

Bilingualism: Language and Cognition

<http://journals.cambridge.org/BIL>

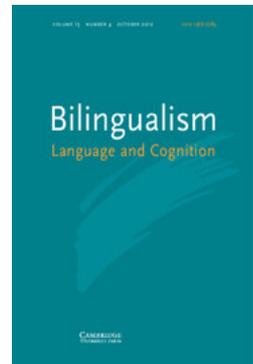
Additional services for *Bilingualism: Language and Cognition*:

Email alerts: [Click here](#)

Subscriptions: [Click here](#)

Commercial reprints: [Click here](#)

Terms of use : [Click here](#)



The effect of dual task demands and proficiency on second language speech production

MATHIEU DECLERCK and JUDIT KORMOS

Bilingualism: Language and Cognition / Volume 15 / Issue 04 / October 2012, pp 782 - 796
DOI: 10.1017/S1366728911000629, Published online: 04 January 2012

Link to this article: http://journals.cambridge.org/abstract_S1366728911000629

How to cite this article:

MATHIEU DECLERCK and JUDIT KORMOS (2012). The effect of dual task demands and proficiency on second language speech production. *Bilingualism: Language and Cognition*, 15, pp 782-796 doi:10.1017/S1366728911000629

Request Permissions : [Click here](#)

The effect of dual task demands and proficiency on second language speech production*

MATHIEU DECLERCK

Rheinisch-Westfälische Technische Hochschule
Aachen University, Institute of Psychology

JUDIT KORMOS

Department of Linguistics and English Language,
Lancaster University

(Received: February 20, 2011; final revision received: September 21, 2011; accepted: September 27, 2011; first published online 4 January 2012)

In this study we examined how the introduction of a parallel finger-tapping task influences second language (L2) speech encoding mechanisms and monitoring processes, and how the level of proficiency impacts the efficiency and accuracy of L2 performance under single and dual task conditions. The results indicate that imposing dual task demands had a negative effect on the accuracy of lexical selection and the efficiency of error-correction processes. We argue that this can be explained with reference to attentional bottleneck effects on lexical selection processes and on monitoring. The findings also reveal that the level of L2 competence influenced both the speed and the accuracy of speech encoding processes and the efficiency of monitoring.

Keywords: second language production, dual task, monitoring, second language proficiency

Introduction

The role of attention in second language (L2) speech production has been at the forefront of investigations in task-based research, which aims to answer the question how particular characteristics of a task affect L2 speakers' allocation of attention to different aspects of task performance. Task-based research has primarily concentrated on the influence of task complexity on different output measures such as complexity, accuracy and fluency (for a recent discussion see Norris & Ortega, 2009) and the role of task characteristics on the underlying speech production mechanisms has only been discussed in a few studies (e.g., Kormos, 2011; Robinson, 2007; Skehan, 2009). Moreover, research on the impact of task demands on monitoring processes has been scarce despite the fact that monitoring can enhance the efficiency of L2 acquisition because it involves both attention and conscious processing as well as producing output (de Bot, 1996; Gilabert, 2007; Izumi, 2003; Kormos, 2000).

In the field of task-based research, a large number of studies have investigated how increased attentional demands affect L2 performance through the provision of planning time (for a review see Ellis, 2005). Manipulating planning time, however, influences not only the availability of attentional resources but also

the focus of attention and the retrieval of prior content and linguistic knowledge (Ortega, 1999). Consequently, planning time studies might not give a direct insight into how increased attentional demands influence speech production processes. Our study is, to the best of our knowledge, the first one in the L2 field in which task demands were experimentally manipulated by asking students to perform a parallel finger-tapping task while speaking.

Studies on dual task demands in speech production research are scarce (see e.g., Cook & Meyer, 2008; Ferreira & Pashler, 2002; Hartsuiker & Barkhuysen, 2006; Roelofs, 2008) and the majority of this line of research has used picture naming tasks (for exceptions see e.g., Hartsuiker & Barkhuysen, 2006; Oomen & Postma, 2002). Consequently, little is known about how interference with a parallel non-verbal task affects speech production processes beyond the processes of lexical encoding. The use of a simple sentence-production task as in our research has the potential to yield insights into how additional attentional demands through a kinaesthetic task impact syntactic encoding and monitoring processes and the overall efficiency and speed of speech production.

In this paper we review the psycholinguistic processes of speech production and monitoring and discuss the role of attention in encoding and monitoring L2 speech. Next we describe the procedures of our research, which is followed by the presentation and discussion of the results. The paper concludes with a summary of the findings and a discussion of their relevance to the field of second language acquisition.

* The first author was supported by a grant funded by the Hungarian Scholarship Board (MOB-193-1/2009). We thank the three anonymous reviewers of this paper for their detailed comments and useful suggestions.

Address for correspondence:

Judit Kormos, Lancaster University, Department of Linguistics and English Language, Bailrigg, County South LA1 5FB, UK
j.kormos@lancaster.ac.uk

Review of the literature

The psycholinguistic processes of speaking and monitoring

Speech production has four key components: (i) conceptualization, that is, planning what one wants to say; (ii) formulation, which includes the grammatical, lexical and phonological encoding of the message; (iii) articulation, in other words, the production of speech sounds; and (iv) self-monitoring, which involves checking whether the produced spoken output is correct and appropriate (Levelt, 1983). Information between the stages of processing is assumed to flow based on the mechanisms of activation spreading, that is, via networks of interconnected units that exchange simple signals via the connections they have with each other (Hebb, 1949).

The conceptualization of monitoring processes constitutes a major difference between the two major theories of first language (L1) speech production: spreading activation theory (this name is somewhat misleading because as just mentioned, both models assume that the way information is transmitted in the speech processing system is activation spreading) (e.g., Dell, 1986) and modular theories (e.g., Levelt, 1989, 1999; Levelt, Roelofs & Meyer, 1999). Spreading activation theories allow for the backward flow of activation from a lower level of processing to the higher level, whereas in modular theories, activation can only spread forward. Consequently, in spreading activation theory, if an error occurs in one specific process, a warning signal is immediately issued, and activation flows backward to the previous level, from which processing will restart. Therefore, in this theory, monitoring is assumed to be an inherent feature of the perception and production processes, and no separate monitoring device is postulated (Dell, 1986).

