replicate Almor's (2008) main effect of Conversation is not due to insufficient power, we used the *MSes* reported by Almor to perform a power analysis using G*Power (Buchner et al., 2012). This analysis indicated that 12 participants are required in order to achieve .80 power ($\alpha = .05$) in our experiment to detect the main effect of conversation observed by Almor ($\eta_p^2 = .06$). Just as in E1, the much higher power in our study than in Almor's reflects the higher number of observations per participant in our study ($M = 222.19$, $SD = 39.96$), than in Almor's (exactly 47 observations per participant). Therefore, the lack of a main effect of Conversation in this experiment is not a result of low power. Again, because Almor did not report effect sizes for individual comparisons, we could not perform power analyses for individual contrasts.

For the data in E2 we also ran two sets of GCAs. Our first set of analyses (Analysis 2a) was identical to Analysis 1a described above. Table 5a shows the results of model comparison according to the AIC weight of evidence criterion (Long, 2012). As can be seen in the table, the quadratic model was the most plausible as it had probability close to .98 among the compared models. Table 5b shows the results of model comparison according to log likelihood criteria, which also found that the quadratic model provided a better fit than the linear model, but none of the higher order models improved this fit. Therefore, the full quadratic model was chosen as the best fitting model, indicating that all the predictors interacted to affect the change and rate of change in distance to target. Figure 5 shows the model fitted to the observed data. The full parameter estimates of the model can be found in

Online Supplement 3. As is indicated by the inclusion of the interaction effects in the chosen model, performance changed across different conditions. As is apparent in the graph, there was little difference between the Listen and Talk conditions at the beginning and end of segments of conversation when the target was moving at Low or Medium speeds. In the High target speed condition, however, there were differences such that at segment beginning, performance improved in Listen segments but deteriorated in Talk segments, and at segment ends this pattern reversed. The presence of the quadratic interaction in the chosen model appears to reflect the nonlinear change in performance at the end of Listen segments in both the Medium and High speed conditions.

Our second set of analyses (Analysis 2b) was the same as Analysis 1b described above. Table 6a shows the results of model comparison according to the AIC weight of evidence criterion (Long, 2012). As can be seen in the table, the cubic model was the most plausible as it had a probability of .55 among the compared models. Although in this analysis the cubic model was identified as the most probable, the quartic model was also determined to be somewhat probable. This was confirmed by the maximum likelihood model comparison (Table 6b) in which the cubic model was found to fit the data significantly better than the quadratic model and the quartic model marginally significantly better than the cubic model. Because the quartic model did not add any new information, we chose the cubic model for these data. Figure 6 shows the model fitted to the observed data. The model coefficients are shown in Online Supplement 4. As shown by model and graph, at the end of segments, during Low and Med speeds Prepare, Listen, and Talk segments pattern very similarly, with the exception of a bump in detriment at the end of Listen segments during Med speeds, possibly an effect of anticipating the end of the conversation partner's turn. During High speeds, Prepare patterns similarly to Listen.

Discussion

The conversations in this experiment clearly differed from those in E1. Our between-experiment analyses of post-experiment surveys showed that there were overall differences in conversational quality. Participants understandably rated conversations with the confederate as less natural and typical than conversations with a friend. Unlike in E1, in this experiment there were no differences between the duration of Listen and Talk segments, indicating that the difference found in E1 could have reflected the confederate's conversational style.

Other aspects of the conversations were more comparable to E1. Importantly, there were no differences in self-rated distraction between participants of our two experiments, indicating that differences in results may be attributable to the naturalness and typicality of these conversations but not overall greater distraction. Also, the number of Talk and Listen segments was comparable across the two experiments. The similarity of interruption patterns across both experiments may indicate that, in general, participants, as a result of dual-tasking, were taking a more passive role in the conversation. Although we found that in this experiment participants were producing more disfluencies than partners, this was only a borderline effect and, as such, should be treated with caution. In general, speech rate and disfluency production were identical to those of participants in E1.

In this experiment we were also able to compare ratings of the participants and their partners. We found that participants and partners largely agreed with one another on the naturalness and typicality of their conversations. They only differed in their agreement of how distracted the participant was, with participants on average rating themselves as being more distracted than how their partners rated them.

The effect conversation had on the visuomotor task in this experiment was more subtle than in E1. The lack of a main effect of Conversation is compatible with previous research (Horrey & Wickens, 2006), specifically in that conversation has little impact on a concurrent tracking task. However, our planned pairwise comparisons revealed that Talk was more detrimental than Listen, and Overlap, Pause, and Prepare were as detrimental to the concurrent task as Talk. Thus, although the effect appears to be somewhat smaller than what was found in E1, the main result that talking is worse than listening remains a reliable finding.

GCAs revealed similar patterns to those found in E1 except that differences between conversational segments were not apparent during Low and Med speeds but rather only became clear in the High speed condition, which showed a pattern very similar to what was found in E1. This may be a result of participants requiring fewer resources for speech planning and monitoring for the conversations in this experiment. Interestingly, Prepare patterned more closely to Listen than Talk segments in this experiment, whereas in E1 Prepare seemed to have behaved more similarly to Talk than Listen.

GENERAL DISCUSSION

By having participants perform a tracking task while conversing with a friend or a confederate, we were able to record fine-grained, dynamic changes in attention demands during conversation. The two experiments we reported have shown that the attentional demands associated with natural conversation change rapidly. Specifically, we observed that during talking segments, tracking performance declined at the beginning and improved at the end and that during listening segments, performance improved at the beginning and declined at the end. To the best of our knowledge, this is the first demonstration of the dynamical changes in attentional demands during different stages of conversation at a millisecond time scale.

The specific pattern of these dynamical changes indicates that speech planning is the most resource-demanding aspect of conversation. This is further supported by the similarity between tracking task performance during Pause and Talk segments, and by the gradual decline in tracking performance that occurs during Prepare segments. These show that the detrimental effects of talking actually occur even before and without any acoustic instantiation of speech. Indeed, our findings indicate that the attentional detriment associated with speech planning is spread across talking, preparing, and listening segments. This has some important implications for theories of language comprehension and production, which we will review in more detail below.

We also found that increasing the difficulty of the tracking task exaggerated the interference from conversation. Because increasing tracking task difficulty increases monitoring requirements, this finding indicates that the cognitive demands of speech monitoring overlap with those of speech planning, and therefore that the two processes tap shared resources. This provides useful constraints for applying theories of dual-task performance (e.g., Wickens, 2008) to multitasking during conversation. This is also reviewed in more detail below.

Finally, we believe that the approach of recording continuous measures of secondary task performance will be a valuable tool for other researchers interested in studying how attention shifts across multitasking scenarios. Our results indicate that this approach is most profitably used in conjunction with an advanced statistical analysis of dynamical change, such as the GCAs that we employed in this work.

We note that although one of our findings, that talking is more attention demanding than listening, has been reported before, previous research has only shown this for non-naturalistic linguistic tasks that either involve no true linguistic communication (Kunat et al., 2008) or at best, involve an interaction with a pre-recorded script (Almor, 2008). Indeed, our fine-grained temporal analyses allowed us to explain *why* talking is more demanding than listening. Specifically, we claim that talking involves simultaneous planning and monitoring, both of which require considerable attentional resources.

