Cognition 136 (2015) 304–324

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Variation in dual-task performance reveals late initiation of speech planning in turn-taking

Matthias J. Sjerps [a,b,c,*], Antje S. Meyer [a,b]

[a] Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands
[b] Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University, Nijmegen, The Netherlands
[c] Department of Linguistics, University of California Berkeley, Berkeley, California, USA

ARTICLE INFO

Article history:
Received 21 August 2013
Revised 29 September 2014
Accepted 20 October 2014

Keywords:
Dual-task
Turn-taking
Speech production
Speech perception
Eye-tracking

ABSTRACT

The smooth transitions between turns in natural conversation suggest that speakers often begin to plan their utterances while listening to their interlocutor. The presented study investigates whether this is indeed the case and, if so, when utterance planning begins. Two hypotheses were contrasted: that speakers begin to plan their turn as soon as possible (in our experiments less than a second after the onset of the interlocutor's turn), or that they do so close to the end of the interlocutor's turn. Turn-taking was combined with a finger tapping task to measure variations in cognitive load. We assumed that the onset of speech planning in addition to listening would be accompanied by deterioration in tapping performance. Two picture description experiments were conducted. In both experiments there were three conditions: (1) Tapping and Speaking, where participants tapped a complex pattern while taking over turns from a pre-recorded speaker, (2) Tapping and Listening, where participants carried out the tapping task while overhearing two pre-recorded speakers, and (3) Speaking Only, where participants took over turns as in the Tapping and Speaking condition but without tapping. The experiments differed in the amount of tapping training the participants received at the beginning of the session. In Experiment 2, the participants' eye-movements were recorded in addition to their speech and tapping. Analyses of the participants' tapping performance and eye movements showed that they initiated the cognitively demanding aspects of speech planning only shortly before the end of the turn of the preceding speaker. We argue that this is a smart planning strategy, which may be the speakers' default in many everyday situations.

1. Introduction

A hallmark of natural conversation is turn-taking, with interlocutors alternating in adopting the roles of listener and speaker. Speakers normally manage to coordinate their contributions to a conversation in such a way that their utterances follow smoothly on from each other, rather than overlapping or being separated by long pauses (Sacks, Schegloff, & Jefferson, 1974). For instance, Stivers et al. (2009) analysed a corpus of polar (yes/no) question–answer sequences in ten languages and found that the average interval between a question and the answer was around 200 ms. Data from Dutch corpora containing a range of different utterance types has provided a similar estimate (Heldner & Edlund, 2010). Many authors have stressed that natural conversations are characterised by smooth transitions of turns (Sacks et al., 1974; Wilson & Wilson, 2005). Moreover, there is good evidence that inter-turn intervals can convey meaning; for instance, a

* Corresponding author at: Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University, Nijmegen, The Netherlands. Tel.: +31 243521207.
E-mail address: m.j.sjerps@gmail.com (M.J. Sjerps).

http://dx.doi.org/10.1016/j.cognition.2014.10.008
0010-0277/

long pause before an answer to a request may indicate reluctance to comply (Jefferson, 1989; Roberts & Francis, 2013; Roberts, Margutti, & Takano, 2011). Thus, speakers know how to time their contributions appropriately.

However, very little is known about the way this timing is achieved. Short inter-turn intervals and occasional overlaps of turns indicate that speakers often begin to plan their utterance while still listening to the other person (De Ruiter, Mitterer, & Enfield, 2006; Levinson, 2013; Sacks et al., 1974). This is because planning a single content word (e.g., a name of a picture) may take close to a second (Indefrey & Levelt, 2004; Strijkers & Costa, 2011) and initiating a simple descriptive utterance, such as "The donkey kicked the man", may take about two seconds (Gleitman, January, Nappa, & Trueswell, 2007; Griffin & Bock, 2000).

But when do speakers begin to plan their turns? Do they typically begin to plan an utterance as soon as they have a rough idea what they might say, or do they only begin to plan when they can anticipate that the interlocutor's turn is about to end? If utterance planning and listening indeed co-occur in time, how do speakers distribute their processing resources across these tasks? And how do concurrent listening and speech planning affect each other? Addressing these and related questions is crucial for understanding how we speak and comprehend speech in everyday contexts. So far, however, most experimental psycholinguistic work has concerned monologues and little is known about the way listening and speech planning are coordinated in everyday conversations. In the present study, we developed a new paradigm to assess the coordination of speaking and listening in a simple turn-taking task.

