Linguistic dual-tasking:

Understanding temporal overlap between production and comprehension

ISBN: 978-94-92910-02-8

Cover picture: Emma Ward and Amie Fairs

Printed and bound by Ipskamp Drukkers b.v.

Linguistic dual-tasking:
Understanding temporal overlap between production and comprehension

Proefschrift
ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van de decanen
in het openbaar te verdedigen op vrijdag 7 juni 2019
om 10.30 uur precies

door
Amie Leigh Fairs
geboren op 19 juni 1990
te Portsmouth, Verenigd Koninkrijk

Promotor:

Prof. dr. Antje S. Meyer

Copromotoren:

Dr. Sara Bögels

Dr. Alexis Hervais-Adelman (Universität Zürich, Zwitserland)

Manuscriptcommissie:

Prof. dr. Peter Hagoort

Prof. dr. Marina Laganaro (Université de Genève, Zwitserland)

Dr. Vitória Piai

This research was supported by the The Max Planck Society for the Advancement of Science, Munich, Germany.

For Florent
This thesis would not have been finished without you

Table of Contents

1 General introduction

Every day, people speak or listen while they are doing something else. For instance, people talk when cooking dinner, or listen to the radio when driving. Despite the ease with which we do these two tasks, a lot of evidence suggests that carrying out two demanding tasks at once hinders at least one of the tasks (for review, see Fischer & Plessow, 2015; Pashler, 1994).

We already know that this happens with a language task and a non-language task, like speaking when driving (Strayer & Johnston, 2001). However, what happens if the two tasks are language-related? For example, how does a person plan what they want to say while they listen to someone else? Can these two tasks – speech planning and comprehension – be carried out together? This thesis seeks to investigate this question, specifically investigating constraints on simultaneous word production and comprehension with simple linguistic tasks.

Despite conversation being one of the most common ways people communicate (Levinson, 2016), little work has investigated how people manage to have such smooth conversations. Times between turns are about 200ms in many languages around the world (Heldner & Edlund, 2010; Stivers et al., 2009), yet estimates on how long it takes to produce a word from thinking of the word to uttering it are at least 600ms (Indefrey & Levelt, 2004; Indefrey, 2011). This suggests that people start planning speech while still comprehending incoming speech from their interlocutor. Experimental evidence also suggests there is overlap between planning speech and comprehending speech, with some studies finding that speech planning begins during the preceding turn (Bögels, Magyari, & Levinson, 2015; Bögels, Casillas, & Levinson, 2018; Barthel, Sauppe, Levinson, & Meyer, 2016), and others finding that it begins late in the preceding turn, but still in overlap with incoming speech (Boiteau, Malone, Peters, & Almor, 2014; Sjerps & Meyer, 2015). Recent theories of conversation also model overlap in these two tasks (Levinson & Torreira, 2015). Taken together, there is evidence that at a broad level speech planning processes and comprehension processes can occur in overlap.

Despite the evidence that shows broad overlap of planning and comprehension processes, how these processes are coordinated at a micro level is unclear. Specific processes in planning and comprehension could be carried out in parallel, or there could be rapid switching between planning processes and comprehension processes. It is also possible that there is a combination of these two strategies, such that some planning and comprehension processes can be performed in parallel and some serially. These kinds of strategies are investigated in this thesis, where participants are forced to carry out planning and comprehension processes in a dual-task, to determine which processes can be carried out in parallel and which serially.

A separate strand of psychological research has modelled how people carry out two tasks at the same time. Typically, the two tasks are quite simple. There are three main theories of dual-tasking: the structural bottleneck theory (Pashler, 1984, 1994, 1998), the capacity sharing theory (Kahneman, 1973; Navon & Miller, 2002; Tombu & Jolicœur, 2003) , and the strategic bottleneck theory (D. E. Meyer & Kieras, 1997). The structural bottleneck and capacity sharing theories are described in more detail in Chapter 2. All theories describe tasks as being composed of three separate stages: a perception stage, a central task processing stage, and a response execution stage. The perception and execution stages can run in parallel with any other stage. However, the central stage requires capacity[1] and as such cannot be easily combined with other stages. Under the structural bottleneck theory, there is a bottleneck at this central stage, so only one task can occupy this stage at a time. This means that central task processes are only carried out on one task at a time, and hence task processing is serial. Under the capacity sharing theory, both tasks share capacity for task processing. Because capacity is limited, the central stages of two tasks can be carried out in parallel but with less capacity assigned to each task, and so processing in both tasks will be slower. Under the strategic bottleneck theory, any stage can run alongside any other stage. However, due to task demands or for strategic reasons, tasks may be scheduled in a serial way.