We now turn to discuss in detail the theoretical implications of our findings for theories of language comprehension and production, and for dual-task research.

Language Comprehension and Production during Dual-Task Scenarios

The vast differences we have shown between tracking performance during listening and talking have important implications for models of language comprehension and production. For comprehension, our results support theories that argue for flexibility in the effort allocated to comprehension. One example is the *good enough processing* theory of sentence comprehension (Christianson et al., 2006; Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007), which states that comprehenders process linguistic information only to the extent necessary for deriving a possible interpretation. According to this approach, full syntactic parsing of the input is effortful and only occurs in special circumstances, such as when the derived interpretation is deemed wrong due to inconsistency with preceding context or upon receiving successive input. Another example is the minimalist approach to discourse processing proposed by McKoon and colleagues (McKoon & Ratcliff, 1990; McKoon & Ratcliff, 1989a; McKoon & Ratcliff, 1989b). According to this approach, discourse relations such as inferences and co-reference are not always computed by comprehenders. Whether or not such relations are processed depends on the comprehender's goals (Greene, McKoon & Ratcliff, 1992) and level of engagement (Love & McKoon, 2011) as well as on processing difficulty. The main tenet of both theories is that the depth of processing during comprehension varies as a function of the effort required for processing, the available resources, and the goals of the comprehender. While listening, when resources are required for a secondary task, for example in the difficult conditions of target tracking, interlocutors can allocate resources to the secondary task. This was indicated in participants' tracking performance at the beginning and end of listening segments. Note that according to these theories, fewer resources do not entail reduced comprehension but simply that successful comprehension can be achieved with fewer resources and that the incentive to divert resources away from comprehension varies according to the factors we discussed.

In contrast to comprehension, production is much more effortful, requiring precise planning, lexical access, syntactic formulation, motor execution, and monitoring at all levels (Postma, 2000). Even the monitoring aspect of production (which could be argued to be a similar mechanism as that employed during comprehension) appears to consume more resources during production than comprehension, likely because every stage of linguistic production requires the use of the monitor or separate monitors, whereas a conversational partner's linguistic input may not be scrutinized with such rigor. Other authors have similarly argued (as we have) that the planning and monitoring stages of speech production pose the greatest demands on attention (e.g., Blackmer & Mitton, 1991; Ferreira & Pashler, 2002; Oomen & Postma, 2002). Our data show the progress from only speech planning, to the more demanding combination of ongoing speech and planning, and finally to the most attentionally demanding mixture of speech, planning, and monitoring. As more processes are combined, the demands on attention increase, leading to worse performance on the concurrent tracking task. One implication of this interpretation is that speech planning itself is not what leads to worse performance but that the detriment is a result of planning

overlapping with other processes. This is mirrored in our results which showed that the difference between listening and speaking varied as a function of the difficulty of the visual task. When the visual task was more demanding, more attention was gradually directed towards the visual task during listening and the opposite was true during speaking.

There were some intriguing differences between our two experiments. The main difference is that generally E2 appeared to be easier: participants from E2 said the conversations were more natural and typical than the participants from E1; there was a main effect of Conversation on the tracking task in E1 but not in E2; and GCA graphs show that the effect of talking was more exaggerated in E1 than in E2. However, we must tread carefully when interpreting these apparent differences. For one thing, E1 only used one confederate. Thus, the results may only be applied to situations in which someone is talking with a lab confederate or possibly even just the individual we used in our experiment. E2, on the other hand, shows reliable differences in tracking task performance when participants are conversing with different people: their friends. Previous research (Savitsky et al., 2011) has shown that friends are not as vigilant at tracking one another's common ground (Clark & Brennan, 1991) as they are when speaking with a stranger. For example, participants tended to make more egocentric mistakes when following directions given by a friend than when following directions from a stranger. Savitsky et al. interpreted their findings as demonstrating an overestimation of the success of communicating with a close friend, something they term the *closeness-communication bias*. However, another possible interpretation is that intimates feel less impetus to invest the cognitive resources necessary for audience design when communicating with a friend because they do not attach high cost to temporary communication failures and corrections. This means that interlocutors may choose to invest more or less resources in the communication depending on their level of intimacy or perhaps on their overall level of comfort. While Savitsky's study shows that this bias occurs in discourse-level comprehension, we expect it would easily translate into production as well. If participants, in speaking with friends, do not allocate as many resources to monitoring their speech (and discourse content especially) as they do when speaking with strangers, we would expect conversations with friends to be less detrimental to a concurrent task than conversations with strangers (or in our case conversations with a confederate). If indeed this *closeness-communication bias* did manifest during speech production, we might expect it to look similar to the observed differences between Prepare and Talk with Listen that we saw across our experiments. However, because the design of our study does not permit us to directly test these claims, disentangling these issues remains open for future research.

Dual-Task Attention Model

Wickens's (2008) model of dual-task attention provides a useful framework for interpreting the results from our experiment. According to that model, detriments to the concurrent visuomotor task should be most apparent when the two tasks match along three possible dimensions: stage of processing, code of processing, and modality. This is indeed the case in the present study, as can be seen in the differences between talking and listening. We had earlier argued that talking involves planning, response, and monitoring, while listening involves monitoring (and this only minimally). Both planning and monitoring increase attentional load along the cognitive processing stage, while the mechanical aspects of speech tax the response processing stage. In our experiments, we observed increasing interference during planning and then a continued increase in interference with the addition of speaking and monitoring. Listening, on the other hand, only makes use of the cognitive processing stage. In other words, listening, which overlapped the least with the tracking task in terms of Wickens's model's processing dimensions, interfered with tracking the least.

Wickens's (2008) model not only allows us to describe our findings in a way that is compatible with the literature on dual-task performance, but also helps highlight some important implications of our study for a topical problem in society: that of simultaneously driving and talking on cell phones. Consider what the tracking task results indicate. Differences in distance to target indicate poor monitoring and response capability. This could translate into serious (possibly fatal) consequences in an actual driving scenario, which also requires constant monitoring and response modulation, as well as the ability to detect and respond to the sudden appearance of obstacles on the road. Our results suggest that addressing this issue would require identifying the specific reasons and time course of attentional overlap. For example, we think that driving while talking could be made safer by automatically reducing the threshold of various sensor alerts (e.g., going out of lane, distance to other cars, etc.) when the driver is speaking or is likely to start speaking.

The same logic and dual-task analysis that we have employed here, could also be applied in dual-task scenarios involving two nonlinguistic tasks, for example using a combination of visuomotor tasks such as tracking and signal detection, a scenario that represents other types of likely distractions (aside from conversation) encountered when driving a car (e.g., Wickens & Liu, 1988). Using GCA to model tracking behavior as a go/no-go signal appears (resembling signs and other possible road events) could show the fine-grained effects of nonlinguistic monitoring and response planning. To the best of our knowledge a fine-grain temporal analysis of the influence of dual-task performance has not yet been performed. Given the analytical tools available and the nature of attention's constant flux, the goal of dual-task research should be to model these fluctuations over time, and not describe them discretely.