In modular models of speech production and monitoring such as Levelt's (1989) Perceptual Loop Theory, linguistic encoding errors are not noticed at the level they are made, but only once the erroneous fragment of speech has been phonologically encoded or later when it is articulated. Hence in this view, bits of message that contain an error need to be encoded again from the level of conceptualization. In this model the same mechanism is applied for checking one's own message as for the perception and checking of other speakers' utterances. In order to avoid the necessity for reduplication of knowledge, in the Perceptual Loop Theory the speech comprehension system is used for attending to one's own speech as well as to others. The outcome of the production processes is inspected in three monitoring loops (i.e., direct feedback channels leading back to the monitor). The first loop involves the

comparison of the preverbal plan, which contains the conceptual specifications for the message to be conveyed, with the original intentions of the speaker before the plan is linguistically encoded. In this loop the preverbal plan might need modification because the speaker might find that the formulated message is not appropriate in terms of its information content or in the given communicative situation. In the second loop the phonetic plan (i.e. 'internal speech') is checked before articulation, which is also called 'covert monitoring' (Postma, 2000; Postma & Kolk, 1992, 1993). In this stage the speaker might notice an error such as a wrongly selected word before it is actually uttered. Finally, the generated utterance is also scrutinized after articulation, which constitutes the final, external loop of monitoring. When perceiving an error or inappropriacy in the output in any of these three loops, the monitor issues an alarm signal, which, in turn, triggers the production mechanism for a second time. In Perceptual Loop Theory monitoring requires conscious attention and is supposed to be constrained by the processing constraints of working memory, which is responsible for the regulation of attention (Oomen & Postma, 2002).

Mechanisms of L1 and L2 monitoring and self-repair behavior share a number of similarities, in that the distribution and detection of self-repairs display an analogous pattern in the processes of L1 and L2 acquisition and production (Kormos, 2000; van Hest, 1996). Nevertheless, due to lack of automaticity in L2, monitoring in L1 differs from monitoring in L2 as regards the amount of attention available for error detection. Furthermore, due to the fact that the L2 speakers' system of knowledge might be incomplete and their production mechanisms are not fully automatic, certain repair mechanisms such as corrections performed in cases when L2 users are uncertain about the correctness of their utterance, occur in L2 speech that are not – or only very rarely – observable in L1 production.

The role of attention in speech production

For a long time, attention was regarded as a single resource pool that cannot be divided between two simultaneous tasks (Broadbent, 1958). However, research evidence from a number of studies suggested that there are multiple resource pools for different processing modalities (Kahneman, 1973; Navon, 1984; Sarno & Wickens, 1995; Shah & Miyake, 1996; Wickens, 2007). The existence of multiple attentional resources, such as the separate attention pools for auditory and visual information, can explain for example how one can listen to music while typing a letter. Performance on concurrent tasks was found to deteriorate if both tasks draw on the same pool of attentional resources (Wickens, 1980). In a number of recent theories of attentional resources, the

traditional assumption that attentional capacity is limited has been questioned. Interference models of attention (e.g., Navon, 1989; Sanders, 1998) explain deteriorating performance in dual- or multi-task situations by proposing that reduced performance is not caused by the limited capacity of the attentional system but is the result of interference and cross-talk caused by shifts of attention between tasks. Higher task demands, characterized by a greater amount of stimuli/input and many or similar response alternatives, lead to competition for attention between the same types of information units and memory traces and to cross-talk between them (Sanders, 1998).

More recently, however, it has been shown that interference between processes drawing on potentially different resource pools of attention can also exist (e.g., between driving a car and having an intense discussion). In their study, Ferreira and Pashler (2002) provided empirical evidence for this by demonstrating that the introduction of a parallel auditory discrimination task slowed down picture-naming considerably. Their experiments also indicated that parallel processing demands only affected the stages of speech production which required conscious response selection, that is, retrieving the appropriate lemma and the phonological form of the word corresponding to the concept to be named, but not phoneme selection, which is an automatic response-execution process.

Attentional control and regulation is associated with processes in working memory (Gathercole, 1999), which combine storage with the processing and manipulation of information (Baddeley, 2003). The working memory model comprises a multi-component memory system consisting of the central executive, which coordinates two modality-specific subsystems, the phonological loop and the visuo-spatial sketchpad. The visuo-spatial sketchpad works with visual and spatial information, while the phonological loop is specialized in the manipulation and retention of speech. The episodic buffer stores the intermediary products of phonological and/or visual working memory processes temporarily and assists in their integration (Baddeley, 2000). The central executive has several functions, including attentional control, directing the flow of information through the system and planning (Gathercole, 1999). The central executive is also responsible for regulating interference effects between competing stimuli (Kane & Engle, 2003). A large number of models in attentional control postulate a bottleneck effect in central processing due to the inherent limitations of the capacity of working memory (for a review see Pashler & Johnston, 1998).

In L1 speech production two processes are subject to conscious attentional control: conceptualization and monitoring; whereas linguistic encoding mechanisms are largely automatic (Levitt, 1989). Therefore,

attention plays a role in planning one's message and checking whether one has encoded the intended message appropriately and accurately, hence planning and monitoring processes might interfere with each other (Horton & Keysar, 1996; Levitt, 1983). High conceptual demands on monolingual speakers might cause interference with monitoring processes, and consequently certain errors might not be noticed either before or after articulation (Levitt, 1983). Similarly, if there is a parallel task that depletes attentional resources as in Oomen and Postma's (2002) study, which our research replicates, the efficiency of error detection might decrease substantially. As L1 speakers can monitor their encoded message at a high speed, certain errors might be intercepted in the pre-articulatory phase of monitoring (Hartsuiker & Kolk, 2001; Postma, 2000; Postma & Kolk, 1992, 1993). If, however, speakers' attentional resources are used primarily for planning their utterances, inaccuracies might not be corrected before they are articulated, and sometimes not even after articulation either; therefore one might find a higher number of errors in the output. Furthermore, L1 lexical selection and retrieval processes might also require certain levels of consciousness and might be prone to interference effects (Ferreira & Pashler, 2002).

As we have seen, attention influences a number of areas of L1 speech production, but in L2 speech processing, its role is even more complex. For an L2 speaker not only the stages of planning and monitoring are subject to attentional control, but depending on the level of proficiency, L2 speakers display varying degrees of automaticity in linguistic encoding. For an L2 speaker it might require a conscious search mechanism to retrieve the appropriate lemma matching the activated concept and to perform the ensuing syntactic and morphological encoding procedures (for a detailed discussion of L2 speech production procedures see Kormos, 2006). Speaking is an online activity that takes place under time constraints; hence L2 speakers often need to balance fluency with the complexity and grammatical accuracy of their message (Kormos, 2011; Robinson, 2001, 2003, 2005; Skehan & Foster, 2001). This explains why there are trade-off effects in accuracy and fluency if cognitive demands of a given task are high. If conceptualizing the message requires particular attention on the part of the speaker, fewer resources will be available for lexical, syntactic and phonological encoding as well as for monitoring, which might result in more errors in the students' output and in reduced fluency.

The role of attention in monitoring

Whereas the influence of attentional demands on L2 speech production has been widely researched, the role of attention in speech monitoring has been a slightly

neglected area of investigation. Most of the early studies in this field were only concerned with establishing the distribution of various types of self-repairs, and did not attribute high importance to the discussion of the allocation of attention. These studies mostly used raw percentages of occurrence rather than standardized frequency data (e.g., Fathman, 1980; Lennon, 1984; van Hest, 1996) to investigate what type of error L2 speakers' monitor is sensitive to and failed to consider the ratio of repaired errors. Despite these problematic issues of research methodology, researchers of L2 production assumed that L2 learners pay considerably more attention to lexical appropriacy than to grammatical accuracy (e.g., Fathman, 1980; Lennon, 1984; Poulisse, 1993; Poulisse & Bongaerts, 1994; van Hest, 1996). Investigations concerning L2 self-repairs also revealed that the frequency of repairs concerning the information content of the message varies across different types of tasks (Poulisse 1993; van Hest, 1996).