In order to test how people carry out two tasks simultaneously, and whether the central stages of tasks can run in parallel, much work makes use of the psychological refractory period (PRP) paradigm (Telford, 1931). This paradigm is also used in Chapters 3, 4, 5 and 6 in this thesis. Under this paradigm, two different tasks are given to participants. The task 1 and task 2 stimuli are presented with

[1] I define capacity as a form of attention (e.g., Roelofs & Piai, 2011), but make no claims about what kind of attention is used. The dual-tasking literature is often vague in what 'capacity' actually is.

different stimulus onset asynchronies (SOAs) to force overlap between the tasks, or to allow the tasks to be executed serially. Participants are told to respond as quickly as possible to both tasks, and to respond to task 1 before task 2. This paradigm tests whether there is interference between tasks, and which aspects of the tasks are part of the 'central' stage of processing.

The dual-tasking theories described above relate to language processing because some research has investigated whether different language processes require capacity, using the PRP paradigm. The evidence shows that in production, most stages require capacity. This includes conceptual to lemma information flow (Mädebach, Jescheniak, Oppermann, & Schriefers, 2011), lemma selection (V. S. Ferreira & Pashler, 2002; Piai, Roelofs, & Schriefers, 2014), word form activation (V. S. Ferreira & Pashler, 2002), and phonemic encoding (Cook & Meyer, 2008). In comprehension, phonemic encoding does not require capacity (Gaskell, Quinlan, Tamminen, & Cleland, 2008), but word form activation and semantic activation do (Cleland, Tamminen, Quinlan, & Gaskell, 2012; Relander, Rämä, & Kujala, 2009). Therefore, I ask whether all production stages and some comprehension stages would be part of the 'central' stages described in dual-tasking theories. This question is explored more in Chapter 2.

If these processes require capacity, they should be unable to run in parallel. Thus, dual-tasking models predict that production and comprehension processes should either run serially, or run in parallel but in a hindered way. This would suggest that speaking and listening in conversation would not overlap. Yet, as described above, there is evidence that they do so. Thus, I ask how production and comprehension processes work together. If the two tasks are language-related, do I find the same patterns of interference that are found in a typical dual-task (not involving two language tasks)? These questions are experimentally tested in Chapters 3, 4 and 5.

Research into dual-tasking has already found two factors which can influence dual-task performance: practice effects and cross-talk effects. Both of these factors may play a role when coordinating speech planning and comprehension processes. Practice effects refer to the fact that when people have a lot of practice with two tasks, coordination of the tasks becomes streamlined. This results in extremely reduced, or even eliminated, dual-task interference (Allen, Ruthruff, Elicker, & Lien, 2009; Kamienkowski, Pashler, Dehaene, & Sigman, 2011; Logan, Miller, & Strayer, 2011; Strobach, Frensch, Müller, & Schubert, 2012; Strobach, Liepelt, Pashler, Frensch, & Schubert, 2013; Strobach, Salminen, Karbach, & Schubert, 2014; Van Selst, Ruthruff, & Johnston, 1999). Practice effects can even transfer to other tasks (e.g., Liepelt, Strobach, Frensch, & Schubert, 2011; Strobach, Frensch, & Schubert, 2012). Practice effects may play a role in language because every day people speak and listen to thousands of words. Carrying out production and comprehension in parallel every day would give people a lot of practice with coordinating these tasks. Because of this extensive practice, carrying out two linguistic tasks together may be much easier than carrying out other kinds of dual tasks.

The other factor is cross-talk. Cross-talk arises when two tasks are similar and influence one another (Alards-Tomalin, Walker, Nepon, & Leboe-McGowan, 2017; Fargier & Laganaro, 2016; Hommel, 1998; Logan & Schulkind, 2000; Navon & Miller, 1987; Wickens, 2008), and hence 'talk' to each other. Because stages in language production and comprehension may be shared, these shared representations may influence one another. This would result in the representations for production and comprehension affecting one another, which may cause problems with speaking or listening. Thus, we may expect that the two representations are affected, which would lead to greater interference between the tasks. An alternate prediction is that the cognitive system tries to keep these representations separated so that they cannot 'talk' to each other, which may result in a more serial processing strategy such that planning and comprehension are sequential processes.