Conclusions

Our two experiments show that studies using dual-task designs need to pay close attention to the choice, assessment and analysis of the tasks. We have four main points regarding these issues. First, designs using natural conversation, despite requiring a large amount of coding work, are preferable over ones that rely on unnatural tasks (e.g., word games, listening to narratives) because the former more closely mimics real-life dual-task scenarios, while the latter may produce results that are difficult to relate to real-world phenomena. Second, when dealing with natural conversation, it is important to respect the variability of attentional demands in a conversation by dissecting the data in a proper, theoretically driven manner. Studies that ignore this variability will not be able to detect effects with any reliability. We have suggested one such coding scheme in this paper and hope that other investigators will use and further develop this scheme. Third, although this was only suggested by our data, a factor as simple as the identity of the interlocutor, the degree of intimacy between interlocutors, or the naturalness of the conversation may have a major impact on the concurrent visuomotor task. Our fourth and final point is that we believe that the use of GCAs to analyze dual-task data involving continuous measures is an exciting new avenue for scientific investigation. This has recently been used in order to analyze eye-tracking data (Mirman, Dixon, & Magnusson, 2008; Peters, Boiteau, & Almor, in submission), but it is similarly appropriate for tracking tasks and other dual-task data. Previous studies have fallen short in being able to precisely describe how humans manage dual-task scenarios online. This new method of analysis allows us to see a much clearer picture of real-time attentional allocation and shows the benefits of properly analyzing rich datasets. These analyses also revealed the fluctuating nature of attention during conversation, both across and within conversational segments. Simple comparisons of averages are insufficient to capture dual-task performance with any degree of accuracy and may lead to false null results. Even when there were differences (as was the case with speaking and listening segments), the exact nature of these differences remained unclear following the comparisons of averages. Indeed,

only through using GCA were we able to uncover the effect of speech planning on concurrent performance of the visuomotor task.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was partially funded by NIH R21AG030445 and NSF BCS0822617 grants. The authors wish to thank Kristen Sloan and other aLab members for their help with data collection and coding.

References

- Alm H, Nilsson L. Changes in driver behavior as a function of hands-free mobile phones: A simulator study. *Accident Analysis & Prevention*. 1994; 26:441–451. [PubMed: 7916852]
- Almor A. Why does language interfere with vision-based tasks? *Experimental Psychology*. 2008; 55:260–268. [PubMed: 18683623]
- Arnold JE, Kahn JM, Pancani GC. Audience design affects acoustic reduction via production facilitation. *Psychonomics Bulletin & Review*. 2012; 19:505–512.
- Baddeley, AD. *Working memory*. Oxford: Oxford University Press; 1986.
- Baddeley AD, Logie R, Bressi S, Della Sala S, Sinnler H. Dementia and working memory. *Quarterly Journal of Experimental Psychology*. 1986; 38A:603–618. [PubMed: 3809575]
- Bates D, Maechler M, Bolker B. lme4: Linear mixed-effects models using Eigen and syntax. 2011 R package version 0.999375-39.
- Becic E, Dell GS, Bock K, Garnsey SM, Kubose T, Kramer AF. Driving impairs talking. *Psychonomic Bulletin & Review*. 2010; 17(1):15–21. [PubMed: 20081155]
- Blackmer ER, Mitton JL. Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*. 1991; 39:173–194. [PubMed: 1841032]
- Bock, K. Sentence production: From mind to mouth. In: Miller, JL.; Eimas, PD., editors. *Handbook of perception and cognition*. Vol. 11. Orlando, FL: Academic Press; 1995.
- Bortfield H, Leon SD, Bloom JE, Schober MF, Brennan SE. Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*. 2001; 44(2):123–147. [PubMed: 11575901]
- Buchner, A.; Erdfelder, E.; Faul, F.; Lang, A. G*Power (Version 3.1.5)[Computer program]. 2012. <http://www.psych.uni-duesseldorf.de/aap/projects/gpower/>
- Caird JK, Willness CR, Steel P, Scialfa C. A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention*. 2008; 40(4):1282–1293. [PubMed: 18606257]
- Chafe, WL. Linguistic differences produced by differences between speaking and writing. In: Olson, DR.; Torrance, N.; Hildyard, A., editors. *Literacy, language, and learning: The nature and consequences of reading and writing*. Cambridge, UK: Cambridge University Press; 1985.
- Chafe, WL. Cognitive constraints on information flow. In: Tomlin, R., editor. *Coherence and Grounding in Discourse*. Amsterdam: John Benjamins; 1987.
- Christianson K, Williams CC, Zacks RT, Ferreira F. Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes*. 2006; 42(2):205–238. [PubMed: 17203135]
- Clark HH. Managing problems in speaking. *Speech Communication*. 1994; 15:243–250.
- Clark, HH. *Using language*. Cambridge, UK: Cambridge University Press; 1996.
- Clark, HH.; Brennan, SA. Grounding in communication. In: Resnick, LB.; Levine, JM.; Teasley, SD., editors. *Perspectives on socially shared cognition*. Washington: APA Books; 1991.
- D'Esposito M, Detre JA, Alsop DC, Shin RK, Atlas S, Grossman M. The neural basis of the central executive system of working memory. *Nature*. 1995; 378:279–81. [PubMed: 7477346]

- Drews FA, Pasupathi M, Strayer DL. Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology: Applied*. 2008; 14:392–400. [PubMed: 19102621]
- Ferreira F, Bailey KGD, Ferraro V. Good-enough representations in language comprehension. *Current Directions in Psychological Science*. 2002; 11:11–15.
- Ferreira F, Christianson K, Hollingworth A. Misrepresentations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research*. 2001; 30(1):3–20. [PubMed: 11291182]
- Ferreira F, Patson ND. The ‘good enough’ approach to language comprehension. *Language and Linguistics Compass*. 2007; 1(1–2):71–83.
- Ferreira VS, Pashler H. Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2002; 28:1187–1199.
- Forster MR, Sober E. How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *The British Journal for the Philosophy of Science*. 1994; 45:1–35.
- Forster, MR.; Sober, E. AIC scores as evidence—a Bayesian interpretation. In: Forster, MR.; Bandyopadhyay, P., editors. *The philosophy of statistics*. New York: Kluwer; 2011. p. 1-25.
- Frazier L, Rayner K. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*. 1982; 14:178–210.
- Gennari S, MacDonald MC. Linking production and comprehension processes: The case of relative clauses. *Cognition*. 2009; 111(1):1–23. [PubMed: 19215912]
- Gibson E. Linguistic complexity: Locality of syntactic dependencies. *Cognition*. 1998; 68:1–76. [PubMed: 9775516]
- Gibson E, Pearlmutter N. Constraints on sentence comprehension. *Trends in Cognitive Science*. 1998; 2:262–268.
- Greene SB, McKoon G, Ratcliff R. Pronoun resolution and discourse models. *Language*. 1992; 67:439–474.
- Horrey WJ, Wickens CD. Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors*. 2006; 48:196–205. [PubMed: 16696268]
- Horton WS, Keysar B. When do speakers take into account common ground? *Cognition*. 1996; 59:91–117. [PubMed: 8857472]
- Insurance Institute for Highway Safety, Highway Loss Data Institute. Cell phone laws. 2012. Retrieved on November 11, 2012 from <http://www.iihs.org/laws/cellphonelaws.aspx>
- Israel JB, Chesney GL, Wickens CD, Donchin E. P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology*. 1980; 17:57–70.
- Just MA, Carpenter PA, Keller TA, Emery L, Zajac H, Thulborn KR. Interdependence of nonoverlapping cortical systems in dual cognitive tasks. *NeuroImage*. 2001; 14:417–426. [PubMed: 11467915]
- Just MA, Keller TA, Cynkar J. A decrease in brain activation associated with driving when listening to someone speak. *Brain Research*. 2008; 1205:70–80. [PubMed: 18353285]
- Kunar MA, Carter R, Cohen M, Horowitz TS. Telephone conversation impairs sustained visual attention via a central bottleneck. *Psychonomic Bulletin & Review*. 2008; 15:1135–1140. [PubMed: 19001580]
- Kubose TT, Bock K, Dell GS, Garnsey SM, Kramer AF, Mayhugh J. The effects of speech production and speech comprehension on simulated driving performance. *Applied Cognitive Psychology*. 2006; 20:43–63.
- Lavie N. Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*. 2005; 9:75–82. [PubMed: 15668100]
- Levelt WJM. Monitoring and self-repair in speech. *Cognition*. 1983; 14:41–104. [PubMed: 6685011]
- Levy J, Pashler H, Boer E. Central interference in driving: Is there any stopping the psychological refractory period? *Psychological Science*. 2006; 17:228–235. [PubMed: 16507063]
- Logie RH, Cocchini G, Della Sala S, Baddeley AD. Is there a specific executive capacity for dual-task coordination? Evidence from Alzheimer's disease. *Neuropsychology*. 2004; 18:504–13. [PubMed: 15291728]