Kormos' (2000) study, which investigated the distribution and frequency of self-repairs and the correction rate of errors in the speech of Hungarian learners at three levels of proficiency (pre-intermediate, upper-intermediate, advanced) and native speakers of Hungarian, was specifically devoted to the examination of the role of attention in monitoring L2 speech. The global distribution of self-repairs in the research showed that in an information exchange task Hungarian L2 learners paid approximately equal attention to the appropriacy and adequacy of the informational content of their utterance as to linguistic accuracy. The analysis of the correction rate of lexical and grammatical errors seemed to indicate a similar tendency. Kormos' (2000) study also confirmed that with the development of language proficiency, L2 speakers' attention in monitoring shifts from lower level linguistic errors to problems arising at the discourse level (see also Gilabert, 2007 for a similar finding).

In recent task-based investigations, self-repairs were also analyzed to gain an insight into how different task-conditions affect students' output. Yuan and Ellis (2003) found that participants who were provided with online planning time made more reformulations and self-repairs than those whose performance was limited by time constraints. In contrast, both Gilabert's (2006) and Yuan and Ellis' (2003) studies suggested that the frequency of self-repairs did not increase significantly when students were provided with pre-task planning time. Gilabert's (2007) research revealed that when tasks increased in complexity, students corrected a higher proportion of errors, which indicated that the tasks employed in his study had the potential to direct students' attention to accuracy. It has to be noted, however, that the frequency of self-repairs and the correction rate of errors might not be a direct indicator of overall linguistic accuracy as it does not take into account cases when students make no

or few errors that remain uncorrected. These variables should, rather, be seen as indicators of the efficiency of monitoring processes.

The aims of the present study

As shown in the review of the literature, there is a scarcity of studies that have investigated the effect of various task characteristics on L2 monitoring processes. Moreover, to our knowledge, no study in the L2 field has manipulated the attentional demands of language production tasks experimentally. Therefore, the primary aim of the study presented in this paper was to investigate how dual task demands affect L2 speech production processes. For this purpose, we introduced a finger-tapping task, in which students had to hit keys on a keyboard randomly while speaking. Studies in experimental psychology have suggested that this task requires a large amount of central attention and can disrupt the operation of the central executive, which is responsible for the regulation of attentional control (e.g., Baddeley, Emslie, Kolodny & Duncan, 1998). Consequently, we hypothesized that the finger-tapping task might interfere with speech production processes that require attentional control in L2, most importantly, linguistic encoding and monitoring. Interference with these speech encoding processes might cause higher number of errors in lexis and syntax. It was also hypothesized that the parallel finger-tapping task might decrease the efficiency of monitoring processes and the speed with which errors are detected and corrected. Furthermore, as a global effect of the dual task, reduction in the fluency of the utterance might also be observed.

In our study we investigated two distinct groups of L2 speakers: intermediate and advanced language learners with the aim of analyzing how interference with speech production processes might vary at different levels of L2 competence. We were also interested in analyzing proficiency-related variation in the accuracy and speed of speech production in general and in monitoring in particular.

The research questions of our study were the following:

1. How do dual task demands influence fluency and accuracy of L2 speech production?
2. How do dual task demands influence the processes and speed of monitoring in L2?
3. How do the fluency and accuracy of speech production and the speed and efficiency of monitoring differ at intermediate and advanced levels of L2 competence?
4. Is there an interaction between dual task demands and proficiency in L2 speech production and monitoring?

Method

Participants

Twenty native speakers of Hungarian participated in the study. Ten of them were paid first-year university students (first-year Bachelors degree in English). Their age ranged from 18–23 years ($M = 19.2$ years). Their score on the paper-based version of the Oxford Quick Placement test was between 46 and 58 points out of 60. Based on the equivalence scheme provided by the testing manual, this indicates that their level of proficiency was at the C1 level of the Common European Framework of Reference (CEFR) (Council of Europe, 2001), which can be regarded as an advanced level of L2 competence ($M = 51.7$, $SD = 3.74$). The other ten participants were secondary school students aged 16–17 years ($M = 16.8$ years), who gained between 30 and 36 points on the Oxford Quick Placement Test, which was established to be equivalent to the B1 level in the CEFR based on the testing manual, and could thus be considered intermediate range of language proficiency ($M = 33.7$, $SD = 2.12$). An independent samples t -test demonstrated that the difference in proficiency test scores between the groups was significant ($t(8) = -13.9$, $p < .001$).

Both groups consisted of five male and five female students. One of the secondary school participants, however, had to be excluded after data analysis as he did not produce random finger-tapping in the single and dual task condition.

Materials

Speech samples were collected from the participants using a network description task, which consisted of 32 trials, 16 trials in the single task condition and 16 trials in the dual task condition, in which participants were required to perform a finger-tapping task in parallel. In each of these trials the participants were presented with a visual network of lines that connected eight different pictures of objects. Each network contained different objects to describe. Similar networks were used by Levelt (1983), Martin, Weisberg and Saffran (1989) and Oomen and Postma (2001, 2002). An example of a network is given in Figure 1. The networks were presented using the Microsoft PowerPoint program.

The participants had to describe the path of a red dot, which moved differently over each network. The description of the movement of the dot had to contain the direction (up, down, left or right), the position and shape of the line (upper, lower, right or left curved line or diagonal line or straight line) and the next object in a sentence (e.g., the dot goes left over the straight line to the flower). This task did not exert substantial conceptualization demands on the learners as they had to

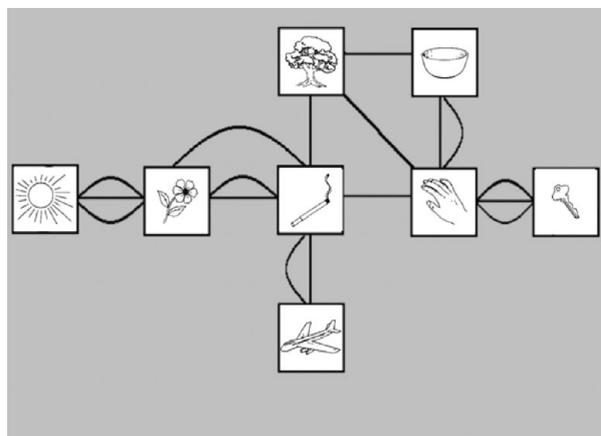


Figure 1. Example of a network trial.

describe predetermined content and were not required to plan the information content of their message themselves. The task, however, made substantial demands on lexical encoding as students needed to activate the lemmas matching and characterizing the objects and the direction of movement. Most of the objects were high-frequency lexical items with approximately 10% of the words being outside the 1000 most frequent words in English (West, 1953). With regard to syntactic encoding, the task induced the use of complex noun phrases and simple clausal structures. The appendix contains two examples of performance by an intermediate and advanced learner.