In summary, we see that people are able to plan and comprehend speech at the same time. However, research investigating dual-tasking would suggest that there are limitations on how people are able to coordinate language production and comprehension at a micro level, where there may be serial or parallel processing of planning and comprehension. Thus, I ask: what are the constraints on how people produce and comprehend language in a dual-task?

In Chapter 2 I present a review of the research field, drawing together ideas from dual-tasking theories and language processing theories to create a working model of how production and comprehension could be coordinated in parallel. I discuss evidence of capacity constraints in psycholinguistic processing, and how these inform coordinated production and comprehension. Many of the ideas touched on in this introduction are explored in greater depth in this chapter. This review also generates predictions for constraints on the concurrent processing of production and comprehension, some of which are tested in later chapters.

In Chapter 3 I tested whether carrying out two linguistic tasks concurrently results in greater (because of cross-talk), lesser (because of practice), or the same level of interference as carrying out one linguistic and one non-linguistic task. In

Experiment 1, participants carried out two dual-tasks consisting of picture naming and syllable or tone identification. I tested whether there would be additional linguistic interference on top of normal dual-task interference in the picture naming + syllable identification condition. In Experiment 2, I tested whether any linguistic effects were truly linguistic or based on acoustic complexity differences between syllables and tones. In this experiment, I kept the auditory sound the same but manipulated the knowledge of whether the sound was linguistic or not. Participants again carried out a picture naming + sound identification task, with the linguistic instruction a between-participant variable.

In Chapter 4 I again tested the linguistic interference hypotheses while keeping acoustic complexity constant, but in a within-participant design. Participants listened to acoustically matched auditory stimuli – synthesised vowels as the linguistic stimuli and musical rain versions of these vowels as the non-linguistic stimuli (Uppenkamp, Johnsrude, Norris, Marslen-Wilson, & Patterson, 2006). In two sessions participants combined identification of these sounds with a non-linguistic size judgement task or a linguistic picture naming task. I aimed to determine whether the linguistic interference effect was still found when acoustic complexity was kept constant.

In Chapter 5, I investigated how a different language process, lexical access, is affected in a dual-task. I tested whether lexical access could be carried out in parallel with another linguistic task, and whether the type of task affected dual-task interference. Participants carried out two different types of dual-task. One was a typical dual-tasking experiment where responses were given to both stimuli, and the other was a task choice experiment (Besner & Care, 2003), where a decision made on the auditory stimulus (task 1) determined what response to give to the visual stimulus (task 2). Participants heard the same syllables and tones as in Chapter 3, and carried out picture naming with visual distractors. I measured semantic interference in picture naming latencies to determine whether lexical access was carried out at the same time as the other task. I asked whether the semantic interference effect would be affected in any way by carrying out a concurrent linguistic task compared to a non-linguistic task.

In Chapter 6 I tested linguistic dual-tasking in an MEG experiment. Participants carried out the dual-task of syllable identification and picture naming while the magnetic activity in their brain was measured. I wanted to determine the cognitive dynamics of dual-tasking with two linguistic tasks compared to carrying out only one of the tasks, inspired by recent work by Marti, King, and Dehaene (2015). The evidence from that study suggests that the two prominent dual-tasking theories (the structural bottleneck theory and the capacity sharing theory) cannot account for the cognitive behaviour in a dual-task. As the behavioural evidence from previous chapters suggests that there is some interference between two linguistic tasks, I wished to discover the cognitive dynamics of this type of dual-task, and whether the patterns were similar to Marti et al. (2015). I used a machine learning technique, temporally-generalised multivariate pattern analysis (MVPA), generalised across condition, and applied this to the data. This technique meant that I was able to train a classifier to discriminate between no task and syllable identification, another classifier to discriminate between no task and picture naming, and then I tested both of these classifiers on their ability to determine activity in the dual-task condition. I asked whether classifiers were able to decode activity in the dual-task and how this activity pattern was different in comparison to when the tasks were carried out as single tasks.

In the final chapter of this thesis, I bring together all results and discuss how these findings relate to the wider literature of coordinating production and comprehension. I relate the findings to the model proposed in Chapter 2, and to theories of dual-tasking. I also discuss methodological implications arising from the thesis and suggest future research directions.