- Long, JD. Longitudinal data analysis for the behavioral sciences using R. Los Angeles: Sage Publications, Inc; 2012.
- Love J, McKoon G. Rules of engagement: Incomplete and complete pronoun resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2011; 37:874–887.
- MacDonald, MC. Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In: MacWhinney, B., editor. *The emergence of language*. Mahwah NJ: Erlbaum; 1999.
- MacDonald MC, Pearlmutter NJ, Seidenberg MS. Lexical nature of syntactic ambiguity resolution. *Psychological Review*. 1994; 101(4):676–703. [PubMed: 7984711]
- McKoon G, Ratcliff R. Inferences about contextually defined categories. *Journal of Experimental Psychology: Language, Memory, and Cognition*. 1989a; 15(6):1134–1146.
- McKoon G, Ratcliff R. Semantic associations and elaborative inference. *Journal of Experimental Psychology: Language, Memory, and Cognition*. 1989b; 15(2):326–338.
- McKoon G, Ratcliff R. Dimensions of inference. *The Psychology of Learning and Motivation*. 1990; 25:313–328.
- Mirman D, Dixon JA, Magnuson JS. Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*. 2008; 59:475–494. [PubMed: 19060958]
- Oomen CCE, Postma A. Limitations in processing resources and speech monitoring. *Language and Cognitive Processes*. 2002; 17(2):163–184.
- Peters SA, Boiteau TW, Almor A. Pronouns, definite references, and semantic relations in spoken language comprehension. in submission.
- Postma A. Detection of errors during speech production: A review of speech monitoring models. *Cognition*. 2000; 77:97–131. [PubMed: 10986364]
- Rakauskas ME, Gugerty LJ, Ward NJ. Effects of naturalistic cell phone conversations on driving performance. *Journal of Safety Research*. 2004; 35:453–464. [PubMed: 15474548]
- Ramanarayanan V, Bresch E, Byrd D, Goldstein L, Narayanan SS. Analysis of pausing behavior in spontaneous speech using real-time magnetic resonance imaging of articulation. *The Journal of the Acoustical Society of America*. 2009; 126(5):160–165.
- Rom DM. A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika*. 1990; 77(3):663–665.
- Sacks H, Schegloff EA, Jefferson G. A simplest systematics for the organization of turn-taking for conversation. *Language*. 1974; 50(4):696–735.
- Salthouse TA, Fristoe NM, Lineweaver TT, Coon VE. Aging of attention: Does the ability to divide decline? *Memory and Cognition*. 1995; 23:59–71. [PubMed: 7885266]
- Savitsky K, Keysar B, Epley N, Carter T, Swanson A. The closeness-communication bias: Increased egocentrism among friends versus strangers. *Journal of Experimental Social Psychology*. 2011; 47:269–273.
- Schegloff EA. Overlapping talk and the organization of turn-taking for conversation. *Language in Society*. 2000; 29:1–63.
- Shallice T, McLeod P, Lewis K. Isolating cognitive modules with the dual-task paradigm: Are speech perception and production modules separate? *Quarterly Journal of Experimental Psychology*. 1985; 37:507–532. [PubMed: 4081099]
- Strayer DL, Drews FA, Johnston WA. Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied*. 2003; 9:23–32. [PubMed: 12710835]
- Strayer DL, Johnston WA. Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science*. 2001; 12:4462–4466.
- Trueswell JC, Tanenhaus MK, Kello C. Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1993; 19(3):528–553.
- Wickens, CD. Processing resources in attention. In: Nickerson, R., editor. *Human Skills*. 2. New York: Wiley and Sons; 1984. p. 71-105.

- Wickens CD. Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*. 2002; 3(2):159–177.
- Wickens CD. Multiple resources and mental workload. *Human Factors*. 2008; 50(3):449–455. [PubMed: 18689052]
- Wickens CD, Liu Y. Codes and modalities in multiple resources: A success and a qualification. *Human Factors*. 1988; 30:599–616. [PubMed: 3220489]