Each network task consisted of nine steps of movement. The dot completed its path over each network in 55 seconds, allowing the participants to use a normal speech rate. The timing of the movement of the dot was carefully piloted so that students would not feel a time pressure exerted by the movement of the dot. Each of the line drawings was presented twice – once during the single task condition and once during the dual task condition.

The finger-tapping task consisted of 32 trials: 16 trials in the random generation condition, in which the participants were only required to do the finger-tapping task, and 16 trials in the dual task condition, in which students had to perform the finger-tapping task and the network task simultaneously. For the finger-tapping task the participants used a keyboard with ten white keys. Each white key stood for a finger position. The position of the ten keys was chosen so that students could easily tap on the keys and that the task would not require prior typing knowledge.

Procedures

Before the experiment, the Oxford Quick Placement test was administered to a larger group of university and secondary school students. Participants of the study were selected from the students whose scores on the test fell in

Table 1. *Repair coding scheme.*

Type of repair	Example
Error repair	
Lexical repair	Over an upper . . . over a lower curved
Phonological repair	To the clo . . . to the coat
Syntactic repair	The dot go . . . the dot goes
Appropriateness repair	
Order repair	It's going to . . . on the straight line to the
Appropriate information repair	The nurse . . . the nurse with the syringe
Rephrasing repair	On the right . . . go right

the B1 range for the secondary school students and in the C1 range for the university students.

The experimental set-up followed the procedures used in Oomen and Postma's (2002) study. The participants were tested under three conditions. The first was the random generation condition, in which the students had to press one out of ten keys, each corresponding to a different finger, every second as randomly as possible. They were given two practice trials and 16 experimental trials of 55 seconds each, and when necessary, feedback was given by the experimenter (e.g., when they focused too much on one hand or finger). This task was preceded by a brief explanation of what constituted randomness.

The other two conditions were counterbalanced as the second and third blocks of the experiment. This meant that half of the participants performed the dual condition task first and the single condition second, whereas the other half of the participants were asked to do the single condition task first followed by the dual condition task. Before administering these blocks, a pre-recorded example of the network task was shown to the participants. Both blocks consisted of two practice trials and 16 experimental trials, which lasted 55 seconds each. During both blocks the participants had to perform the network task described above. While this was the only task in the single task condition, in the dual task condition the participants had to perform the network task simultaneously with the finger-tapping task.

The performance of the candidates was transcribed by the first author and recorded by the internal microphone of the computer that was used with the help of the program WaveLab version 5.01a, which has an accuracy of 0.1 ms.

Data analyses

The randomness of the finger-tapping task was calculated for each trial using the Random Number Generation Index (RNG) (Evans, 1978). The RNG scores range between 0.00 and 1.00. The more these scores diverge from 0.00, the less random they can be considered. We used the

RGCalc program (Towse & Neil, 1998) to calculate the RNG scores.

The speech data was transcribed and coded by the first author. The analysis included the identification and classification of different errors (lexical, phonological and syntactic errors), repairs (lexical, phonological, syntactic, order, appropriateness and rephrasing repairs), the amounts of words, the disfluencies (filled pauses and repetitions) and the error to cut-off and the cut-off to repair times.

The coding scheme used in the present study was an adaptation of the scheme used by Levelt (1983) and Kormos (2000). The six different types of repair were separated into two distinct categories: error-repairs, which included the correction of lexical, phonological and syntactic errors, and appropriateness repairs, which were employed when the speaker decided to encode the originally intended information in a modified way. In these repairs the speakers did not correct linguistic errors but reformulated the originally provided information. Three types of appropriacy repair were analyzed in the current study: those in which the order of information was changed (different-order repairs), the message was rephrased (rephrasing repairs), and the information was corrected (appropriate information repair). For the full coding scheme with examples, see Table 1. The raw number of repairs was converted into frequency data by calculating the frequency of repairs over the total amount of spoken words during a block.¹

For the analysis of errors, Lennon's (1991) definition of errors was used. In this definition an error is "a linguistic form or combination of forms, which in the same context and under similar conditions of production, would in all likelihood, not be produced by the speakers' native speaker counterparts" (p. 182). On the basis of this definition, all the possible errors were identified in the transcripts. The coding scheme of the errors was

¹ Overall the occurrence of unnecessary corrections was infrequent (n = 14). Correct forms that were changed into an incorrect form occurred even more rarely.

Table 2. *Error-coding scheme.*

Type of repair	Example
Error	
Lexical error	Over the straight line (instead of lower curve)
Phonological error	Straige line (instead of straight line)
Syntactic error	Over lower curve (instead of over a lower curve)

devised so as to correspond to the categorization of error-repairs. In this project, errors of lexis, syntax and phonology were classified according to where the error occurred in the speech production process. Lexical errors were the result of inappropriate lexical access, whereas syntactic errors were inaccuracies that were the result of faulty grammatical encoding processes. Phonological errors were mispronunciations of words that indicated problems in the phonological encoding and articulating phases of speech production (see Table 2 for examples in each category). The frequency of each error over the total number of spoken words during a block produced was used as the measure of accuracy. The correction rate of errors was also analyzed so as to gain an insight into the efficiency of monitoring processes. The correction rate of each type of error (lexical, syntactic and phonological) was calculated by dividing the number of respective self-repairs by the total number of errors in the given category.

The error to cut-off times were measured from the beginning of the incorrect word until the beginning of the cut-off. The cut-off to repair times were calculated from the beginning of the cut-off until the beginning of the corrected speech. Both the error to cut-off times and the cut-off to repair times were analyzed using the software WaveLab version 5.01a.

In order to assess how the introduction of a parallel task affected the overall fluency of the students' speech, both disfluencies and the pruned speech rate were analyzed. From among the phenomena indicating breakdowns in the speech encoding processes, the frequency of repetitions and filled pauses such as *uhm* and *er* was measured. Speech rate was calculated as the total number of syllables produced per second.

To ensure the reliability of coding, 5% of the spoken data (one block of one participant) was coded by a second rater. The block and the participant were chosen randomly and were presented to the second rater together with the coding scheme. The convergence between the two raters was 75%. This is similar to the 70% of Oomen and Postma (2001), the 76% of Blackmer and Mitton (1991) and the 73% of Levelt (1983).

The 16 trials within each condition were considered as one task and results for the different trials were not analyzed separately.² The statistical analyses used were multiple analyses of variance (MANOVA), with the task-condition constituting the within-subject and the level of proficiency the between-subject factors. In order to test the differences in the frequency of different types of self-repair and the correction ratio of various types of error, repeated measures of analysis of variance were used. Effect sizes were also calculated to assess the importance of the obtained effects. Eta squared values below .06 were regarded as small, below .13 as medium and above .13 as indicating large effect size (Cohen, 1988).