2 Towards a working model of linguistic dual-tasking

Abstract

During conversation, interlocutors converse quickly with short gaps between turns. Evidence from corpus and experimental studies suggests that people are able to plan their response while they are listening to their interlocutor, and thus coordinate planning and comprehension with close temporal overlap. However, a different vein of linguistic research shows that multiple processes in word production and comprehension require processing capacity, and evidence from the wider dual-tasking literature indicates that capacity-demanding tasks are difficult to carry out in overlap. In this review, I aim to bring together the scarce evidence on the temporal coordination of speech planning and comprehension, theoretical contributions from the dual-task literature, and evidence of capacity demands in word production and comprehension to discuss a working model of linguistic dual-tasking. Constraints on how word production and comprehension are discussed with reference to this working model, and predictions the model generates, are presented.

Conversation is the ecological niche of language use (Levinson & Torreira, 2015). During conversation, interlocutors plan what to say to their partner, while listening to their partner's speech stream and mapping these sounds to meaning. Despite the fact that language is used for this dynamic purpose, very little research has investigated the coordination and temporal overlap of both planning speech (language production) and understanding speech (language comprehension).

Here, I review and bring together relevant language and dual-tasking theories and evidence to create a working model describing how word production and comprehension could be coordinated. I begin with research which has investigated the temporal overlap of language production and comprehension in conversation. I then introduce a theory of concurrent word production and comprehension, and review dual-tasking theories relevant to temporal overlap of tasks. Psycholinguistic research investigating capacity allocation to language processes are discussed, and I describe how linguistic dual-tasking may be different from non-linguistic dual-tasking. I end by expanding the production and comprehension model, and generate predictions from this model to help us better understand how production and comprehension may be coordinated. I focus throughout on the possible constraints on simultaneous production and comprehension, and how to align dual-tasking and psycholinguistic theories.

Temporal overlap: When do people start planning speech?

The temporal overlap of production and comprehension arose as an area of interest based on reconciling findings concerning how long word planning takes versus how long gaps between turns are. Evidence suggests that it takes approximately 600ms from conceptual activation to word articulation (Indefrey & Levelt, 2004; Indefrey, 2011)[1]. However, the modal timing between turns is around 200ms (Heldner & Edlund, 2010; Stivers et al., 2009)[2]. This suggests that planning must overlap with comprehension, to ensure that the speech plan can be launched around 200ms after the end of a turn (see Levinson, 2016).

[1]The studies which have generated these timings are almost all picture naming experiments, where single images or simple events are described. How well the timings generalise out to planning in conversation, where there is often a rich contextual history and setting to the conversation, and multi-modal input, is unclear.

[2]Studies which have investigated between-turn timing have focused on responses to yes/no questions, with speech and non-speech responses and body language classed as turn-initiating behaviour (Stivers et al., 2009). These findings may generalise differently to complex conversation.

Recent studies have addressed the question of when in a turn people start to plan speech. There are two possibilities: as early in the turn as possible ('early planning'), or at a later point in the turn ('late planning'). Both planning types result in overlap between production and comprehension, but this overlap is greater with early planning. Early planning is illustrated and discussed in detail by Levinson and Torreira (2015).

To test whether planning begins as early as possible, Bögels et al. (2015) manipulated in an EEG study when in a turn speakers could begin planning. Participants were asked quiz-style questions containing critical words, from which the answer would be known. These words were at early or late points in the questions. Bögels et al. (2015) tested whether people would begin planning at this early point.

Bögels et al. (2015) found a large positivity in the EEG signal, which occurred approximately 500ms after the onset of the critical word. Importantly, the positivities were similar in the early and late planning conditions, but they were shifted in time, demonstrating that participants began planning early when they could. The positivity was localised to the middle and superior temporal gyrus, and the inferior frontal gyrus. These areas have been shown to be part of the language production network (Indefrey & Levelt, 2004). Participants also responded earlier in the early planning condition (M = 640ms) compared to the late planning condition (M = 950ms). This study provides evidence that people can plan in overlap with comprehending. Similar behavioural results have also been found by Barthel et al. (2016).

There is also evidence of late planning. Boiteau et al. (2014) had participants engage in normal conversation while tracking a ball on screen using a mouse. Boiteau and colleagues found that towards the end of natural speech turns, participants were less accurate at tracking the ball on screen. This was especially true with a fast-moving ball. Lower accuracy occurred at the end of a listening turn, indicating that in this case participants planned their upcoming speech fairly late in the turn. One of the strengths of this study is that participants were engaged in normal conversation, lending high ecological validity. A similar result was also found by Sjerps and Meyer (2015) in a more controlled experimental paradigm. These studies demonstrate that speech planning does not always begin as early as possible.