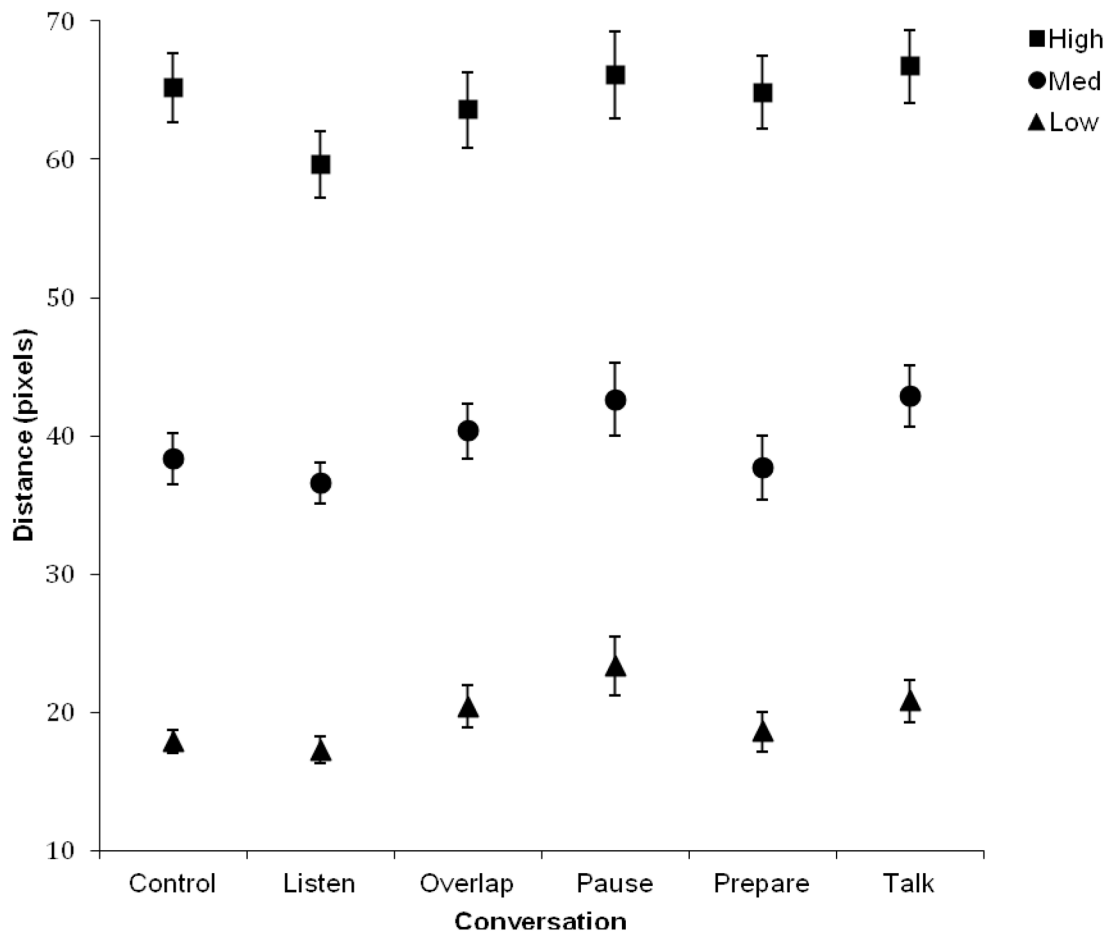


Figure 1. Distance to target during Control, Listen, Overlap, Pause, Prepare, and Talk segments across Low, Med, and High target velocity conditions when participants were paired with a confederate (E1). Symbols show observed means. Error bars show standard error of the mean.

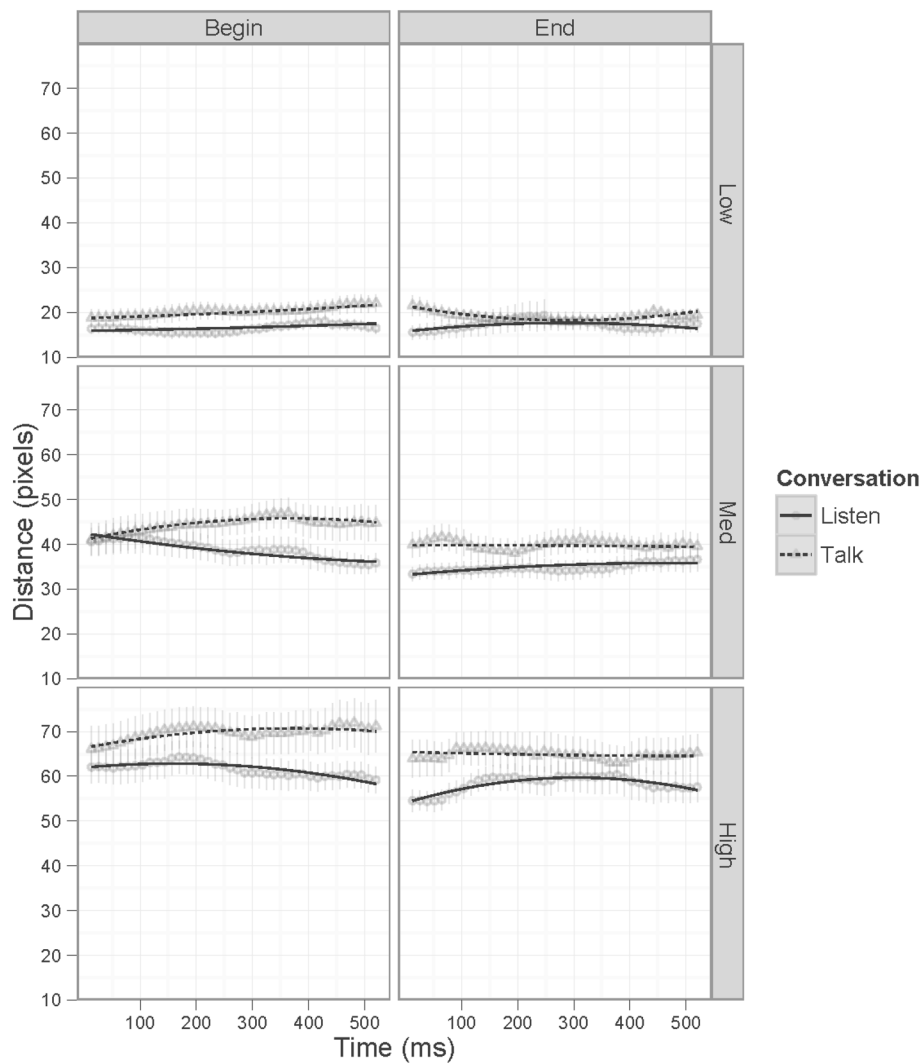


Figure 2. Distance to target during conversations with a confederate (E1) at the beginning and end of Listen and Talk segments in Low, Medium and High target velocity conditions. Symbols show observed means, error bars show standard error of the mean, and lines show the predicted values of the quadratic model.

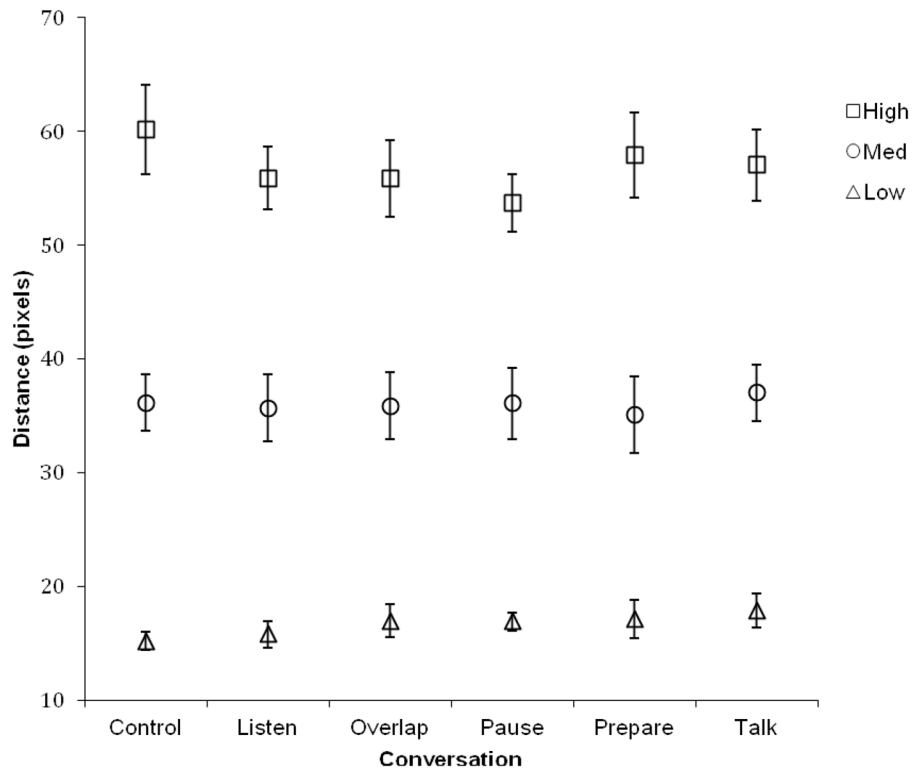


Figure 3. Distance to target during conversations with a confederate (E1) at the end of Listen, Talk and Prepare segments in Low, Medium and High target velocity conditions. Symbols show observed means, error bars show standard error of the mean, and lines show the predicted values of the quadratic model.