Results

As can be seen in Table 3, the RNG values were around .3 in the single condition and around .4 in the dual condition. This represents a significant increase in the dual condition suggesting that finger-tapping behavior became less random when students also had to perform a speaking task. Unfortunately, there are no established cut-off points for randomness. Nevertheless, we can consider finger-tapping random in the dual condition based on the figures provided in Jahanshahi, Saleem, Ho, Drinberger and Fuller's (2006) validation study of randomness measures.³

The level of proficiency had no significant effect on finger-tapping and no interaction between group and condition was found either.

As can be seen in Table 4, advanced language learners used repetitions and filled pauses with frequency similar to that of the intermediate participants. As for the speech rate, advanced students spoke considerably faster than the intermediate learners, but dual task demands did not exert a significant influence on this measure of fluency. The task-condition did not have a significant effect on any of the investigated fluency variables, and no interaction between the task-condition and proficiency was found either.

With regard to errors, the most frequent type of error was grammatical inaccuracies under both conditions (Single condition: $M = 4.34$, $SD = 3.78$; Dual condition: $M = 4.08$, $SD = 4.47$). Lexical errors (Single condition: $M = 2.58$, $SD = 1.91$; Dual condition: $M = 3.89$, $SD = 3.43$) and phonological errors (Single condition: $M = 2.85$, $SD = 3.34$; Dual condition: $M = 3.38$, $SD = 4.45$) occurred with similar frequency. As can be seen in

² For the frequency of error-repairs $F(1,17) = 2.35$, n.s., and for repair ratio $F(1,17) = 0.31$, n.s., no practice effect across trials was found. The frequency of errors, however, decreased significantly across trials ($F(1,17) = 8.40$, $p < .01$). For all the other variables there was not enough data per trial to analyze the practice effect.

³ In this study by Jahanshahi et al. (2006) RNG values ranged between .3 and .4 under single task conditions.

Table 3. The effect of proficiency and task-condition on finger-tapping.

Group		Mean random generation task (SD)	Mean dual task (SD)	F group (η^2)	F condition (η^2)	F condition \times group (η^2)
Random generation	Advanced	0.290 (0.071)	0.420 (0.127)	0.87 (0.05)	44.08** (0.72)	0.14 (0.01)
	Intermediate	0.327 (0.085)	0.443 (0.131)			

** $p < .01$

Table 4. The effect of proficiency and task-condition on dysfluency variables.

Frequency over amount of spoken words during block		Mean single task (SD)	Mean dual task (SD)	F group (η^2)	F condition (η^2)	F group \times condition (η^2)
Repetitions	Advanced	0.51 (0.52)	0.42 (0.53)	0.01 (0.00)	0.08 (0.00)	0.25 (0.01)
	Intermediate	0.53 (0.32)	0.55 (0.48)			
Filled pauses	Advanced	3.00 (6.32)	5.00 (1.18)	0.01 (0.00)	0.51 (0.03)	0.83 (0.04)
	Intermediate	4.08 (5.92)	3.83 (3.41)			
Speech rate (syll/sec)	Advanced	1.91 (0.46)	1.86 (0.45)	6.32* (0.27)	1.86 (0.09)	0.15 (0.01)
	Intermediate	1.40 (0.45)	1.37 (0.36)			

* $p < .05$

Table 5, students produced significantly more lexical errors in the dual task condition, but no other types of error were influenced by the task-condition. On the other hand, the level of proficiency had a large effect on the frequency of all the sub-types of error. No interaction between the level of proficiency and the task-condition could be observed.

Error-repairs (Single condition: $M = 0.89$, $SD = 0.49$; Dual condition: $M = 0.89$, $SD = 0.47$) were more frequent than appropriacy repairs in both conditions (Single condition: $M = 0.24$, $SD = 0.21$; Dual condition: $M = 0.24$, $SD = 0.32$). Both types of repair occurred with identical frequency in the two tasks. As shown in Table 6, dual task demands did not have a significant effect on the frequency of any of the sub-types of self-repair, and the proficiency of the participants did not seem to exert substantial influence on self-repairs either. No interaction between the task-condition and proficiency level was detected in any of the categories of self-repairs. In order to test whether particular repair types occurred statistically more frequently than others, we conducted

repeated measure analysis of variance, which showed that there was a significant difference in the frequency of the different types of self-repair (Single condition: $F(2,17) = 16.65$, $p < .001$, $\eta^2 = .69$; Dual condition: $F(2,17) 25.55$, $p < .001$, $\eta^2 = .75$).

With regard to the overall repair rate of errors, it was found that intermediate learners repaired a significantly smaller proportion of their errors than advanced students, and participants' error-correction rate was also significantly lower in the dual than in the single condition (see Table 7). The findings also show that there was an interaction between task-condition and level of proficiency, and the data reveals that there was a more substantial decrease in the error-correction rate of the advanced participants than in the case of intermediate learners. As for the different sub-types of error, higher-proficiency students corrected a significantly greater proportion of lexical, syntactic and phonological errors. Interestingly, when the sub-types of error were analyzed for task-condition effects, the correction rate did not differ in the dual and single tasks in any of the

Table 5. *The effect of proficiency and task-condition on the frequency of errors.*

Error frequency over amount of spoken words during block	Group	Mean (SD)	Mean (SD)	F group (η^2)	F condition (η^2)	F condition \times group (η^2)
Total error	Advanced	3.97 (3.99)	5.28 (4.26)	13.32** (0.43)	2.06 (0.10)	0.67 (0.00)
	Intermediate	16.25 (12.17)	18.15 (8.77)			
Lexical error	Advanced	1.32 (1.61)	2.03 (2.80)	15.19** (0.47)	7.96* (0.31)	2.65 (0.12)
	Intermediate	3.98 (1.68)	6.10 (3.77)			
Phonological error	Advanced	1.06 (0.84)	1.34 (1.00)	7.32* (0.30)	1.13 (0.06)	0.26 (0.00)
	Intermediate	4.85 (3.98)	5.66 (5.70)			
Syntactic error	Advanced	1.57 (2.82)	2.03 (2.80)	5.28* (0.23)	0.09 (0.00)	0.60 (0.03)
	Intermediate	7.41 (8.62)	6.36 (5.01)			

* $p < .05$; ** $p < .01$

sub-categories, although the eta squared value indicates a large effect size. Interaction between task-condition and level of proficiency was not found in any of the sub-types of error-correction rate. In order to test whether particular repair types occurred statistically more frequently than others, additional repeated measure analysis of variances were conducted. These analyses indicated that there was a significant difference in the repair rate of the different types of error (Single condition: $F(2,17) = 22.26$, $p < .001$, $\eta^2 = .73$; Dual condition: $F(2,17) = 34.63$, $p < .001$, $\eta^2 = .81$). The subsequent paired sample t -tests showed that all the self-repair rates in both conditions were significantly different from each other at the $p < .001$ level, with the exception of the syntactic and phonological repair rates in the single task condition ($t = 1.20$, $p = .24$).