If planning and comprehension overlap, does comprehension suffer? Bögels et al. (2018) tested this question in an EEG study where participants described pictures in response to a question. Similarly to Bögels et al. (2015), answers to questions could be planned early or late. The partner's speech also contained an expected or unexpected final noun. Bögels et al. (2018) investigated whether the N400 effect was reduced in the early planning compared to the late planning condition, hypothesising that comprehension may suffer if attention is devoted to planning (resulting in a smaller N400 effect). They found that in participants who planned early, as indexed by an early positivity and early speech onset latencies, there was a reduced N400 compared to participants who planned later. This suggests that resources used for comprehension were reduced, as they were allocated to planning.

Together, these studies indicate flexibility in when people begin planning, and suggest planning and comprehension may share resources. However, these studies have not looked at the fine-grained planning of speech at different levels of production and comprehension. How these levels may be coordinated in the dual-task of planning and comprehension is unknown.

Model of word production and comprehension

I take the WEAVER++/ARC model of production and comprehension as a working model (Roelofs, 2014). The model accounts for word production and comprehension in normal speech use. This model builds on previous versions of the WEAVER++ model (Roelofs, 1992, 1997, 2003, 2008b, 2008a), which is also assumed by Indefrey and Levelt (2004). It captures many processes assumed by other theories of production (Dell, 1986; W. J. M. Levelt, Roelofs, & Meyer, 1999) and comprehension (McClelland & Elman, 1986; Gaskell & Marslen-Wilson, 1997, 2002). The production process follows a chain of lexical concepts, lemmas, output forms, and output phonemes. The comprehension process follows a chain of input phonemes, input forms, lemmas and lexical concepts (see Figure 2.1). Lexical concepts and lemmas are shared between production and comprehension, but word forms and phonemes are distinct. These separate form and phoneme levels are linked.

Shared and distinct representations in production and comprehension

At the lexical concept and lemma selection levels studies suggest that representations are shared between production and comprehension (Cutting, 1997; Humphreys, Mirković, & Gennari, 2016; Van Assche, Duyck, & Gollan, 2016). This

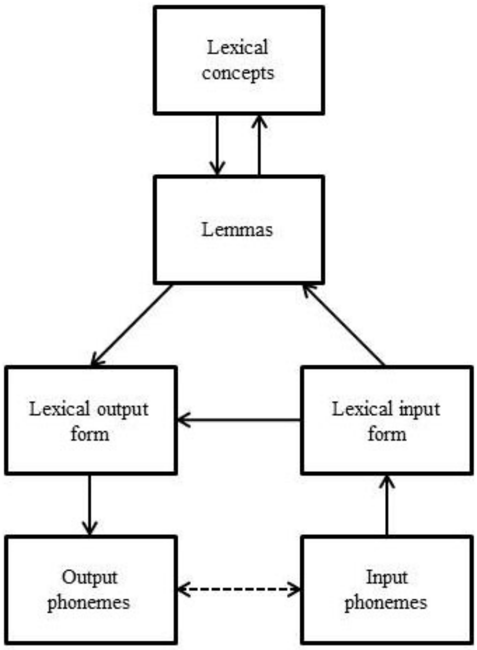

Figure 2.1: Simplified version of WEAVER++/ARC model. Note that lexical concepts and lemmas are shared between production and comprehension, whereas there are separate input and output word forms and phonemes. For the full model see Roelofs, 2014.

has also been found at the neural level (Menenti, Gierhan, Segaert, & Hagoort, 2011; Silbert, Honey, Simony, Poeppel, & Hasson, 2014), where similar brain regions have been found to be activated for semantic processing in production and comprehension. This suggests that at a semantic level the same representations are used in production and comprehension.

A large body of work has also investigated this question using the picture-word interference paradigm. Here, participants name a picture in the context of an auditory or visually presented distractor word (Glaser & Düngelhoff, 1984; Schriefers, Meyer, & Levelt, 1990). Most research has found that when naming a picture with a distractor which is categorically related (for example, a picture of a DOG with the distractor word CAT), naming RTs are slower compared to an unrelated distractor (distractor TABLE; e.g., Damian & Martin, 1999; Schriefers et al., 1990). Explanations of how this effect arises are that at the level of lemma selection, the distractor word activates its conceptual and lemma information, and this competes with the picture name (Roelofs, 1992, 2003). This suggests that these levels are shared between production and comprehension. Note however that the picture name and distractor must be presented almost simultaneously for this effect to occur; longer intervals lead to no effect.