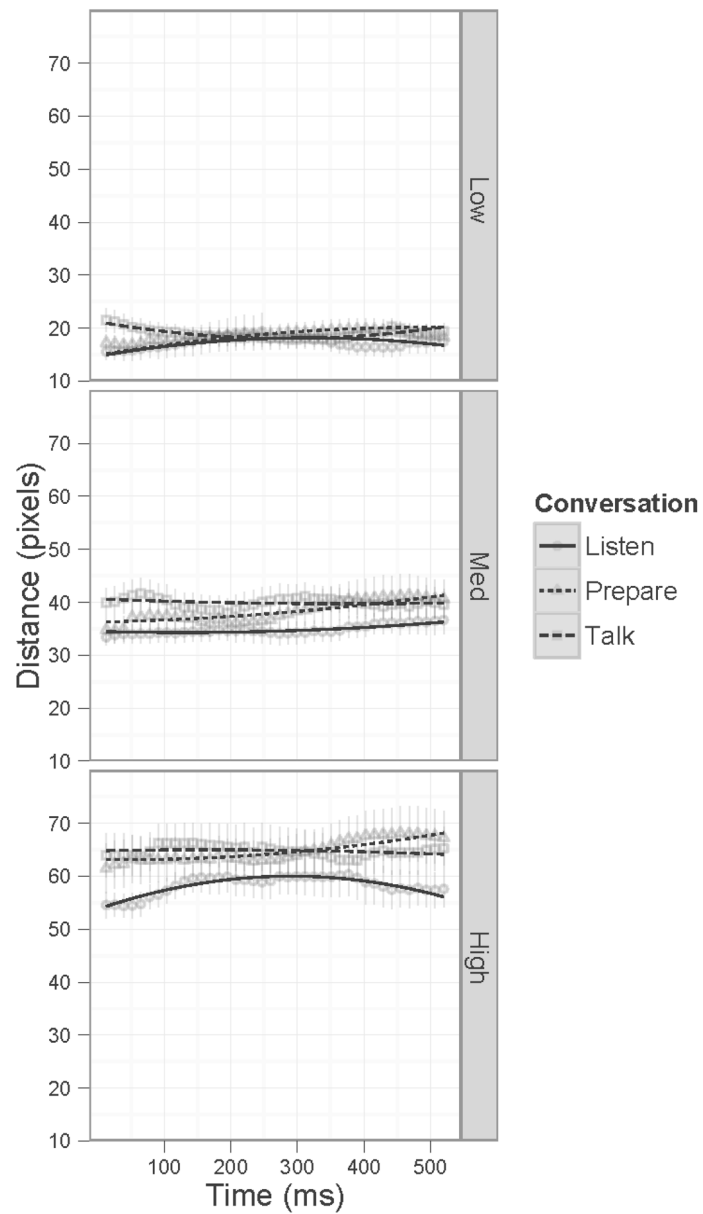


Figure 4. Distance to target during Control, Listen, Overlap, Pause, Prepare, and Talk segments across Low, Med, and High target velocity conditions when participants were paired with a friend (E2). Symbols show observed means. Error bars show standard error of the mean.

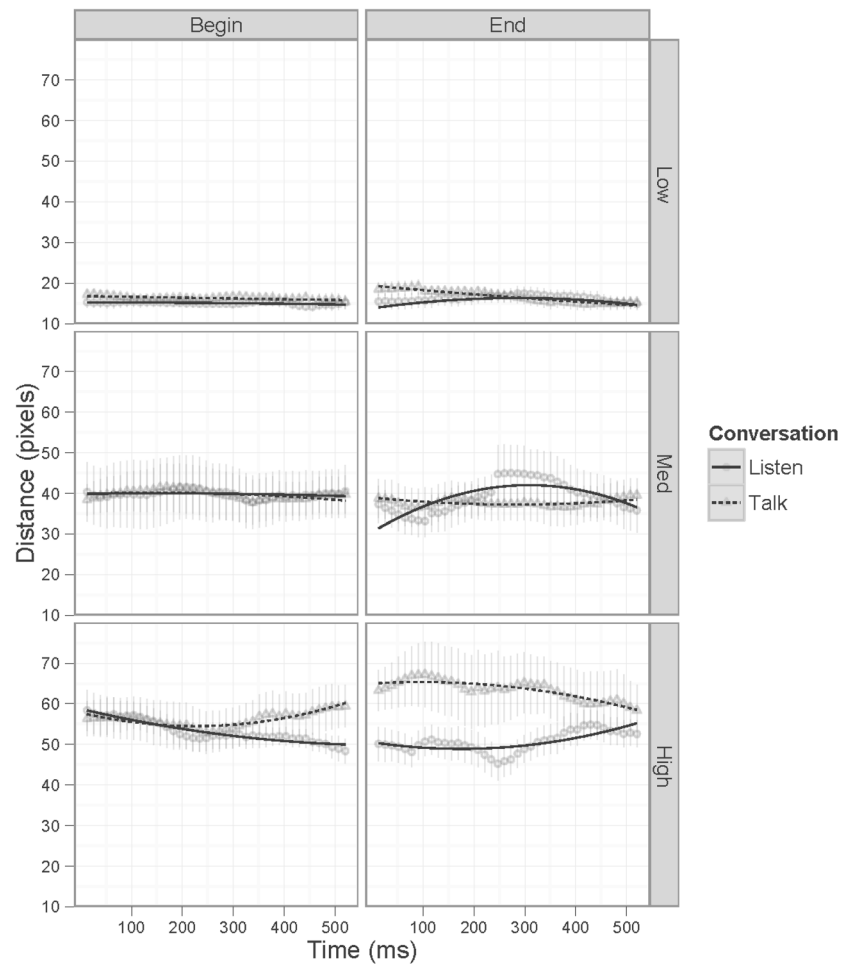


Figure 5. Distance to target during conversations with a friend (E2) at the beginning and end of Listen and Talk segments in Low, Medium and High target velocity conditions. Symbols show observed means, error bars show standard error of the mean, and lines show the predicted values of the quadratic model.