In our research we were also interested in the link between the frequency of different types of error and self-correction because in addition to the correction rate, this might also offer insights into the allocation of attentional resources under the two task-conditions. The correlation of the frequency of the overall frequency errors and self-repairs was $r = .54$ ($p = .01$) in the single task condition, and $r = .42$ ($p = .06$) in the dual task condition. A significant link between the frequency of lexical errors and the frequency of lexical self-repairs was also found in both conditions (Single condition: $r = .46$, $p = .04$; Dual condition: $r = .62$, $p = .004$). For syntax, however,

no relationship between errors and repairs was detected in any of the conditions (Single condition: $r = .06$, $p = .80$; Dual condition: $r = -.20$, $p = .41$), and phonological errors correlated significantly with phonological repairs only in the single task condition ($r = .51$, $p = .02$) but not in the dual task condition ($r = -.12$, $p = .62$).

The repair times of the phonological and syntactic repairs were not analyzed due to the low number of this type of repair. Although the mean error to cut-off times for the advanced group were higher than for the intermediate students, the difference was only nearing the level of significance ($p = .06$; see Table 8). However, the advanced student group was found to be faster in repairing lexical errors. No effect of task-condition was found for any of the repair times.

Discussion

The effect of dual task demands

The results concerning the effect of the dual task demands suggest that the introduction of the finger-tapping task did not cause major changes in the fluency of the utterance, the speed of error-detection and the overall frequency of self-repairs. With regard to the accuracy of the utterance, however, it was found that under dual task conditions participants made significantly more lexical errors and corrected a significantly smaller proportion of

Table 6. *The effect of proficiency and task-condition on the frequency of self-repairs.*

Frequency of repairs over amount of spoken words during block	Group	Mean single task (SD)	Mean dual task (SD)	F group (η^2)	F condition (η^2)	F group \times condition (η^2)
Total number of repairs	Advanced	1.09 (.58)	1.10 (.57)	0.09 (0.01)	0.00 (0.00)	0.00 (0.00)
	Intermediate	1.18 (.75)	1.8 (.70)			
Total number of error-repairs	Advanced	0.76 (0.43)	0.80 (0.37)	1.02 (0.05)	0.00 (0.00)	0.26 (0.01)
	Intermediate	1.03 (0.59)	0.99 (0.61)			
Lexical error-repairs	Advanced	0.49 (0.26)	0.58 (0.24)	1.78 (0.09)	2.85 (0.14)	0.02 (0.00)
	Intermediate	0.72 (0.46)	0.81 (0.53)			
Phonological error-repairs	Advanced	0.18 (0.19)	0.17 (0.18)	0.09 (0.00)	1.33 (0.06)	.89 (0.05)
	Intermediate	0.26 (0.24)	0.14 (0.12)			
Syntactic error-repairs	Advanced	0.08 (0.12)	0.04 (0.05)	0.63 (0.03)	1.09 (0.06)	0.33 (0.01)
	Intermediate	0.04 (0.07)	0.03 (0.04)			
Total number of appropriacy repairs	Advanced	0.33 (0.19)	0.29 (0.40)	1.97 (0.10)	0.00 (0.00)	.18 (0.01)
	Intermediate	0.15 (0.21)	0.19 (0.20)			
Different order repair	Advanced	0.12 (0.10)	0.10 (0.13)	1.49 (0.08)	0.07 (0.00)	1.37 (0.07)
	Intermediate	0.03 (0.07)	0.07 (0.13)			
Appropriacy of information repairs	Advanced	0.10 (0.09)	0.09 (0.16)	1.75 (0.09)	0.12 (0.00)	0.02 (0.00)
	Intermediate	0.05 (0.16)	0.03 (0.05)			
Rephrasing repairs	Advanced	0.09 (0.11)	0.09 (0.15)	0.25 (0.01)	0.04 (0.00)	0.06 (0.00)
	Intermediate	0.06 (0.08)	0.08 (0.10)			

their errors in general. Although the finger-tapping task has a potential effect on the central attention allocation system and consequently might cause interference at every level of speech processing that requires conscious response selection, in our study only two processes were influenced: lexical selection and the efficiency of monitoring were found to be affected by the introduction of the parallel task. There might be several possible

explanations of the lack of an overall condition effect on speech production. On the one hand, we might argue that speech production mechanisms in L2 require substantial attention and conscious processing even in a single task condition, and therefore, introducing a parallel task might not slow down processing to a noticeable extent. Support for this assumption might be found if we compare the speech rate in our study, which was less than two syllables

Table 7. *The effect of proficiency and task-condition on self-repair rate.*

Ratio of repairs (% of errors repaired)	Group	Mean single task (SD)	Mean dual task (SD)	F group (η^2)	F condition (η^2)	F group \times condition (η^2)
Total repair-rate	Advanced	25.64 (12.30)	19.02 (7.78)	15.77** (0.48)	16.14** (0.51)	4.53* (0.21)
	Intermediate	8.64 (6.67)	6.43 (4.24)			
Lexical repair-rate	Advanced	41.12 (11.20)	36.96 (14.80)	15.95** (0.48)	2.48 (0.12)	0.09 (0.01)
	Intermediate	21.81 (15.74)	15.63 (10.07)			
Phonological repair-rate	Advanced	20.55 (21.13)	16.85 (14.53)	7.95* (0.31)	0.51 (0.02)	0.04 (0.00)
	Intermediate	7.91 (8.13)	4.83 (5.65)			
Syntactic repair-rate	Advanced	16.25 (18.51)	6.41 (9.05)	5.54* (0.25)	3.85 (0.19)	2.82 (0.15)
	Intermediate	1.81 (3.70)	1.05 (1.58)			

* $p < .05$; ** $p < .01$ Table 8. *The effect of proficiency and task-condition on the error to repair times.*

Error to repair time	Group	Mean single task (SD)	Mean dual task (SD)	F group (η^2)	F condition (η^2)	F group \times condition (η^2)
Error to cut-off: lexical repairs	Advanced	449.67 (318.46)	423.85 (258.22)	4.24 (0.20)	1.31 (0.07)	0.01 (0.01)
	Intermediate	359.79 (176.06)	322.37 (140.97)			
Cut-off to repair: lexical repairs	Advanced	280.12 (182.21)	292.54 (245.47)	16.46** (0.49)	0.08 (0.00)	0.23 (0.01)
	Intermediate	527.13 (304.74)	512.24 (265.52)			

** $p < .01$

per second in both groups and both conditions, with the figures reported in Oomen and Postma's (2002) study for L1 Dutch speakers, who performed exactly the same tasks as the participants in this study. Their speech rate was 4.0 syllables per second in the dual and 4.1 syllables per second in the single task condition. From this comparison it is apparent that whereas in L1 dual task demands might produce a significant, albeit small, effect on speech rate, due to the slower nature of processing, in L2 speech this effect might not be detected. A similar tendency can also be observed in the case of the cut-off to repair

times, which might serve as indicators of the speed of executing the repair, because performing the correction took substantially more time for the participants of our study than for the L1 speakers in Oomen and Postma's research. On the other hand, the lack of task-condition effect on speech rate might be explained with reference to the fact that speech production involves several processing mechanisms, and working memory limitations might only interfere with specific processes such as lexical encoding and monitoring. This interference may not be noticeable in terms of the overall speed of speech processing but

only in the error-prone functioning of certain encoding and monitoring mechanisms.