In contrast, phonological representations are argued to be distinct from one another, though still loosely coupled (Kittredge & Dell, 2016; Mitterer & Ernestus, 2008; Shallice, McLeod, & Lewis, 1985). Neural evidence has also suggested that lower-level articulatory regions are distinct (Silbert et al., 2014). Evidence for distinct phonological and phonetic forms comes from developmental studies (e.g., Vihman, 2013, 2017; Werker & Tees, 1984) and second language learners (e.g., Best & Tyler, 2007; Van Leussen & Escudero, 2015), showing infants and learners can often perceive phonological and phonetic differences before producing them. However, there is also evidence of motor cortex involvement in both production and comprehension (e.g., Scott, McGettigan, & Eisner, 2009), which suggests that articulatory and/or phonetic representations in production and comprehension are not entirely separable.

Theories intertwining production and comprehension (Hickok & Poeppel, 2004, 2007; Pickering & Garrod, 2004, 2013) also assume parity of representations. Other theories posit a strong relationship between production and comprehension (MacDonald, 2013), and indeed, there is evidence suggesting that the production system is used during comprehension (Martin, Branzi, & Bar, 2018). While the majority of research has shown comprehension affecting production, there is also work demonstrating that production can affect comprehension (Hopman & MacDonald, 2018). Altogether, the experimental and theoretical evidence support shared lexical concept and lemma representations, and distinct form and output representations.

Theories of dual-tasking

A large body of literature has investigated how people dual-task, typically using the psychological refractory period (PRP) paradigm (Telford, 1931). In this paradigm, participants give speeded responses to two tasks in a specified order. Often, the tasks are separated by varying stimulus onset asynchronies (SOAs) allowing different amounts of task overlap. This allows researchers to test which aspects of a task are affected when dual-tasking. I now discuss theories of dual-tasking and the claims they make regarding task processing.

The literature is dominated by two main theories: the response selection bottleneck theory (Pashler, 1994), and the capacity-sharing theory (Navon & Miller, 2002; Tombu & Jolicœur, 2003). These theories are discussed below.

Response Selection Bottleneck theory (RSB)

Pashler (1994; see also Pashler, 1984, 1998) describes task processing in three stages: the perceptual stage, where the task stimulus is perceived; the central stage (also called 'response selection'; note this stage may be further subdivided into response activation and response selection; Hommel, 1998; Lien & Proctor, 2002), where the task is processed; and the execution stage, where the response is given.

Of these stages, only the central stage is subject to a bottleneck, meaning that central processing can only proceed for one task at a time. Perception and execution can proceed in parallel with any other stage. This is visualised in Figure 2.2.

The RSB theory makes three predictions. Firstly, central processing is serial, meaning that task 2 processing will be delayed until central processing in task 1 is finished. The time for task 2 processing is extended by the amount of time the task waits for the bottleneck (until task 1 is finished with central processing).

Secondly, because task 1 proceeds unhindered, there should be no SOA effects in task 1. In fact, task 1 processing should proceed as it does when carrying out task 1 alone. Only task 2 should show SOA effects.

Thirdly, any effects at the central or execution stages of task 2 should remain detectable, but an effect at the perception stage will not be measured with task overlap. This is due to the 'locus of slack' logic: because the perceptual stage of any task can run alongside another stage, the perceptual stage of task 2 can run in parallel with task 1 central processes. Task 2 central processing must wait until task 1 central processing is finished. Thus, any perceptual-level effect is not detected because extended perceptual processing is carried out in the 'slack' (greyed area of Figure 2.2) while task 2 is waiting for bottleneck access.

Capacity-Sharing theory

In contrast to the RSB theory, capacity-sharing theories (Kahneman, 1973; Navon & Miller, 1987, 2002; Tombu & Jolicœur, 2003; Wickens, Vidulich, & Sandry-Garza, 1984; Wickens, 2008) argue that there is no central bottleneck. Rather, a finite amount of capacity is shared across parallel central processing in task 1 and task 2. If all capacity is assigned to task 1, the same predictions as the RSB theory hold. However, sharing capacity results in extended central stages (see Figure 2.2). Perception and execution of a task are not subject to capacity limitations.