The basic idea underlying the study was that initiating speech planning whilst listening to another person should increase the mental load for the speaker, and that this increase in mental load should lead to a performance decrement in a motor task carried out concurrently with the linguistic task. With this approach we built upon results of numerous dual-task experiments showing that performance in a cognitively demanding task deteriorates when it is carried out simultaneously with another cognitively demanding task rather than by itself (Baddeley, 1976; Becic et al., 2010; Bock, Dell, Garnsey, Kramer, & Kubose, 2007; Duncan, 1980; Kemper, Herman, & Lian, 2003; Kemper, Schmalzried, Herman, & Mohankumar, 2011; Lavie, 2005; Lavie, Hirst, de Fockert, & Viding, 2004; Meyer & Kieras, 1997; Pashler, 1984, 1994). There are various accounts of dual-task interference but most of them share the assumption that there is a limit to the overall amount of cognitive resources that can be attributed to concurrent cognitive tasks (Duncan, 1980; Kahneman, 1973; Marois & Ivanoff, 2005; Watanabe & Funahashi, 2014; Wickens, 1980). When capacity needs to be distributed across two tasks (rather than being exclusively dedicated to one task) performance in one or both tasks suffers (Somberg & Salthouse, 1982). Related accounts assume that dual-task interference arises because of limitations to central executive control or monitoring processes (Baddeley & Hitch, 1974; D'Esposito et al., 1995). In addition to these domain-general sources of interference, there may be interference in specific processing components, such as verbal working memory, visual processing, or motor planning, drawn upon by both tasks (Bergen, Medeiros-Ward, Wheeler, Drews, & Strayer, 2013; Pashler, 1994).

Most relevant to the current study are dual-task studies that have shown that speaking and listening are prone to dual-task interference. Much of this work concerned the way listening and speaking (for instance using a mobile phone) can be combined with driving and therefore has used braking, following, or lane-keeping tasks (Becic et al., 2010; Horrey & Wickens, 2006; Kubose et al., 2006; Kunar, Carter, Cohen, & Horowitz, 2008; Strayer, Drews, & Johnston, 2003; Strayer & Johnston, 2001). Other studies were carried out in the context of research on aging and combined linguistic tasks with motor tasks such as walking, finger tapping, or tracking a moving target on a computer screen (Kemper, Herman, & Nartowicz, 2005; Kemper et al., 2003). These lines of research have yielded abundant evidence for dual-task interference between speaking or listening and concurrent non-linguistic tasks. This demonstrates that non-negligible amounts of processing capacity are required for talking and listening (for corroborating evidence from studies using other paradigm see, for instance, (Caplan & Waters, 2013; Cleland, Tamminen, Quinlan, & Gaskell, 2012; Cook & Meyer, 2008; Ferreira & Pashler, 2002; Gordon, Eberhardt, & Rueckl, 1993; Mattys, Brooks, & Cooke, 2009; Papesh & Goldinger, 2012; Roelofs & Piai, 2011). Moreover, these studies have demonstrated that dual-task paradigms are suitable to measure differences in the capacity demands imposed by different linguistic tasks. A common (though not universal) finding is, for instance, that speaking interferes more with secondary task performance, and hence appear to require more capacity, than listening (Almor, 2008; Kubose et al., 2006; Kunar et al., 2008; Recarte & Nunes, 2003) but see (Kubose et al., 2006).

Recently Boiteau, Malone, Peters, and Almor (2014) used a dual-task paradigm to investigate the cognitive demands in turn-taking situations. In their study, participants' primary task was to engage in an unscripted 15-min conversation with a confederate (Experiment 1) or a friend (Experiment 2). The secondary task was a continuous visuomotor task, which consisted of tracking a moving target on a computer screen using the computer mouse. The tracking task was carried out by itself (control condition) and throughout the conversation. The authors recorded the participants' speech rate and fluency in the conversation and their performance in the tracking task, measured as the distance between the target and the cursor. Specifically, they examined the tracking performance in the tracking-only control condition and in 480-ms time windows at the beginning and at the end of utterances the participants heard or produced, and at the ends of pauses preceding or following the participants' utterance onsets. Boiteau and colleagues found that the participants' performance in the tracking task deteriorated in the conversation compared to the control condition. In addition they found that overall the participants' tracking performance was better during listening than during speaking or during the planning pauses preceding their utterances. Further analyses showed that the participants' tracking performance