Parallel task demands were not found to exert a significant influence on the overall accuracy of the students' utterances as measured by the total frequency of errors and the frequency of syntactic and phonological errors, but in the category of lexical errors, significant differences with a large effect size were obtained between the single and the dual conditions. Similar to Oomen and Postma's (2002) study, our research indicates that a parallel task might interfere with lexical selection processes, which played a particularly important role in the successful performance of this task as the syntactic encoding processes required by describing the route of the dot were relatively simple. Both the lack of interaction between task-condition and proficiency and the comparison of results with L1 speakers in Oomen and Postma's research reveal that regardless of the competence level and native-speaker status, interference with the functioning of the central executive might have important consequences for the efficiency of lexical encoding. This conclusion can also be supported by the findings of Ferreira and Pashler's (2002) study, in which it was found that lexical selection processes were subject to attentional bottleneck effects under dual task conditions. Hence, it seems that attentional control is particularly important both in L1 and L2 lexical encoding mechanisms, which involve both search processes for the appropriate lemma and control processes which inhibit the selection of other activated items (Levelt, 1989, 1999; Levelt et al., 1999).

As indicated by the results concerning error-correction rates, the introduction of the finger-tapping task caused deterioration in the overall efficiency of monitoring processes. This lends support to the findings of the previous studies both in L1 and L2 production and to the Perceptual Loop Theory (e.g., Levelt 1983; Kormos, 2000; Oomen & Postma, 2002) that monitoring is a conscious and effortful process that requires considerable attentional resources and is sensitive to influences resulting from parallel task demands. The interaction effect shown by the MANOVA analysis suggests that the monitoring behavior of the advanced and intermediate learners was differentially affected by dual task demands. Nevertheless, if we compare the rate with which the self-correction rate of the two groups of participants was reduced (from 25.64% to 19.02% in the advanced group and from 8.64% to 6.43% in the intermediate group), we can see that the drop in the correction rate is proportionately very similar (approx 25% reduction in correction rate in both groups). Therefore it seems that additional task demands cause proportionately similar reduction in monitoring efficiency both at intermediate and at advanced levels of competence.

The findings with regard to the correction rate of the sub-types of error suggest that the interference caused

by the parallel task in monitoring specific aspects of performance might be relatively small. However, this might be an artefact of the small number of participants and the relatively large standard deviations in the mean frequencies of correction rates.⁴ The results indicate that there is a large main effect of monitoring on overall error-correction rate, which might be explained by the way the overall correction rate is calculated as here the percentage of the total number of corrected errors is taken into account.

The comparison of the results concerning the mean correction rate of lexical and syntactic errors suggests that the erroneous selection of lexical items was most frequently corrected under both conditions, which indicates that learners' attention was particularly focussed on monitoring for lexical choice under both conditions. This is supported by the significant correlation between the frequency of lexical errors and lexical repairs both in the single task and in the dual task condition. The lack of significant links between the frequency of syntactic errors and syntactic self-repairs and the very low correction rate of syntactic errors reveal that the participants' attention might not have been drawn to syntactic accuracy in any of the task-conditions. This might be due to the nature of the task, which did not require the use of complex syntactic structures, as well as to students' recognition that their message was still understandable despite syntactic errors. The different patterns of correlations between the frequency of phonological errors and phonological repairs under the two conditions can indicate that monitoring behavior becomes somewhat random with regard to phonological errors under dual task conditions, whereas under single task conditions students who make more phonological errors correct a higher proportion of their slips of the tongue. This change in the systematicity of phonological error-detection might also result from attentional interference and bottleneck created by the introduction of a parallel task.

Proficiency effects on speech production and monitoring

The level of competence was found to influence nearly every aspect of performance investigated in this study. The results indicate that advanced learners spoke considerably faster than intermediate students, but no significant differences with regard to repetitions and filled pauses were found. This finding points to the fact that speech rate seems to be a more sensitive indicator of the fluency of utterances in L2, as it gives an indication of the overall speed with which speech production processes

⁴ The kurtosis values of the sub-types of error-correction rate were, however, all within the ± 2 range, which suggests normal distribution.

are carried out (see also Kormos & Dénes, 2004, for a similar argument), than the frequency of hesitation phenomena because pausing might be influenced by individual speaking style.

Advanced learners were also found to make significantly fewer errors in every category than intermediate learners, and the differences in the frequency of overall errors between the groups is striking (intermediate learners made nearly four times more errors than advanced speakers). When findings on fluency and accuracy are juxtaposed, it can be concluded that the speech encoding mechanisms of intermediate learners work with considerably less efficiency than those of the advanced participants.

Similar to other studies in the field of self-repairs (e.g., Kormos, 2000; van Hest, 1996), the results of our research also show that the level of L2 competence does not seem to influence the frequency with which students repair themselves. However, the rate of corrected errors was found to be substantially different in the case of the two groups of participants. Advanced speakers repaired a significantly higher proportion of their lexical, syntactic and phonological errors than the intermediate learners, which suggests that, at higher levels of L2 proficiency, monitoring processes function more efficiently. The figures for cut-off to repair times also indicate that advanced learners could perform the correction of noticed errors considerably faster than the intermediate participants, which can be explained with reference to the higher level of automaticity in the encoding processes at higher levels of L2 competence.

Conclusion

In this study we examined how the introduction of a parallel task influences L2 speech encoding mechanisms and monitoring processes, and how the level of proficiency impacts on the efficiency and accuracy of L2 performance under single and dual task conditions. The results indicate that imposing dual task demands has a negative effect on the accuracy of lexical selection and the efficiency of monitoring processes. We have argued that this can be explained with reference to the fact that the introduction of a parallel task that requires kinaesthetic response interferes with the verbal production processes and creates a bottleneck effect in the functioning of the central executive, which is responsible for the management of attentional resources. We have pointed out that the description task employed in this study required that students focus on lexical encoding, and consequently it is at this stage where interference from a parallel task becomes the most noticeable. Our findings in this regard are in line with research investigating attentional bottleneck effects using single word production tasks in

L1 production (Ferreira & Paschler, 2002) and show that lexical selection processes are sensitive to interference from parallel tasks in simple sentence production tasks as well.

The findings with regard to monitoring confirm the results of previous studies in the L1 field which have shown that self-repair processes are sensitive to attentional bottleneck effects. The results indicate that, under dual task conditions, the overall efficiency of monitoring in terms of the percentage of corrected errors decreased significantly in the case of L2 speakers as well. Parallel task demands were found to have a differential effect on intermediate and advanced speakers as the monitoring efficiency of the advanced group decreased more significantly than in the case of the intermediate learner group. The different effect of dual task demands on advanced learners might be caused by different decisions on prioritizing attention in monitoring in the two groups of learners.