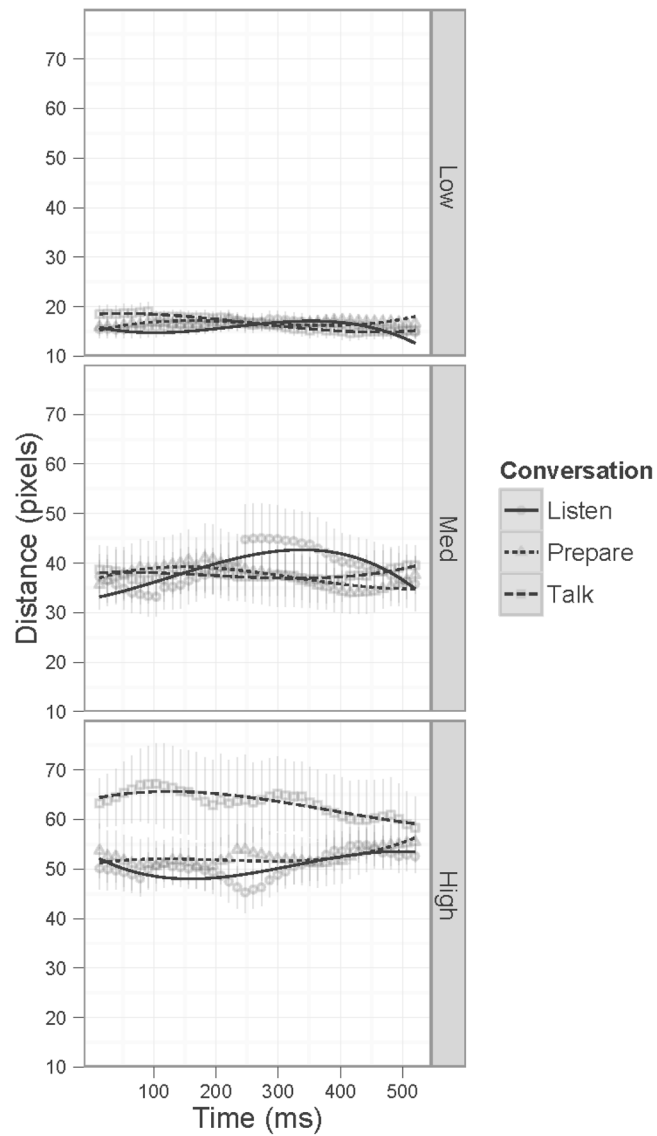


Figure 6. Distance to target during conversations with a friend (E2) at the end of Listen, Talk and Prepare segments in Low, Medium and High target velocity conditions. Symbols show observed means, error bars show standard error of the mean, and lines show the predicted values of the cubic model.

Table 1

Growth curve models used for fitting distance to target for subject i at time point j ($j = 1,40$). All the models we tested included in the first level fixed effects of intercept, linear time (Time) reflecting slope, quadratic time (Time²) reflecting acceleration (rise & fall rate), cubic time (Time³) and quartic Time (Time⁴). The second level equations were used to estimate the effect of condition on the intercept (β_{0i}) and on the time course at the different orders (β_{1i} and β_{2i}). For simplicity, the models presented in the table show one coefficient for each interaction although the models we fitted included individual terms for all the lower level interactions and main effects as well (i.e., each of the γ_{X1} coefficients in the table is in fact a matrix product of a coefficient matrix and a variable selection matrix). Our models always included a random effect of participants on the intercept, slope, and acceleration (ζ_{0i} , ζ_{1i} , and ζ_{2i}), thus allowing both the estimated baseline distance and linear and quadratic rates of change in distance to vary across individuals. Correlations between the random effects were modeled as well.

Model	
[1] Base	$Y_{ij} = \beta_{0i} + \beta_{1i} * \text{Time}_{ij} + \beta_{2i} * \text{Time}_{ij}^2 + \beta_{3i} * \text{Time}_{ij}^3 + \beta_{4i} * \text{Time}_{ij}^4 + \varepsilon_{ij}$ $\beta_{0i} = \gamma_{00} + \zeta_{0i}$ $\beta_{1i} = \gamma_{10} + \zeta_{1i}$ $\beta_{2i} = \gamma_{20} + \zeta_{2i}$ $\beta_{3i} = \gamma_{30}$ $\beta_{4i} = \gamma_{40}$
[2] Intercept	$Y_{ij} = \beta_{0i} + \beta_{1i} * \text{Time}_{ij} + \beta_{2i} * \text{Time}_{ij}^2 + \beta_{3i} * \text{Time}_{ij}^3 + \beta_{4i} * \text{Time}_{ij}^4 + \varepsilon_{ij}$ $\beta_{0i} = \gamma_{00} + \gamma_{01} * \text{Segment-type} * \text{Target-speed} * \text{Conversation} + \zeta_{0i}$ $\beta_{1i} = \gamma_{10} + \zeta_{1i}$ $\beta_{2i} = \gamma_{20} + \zeta_{2i}$ $\beta_{3i} = \gamma_{30}$ $\beta_{4i} = \gamma_{40}$
[3] Linear	$Y_{ij} = \beta_{0i} + \beta_{1i} * \text{Time}_{ij} + \beta_{2i} * \text{Time}_{ij}^2 + \beta_{3i} * \text{Time}_{ij}^3 + \beta_{4i} * \text{Time}_{ij}^4 + \varepsilon_{ij}$ $\beta_{0i} = \gamma_{00} + \gamma_{01} * \text{Segment-type} * \text{Target-speed} * \text{Conversation} + \zeta_{0i}$ $\beta_{1i} = \gamma_{10} + \gamma_{11} * \text{Segment-type} * \text{Target-speed} * \text{Conversation} + \zeta_{1i}$ $\beta_{2i} = \gamma_{20} + \zeta_{2i}$ $\beta_{3i} = \gamma_{30}$ $\beta_{4i} = \gamma_{40}$
[4] Quadratic	$Y_{ij} = \beta_{0i} + \beta_{1i} * \text{Time}_{ij} + \beta_{2i} * \text{Time}_{ij}^2 + \beta_{3i} * \text{Time}_{ij}^3 + \beta_{4i} * \text{Time}_{ij}^4 + \varepsilon_{ij}$ $\beta_{0i} = \gamma_{00} + \gamma_{01} * \text{Segment-type} * \text{Target-speed} * \text{Conversation} + \zeta_{0i}$ $\beta_{1i} = \gamma_{10} + \gamma_{11} * \text{Segment-type} * \text{Target-speed} * \text{Conversation} + \zeta_{1i}$ $\beta_{2i} = \gamma_{20} + \gamma_{21} * \text{Segment-type} * \text{Target-speed} * \text{Conversation} + \zeta_{2i}$ $\beta_{3i} = \gamma_{30}$ $\beta_{4i} = \gamma_{40}$
[5] Cubic	$Y_{ij} = \beta_{0i} + \beta_{1i} * \text{Time}_{ij} + \beta_{2i} * \text{Time}_{ij}^2 + \beta_{3i} * \text{Time}_{ij}^3 + \beta_{4i} * \text{Time}_{ij}^4 + \varepsilon_{ij}$ $\beta_{0i} = \gamma_{00} + \gamma_{01} * \text{Segment-type} * \text{Target-speed} * \text{Conversation} + \zeta_{0i}$ $\beta_{1i} = \gamma_{10} + \gamma_{11} * \text{Segment-type} * \text{Target-speed} * \text{Conversation} + \zeta_{1i}$ $\beta_{2i} = \gamma_{20} + \gamma_{21} * \text{Segment-type} * \text{Target-speed} * \text{Conversation} + \zeta_{2i}$ $\beta_{3i} = \gamma_{30} + \gamma_{31} * \text{Segment-type} * \text{Target-speed} * \text{Conversation}$ $\beta_{4i} = \gamma_{40}$