Our research was conducted with a relatively small number of participants and in a foreign language context, where the target language is not widely used outside classroom contexts. Therefore, further research with a larger sample size and in other contexts would be needed to extend the generalizability of our findings. Furthermore, we used a relatively simple description task so as to ensure that students would be able to perform an additional parallel activity. It would also be interesting to examine how dual task demands affect performance on more complex speaking tasks and on tasks in which the focus varies between linguistic form and meaning.

Appendix. Examples of performances

Advanced participant

It goes up over a straight line to a nurse. After that it goes right over an upper curved line to a bike. After that it goes right over a straight line to a church. Then it goes down over a right curved line to a skirt.

Intermediate participant

It's going left over the lower curved line to the door. It's going right over the upper curved line to a rope. It's going down over the left curved line to the leg. It's going straight down to the wing.

References

- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417–423.
- Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4, 829–839.

- Baddeley, A. D., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *Quarterly Journal of Experimental Psychology*, 51 (A), 819–852.
- Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, 39, 173–194.
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn.). Hillsdale, NJ: Lawrence Erlbaum.
- Cook, A. E., & Meyer, A. S. (2008). Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 886–899.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- de Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning*, 46, 529–555.
- Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.
- Ellis, R. (2005). *Planning and task-performance in a second language*. Amsterdam: John Benjamins.
- Evans, F. J. (1978). Monitoring attention deployment by random number generation: Index to measure subjective randomness. *Bulletin of the Psychonomic Society*, 12 (1), 35–38.
- Fathman, A. K. (1980). Repetition and correction as an indication of speech planning and execution processes among second language learners. In H. W. Dechert & M. Raupach (eds.), *Towards a cross linguistic assessment of speech production*, pp. 77–85. Frankfurt: Peter D. Lang.
- Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 1187–1199.
- Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, 3, 410–419.
- Gilbert, R. (2006). The simultaneous manipulation along the planning time and +/- Here-and-Now dimensions: Effects on oral L2 production. In M. P. García Mayo (ed.), *Investigating tasks in formal language learning*, pp. 44–68. Clevedon: Multilingual Matters.
- Gilbert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching*, 45, 215–240.
- Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject–verb agreement. *Language and Cognitive Processes*, 21, 181–204.
- Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the Perceptual Loop Theory. *Cognitive Psychology*, 42, 113–157.
- Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59, 91–117.
- Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale for the output hypothesis. *Applied Linguistics*, 24, 168–196.
- Jahanshahi, M., Saleem, T. T., Ho, A. K., Dirnberger, G., & Fuller, R. R. (2006). Random number generation as an index of controlled processing. *Neuropsychology*, 20, 391–399.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology General*, 132, 47–70.
- Kormos, J. (2000). The role of attention in monitoring second language speech production. *Language Learning*, 50, 343–384.
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Kormos, J. (2011). Speech production and the Cognition Hypothesis. In P. Robinson (ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance*, pp. 39–60. Amsterdam: John Benjamins.
- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32, 146–164.
- Lennon, P. (1984). Retelling a story in English as a second language. In H. W. Dechert, D. Möhle & M. Raupach (eds.), *Second language productions*, pp. 50–68. Tübingen: Günter Narr.
- Lennon, P. (1991). Error: Some problems of definition, identification and distinction. *Applied Linguistics*, 12, 180–195.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 33, 41–103.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1999). Language production: A blueprint of the speaker. In C. Brown & P. Hagoort (eds.), *Neurocognition of language*, pp. 83–122. Oxford: Oxford University Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Science*, 22, 1–38.
- Martin, N., Weisberg, R. W., & Saffran, E. M. (1989). Variables influencing the occurrence of naming errors: Implications for models of lexical retrieval. *Journal of Memory and Language*, 28, 462–485.
- Navon, D. (1984). Resources – A theoretical soup stone? *Psychological Review*, 91, 216–234.
- Navon, D. (1989). The importance of being visible: On the role of attention in a mind viewed as an anarchic intelligence system; basic tenets. *European Journal of Cognitive Psychology*, 1, 191–213.
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30, 555–578.

- Oomen, C. C. E., & Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research*, 30, 163–184.
- Oomen, C. C. E., & Postma, A. (2002). Limitations in processing resources and speech monitoring. *Language and Cognitive Processes*, 17, 163–184.
- Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, 21, 109–148.
- Pashler, H., & Johnston, J. C. (1998). Attentional limitation in dual-task performance. In H. Pashler (ed.), *Attention*, pp. 155–189. Hove: Psychology Press.
- Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, 77, 97–131.
- Postma, A., & Kolk, H. (1992). The effects of noise masking and required accuracy on speech errors disfluencies and self-repairs. *Journal of Speech and Hearing Research*, 35, 537–544.
- Postma, A., & Kolk, H. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research*, 36, 472–487.
- Poullisse, N. (1993). Slips of the tongue and their correction in L2 learner speech: Metalinguistic awareness and second language acquisition. Presented at the 10th World Congress of Applied Linguistics, Amsterdam.
- Poullisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics*, 15, 36–57.
- Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22, 27–57.
- Robinson, P. (2003). The Cognition Hypothesis of adult, task-based language learning. *Second Language Studies*, 21, 45–107.
- Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics*, 43, 1–32.
- Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching*, 45, 237–257.
- Roelofs, A. (2008). Attention, gaze shifting, and dual-task interference from phonological encoding in spoken word planning. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 1580–1598.
- Sanders, A. F. (1998). *Elements of human performance: Reaction processes and attention in human skill*. Mahwah, NJ: Lawrence Erlbaum.
- Sarno, K., & Wickens, C. D. (1995). The role of multiple resources in predicting time-sharing efficiency: An evaluation of three workload models in a multiple task setting. *International Journal of Aviation Psychology*, 5, 107–130.
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, 125, 4–27.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30, 510–532.
- Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (ed.), *Cognition and second language instruction*, pp. 183–205. Cambridge: Cambridge University Press.
- Towse, J. N., & Neil, D. (1998). Analyzing human random generation behavior: A review of methods used and a computer program for describing performance. *Behavior Research Methods, Instruments, & Computers*, 30, 583–591.
- van Hest, E. (1996). *Self-repair in L1 and L2 production*. Tilburg: Tilburg University Press.
- West, M. (1953). *A general service list of English words*. London: Longman.
- Wickens, C. D. (1980). The structure of attentional resources. In R. S. Nickerson (ed.), *Attention and performance* (vol. VIII), pp. 239–257. Hillsdale, NJ: Lawrence Erlbaum.
- Wickens, C. D. (2007). Attention to the second language. *International Review of Applied Linguistics*, 45, 177–191.
- Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 oral production. *Applied Linguistics*, 24, 1–27.