Model

[6] Quartic

$$Y_{ij} = \beta_{0i} + \beta_{1i} * \text{Time}_{ij} + \beta_{2i} * \text{Time}_{ij}^2 + \beta_{3i} * \text{Time}_{ij}^3 + \beta_{4i} * \text{Time}_{ij}^4 + \varepsilon_{ij}$$

$$\beta_{0i} = \gamma_{00} + \gamma_{01} * \text{Segment-type} * \text{Target-speed} * \text{Conversation} + \zeta_{0i}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11} * \text{Segment-type} * \text{Target-speed} * \text{Conversation} + \zeta_{1i}$$

$$\beta_{2i} = \gamma_{20} + \gamma_{21} * \text{Segment-type} * \text{Target-speed} * \text{Conversation} + \zeta_{2i}$$

$$\beta_{3i} = \gamma_{30} + \gamma_{31} * \text{Segment-type} * \text{Target-speed} * \text{Conversation}$$

$$\beta_{4i} = \gamma_{40} + \gamma_{41} * \text{Segment-type} * \text{Target-speed} * \text{Conversation}$$

Table 2

Average distances to target by conversational condition with standard error in parentheses.

Experiment	Segment Type				
	Control	Talk	Listen	Prepare	Pause
E1	40.47 (1.57)	44.34 (1.98)	37.61 (1.5)	40.44 (1.84)	41.91 (1.95)
E2	37.18 (2.3)	37.82 (2.24)	35.5 (1.97)	36.04 (2.43)	35.92 (2.16)
					38.18 (2.21)

Table 3

Analysis 1a models. (a) Fit information for compared models. (AIC_c – AIC corrected, K – degrees of freedom, Δ – change in AIC_c , W – relative weight of evidence for model among compared models, E – ratio of evidence for model in comparison to the most likely model). (b) Maximum likelihood model comparison. Each model is compared to the one immediately above it using Chi square test of log likelihoods.

(a)	Model	K	AIC_c	Δ	W	E
	quadratic	45	76068.73	0.00	.98	1
	cubic	56	76077.01	8.28	.02	62.68
	slope	34	76078.34	9.61	.01	122.02
	quartic	67	76096.2	27.47	0	920420
	intercept	23	76123.49	54.76	0	$7.78 * 10^{11}$
	base	12	90628.42	14559.69	0	∞

(b)	Model	df	AIC	logLik	χ^2	df	p
	base	12	90628	-45302			
	intercept	23	76123	-38039	14527.01	11	$2.2 * 10^{-***}$
	slope	34	76078	-38005	67.28	11	$4.01 * 10^{-***}$
	quadratic	43	76067	-37991	28.92	11	$6.69 * 10^{-4***}$
	cubic	56	76076	-37982	13.95	11	0.24
	quartic	67	76095	-37981	3.08	11	0.99

Table 4

Analysis 1b models. (a) Fit information for compared models. (AIC_c – AIC corrected, K – degrees of freedom, Δ – change in AIC_c , W – relative weight of evidence for model among compared models, E – ratio of evidence for model in comparison to the most likely model). (b) Maximum likelihood model comparison. Each model is compared to the one immediately above it using Chi square test of log likelihoods.

(a)	Model	K	AIC_c	Δ	W	E
	quadratic	36	56299.94	0	.95	1
	cubic	44	56305.73	5.79	.05	18.11
	quartic	52	56313.8	13.85	0	1019.13
	slope	28	56316.59	16.65	0	4123.91
	intercept	20	56348.08	48.13	0	$2.83 * 10^{10}$
	base	12	67505.45	11205.51	0	∞

(b)	Model	df	AIC	logLik	χ^2	df	p
	base	12	67505	-33741			
	intercept	20	56348	-28154	11173.44	8	$2.2 * 10^{-16}$ ***
	slope	28	56316	-28130	47.59	8	$1.18 * 10^{-7}$ ***
	quadratic	36	56300	-28114	32.79	8	$6.73 * 10^{-5}$ ***
	cubic	44	56305	-28109	10.38	8	0.24
	quartic	52	56313	-28105	8.15	8	0.42

Table 5

Analysis 2a models. (a) Fit information for compared models. (AIC_c – AIC corrected, K – degrees of freedom, Δ – change in AIC_c , W – relative weight of evidence for model among compared models, E – ratio of evidence for model in comparison to the most likely model). (b) Maximum likelihood model comparison. Each model is compared to the one immediately above it using Chi square test of log likelihoods.

(a)	Model	K	AIC_c	Δ	W	E
	quadratic	45	61273.91	0.00	0.98	1
	cubic	56	61281.73	7.82	0.02	49.95
	quartic	67	61288.43	14.52	0	1421.35
	slope	34	61287.9	13.99	0	1091.43
	intercept	23	61324.17	50.26	0	$8.21 * 10^{10}$
	base	12	68255.02	6981.1	0	∞

(b)	Model	df	AIC	logLik	χ^2	df	p
	base	12	68255	-34115			
	intercept	22	61324	-30639	6952.95	11	$2.2e^{-16}$ ***
	slope	34	61288	-30610	58.44	11	$1.8 * 10^{-8}$ ***
	quadratic	45	61273	-30592	36.22	11	$1.55 * 10^{-4}$ ***
	cubic	56	61281	-30584	14.48	11	0.21
	quartic	67	61287	-30577	15.67	11	0.15

Table 6

Analysis 2b models. (a) Fit information for compared models. (AIC_c – AIC corrected, K – degrees of freedom, Δ – change in AIC_c , W – relative weight of evidence for model among compared models, E – ratio of evidence for model in comparison to the most likely model). (b) Maximum likelihood model comparison. Each model is compared to the one immediately above it using Chi square test of log likelihoods.

(a)	Model	K	AIC_c	Δ	W	E
	cubic	44	45614.35	0.00	.55	1
	quartic	52	45616.11	1.76	0.23	2.41
	quadratic	36	45616.16	1.81	0.22	2.47
	slope	28	45631.75	17.41	0	6021.88
	intercept	20	45660.97	46.62	0	$1.33 * 10^{10}$
	Base	12	50797.68	5183.33	0	∞

(b)	Model	df	AIC	logLik	χ^2	df	p
	base	12	50798	-25387			
	intercept	20	45661	-22810	5152.8	8	$2.2e^{-16}$ ***
	slope	28	45631	-22788	45.36	8	$3.15 * 10^{-7}$ ***
	quadratic	36	45616	-22772	31.78	8	$1.02 * 10^{-4}$ ***
	cubic	44	45614	-22763	18.04	8	0.02 *
	quartic	52	45615	-22756	14.52	8	0.07 .