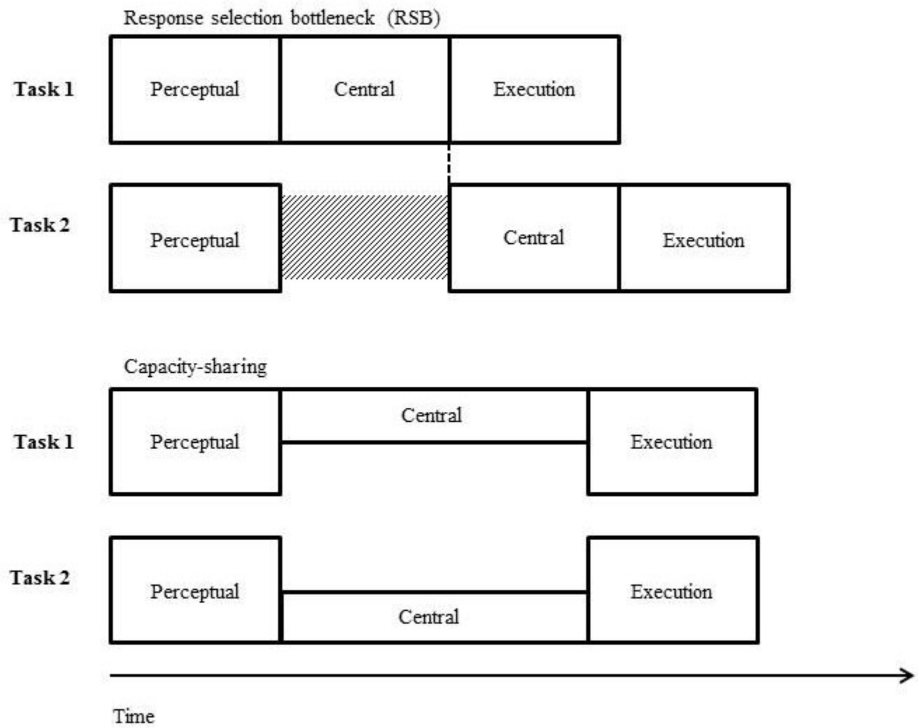

Figure 2.2: Depiction of dual-task processing of the three stages under the RSB theory (top panel) and the capacity-sharing theory (bottom panel). See main text for explanation.

One critically different prediction made by the capacity-sharing theory compared to the RSB theory is that task 1 will be affected by an SOA manipulation if capacity is shared between tasks. Because the central stage of task 1 does not proceed at full capacity, its duration will be extended. Therefore, task 1 and task 2 are affected by overlap of the tasks, not only task 2.

Capacity-sharing theories differ as to whether they posit one form of general capacity (Kahneman, 1973; Tombu & Jolicœur, 2003) or multiple sources (Navon & Miller, 2002; Wickens, 2008). One form of general capacity could be used by multiple tasks, but is limited. Multiple sources are also resource-limited, but they are used by different tasks or processes (see Wickens, 2008). If two tasks do not share resources, they could have perfect time-sharing. Many tasks do share resources however, so realistically the interpretation is that the smaller the overlap in resources, the better the time-sharing.

Capacity-sharing theories also make explicit predictions about cross-talk, referring to when the tasks interfere with one another (see e.g., Hommel, 1998). Under the RSB theory cross-talk is not possible (Navon & Miller, 2002) because of the serial nature of processing. In contrast, under the capacity-sharing theory, tasks which share resources will affect one another because the task representations are simultaneously active. Research investigating cross-talk has mostly focused on similarities in input or output modalities (Alards-Tomalin et al., 2017; Eder, Pfister, Dignath, & Hommel, 2017; Hommel, 1998; Janczyk, Pfister, Hommel, & Kunde, 2014; Janczyk, 2016; Lehle & Hübner, 2009; Lien, Ruthruff, Hsieh, & Yu, 2007; Miller, 2006; Röttger & Haider, 2017; Stelzel & Schubert, 2011). Little research has tested similarity in the central stage. I believe this is for two reasons. Firstly, the central stage of task processing tends to be considered an amodal system. Central representations are separated from perception and execution (Hazeltine, Ruthruff, & Remington, 2006). Secondly, most research on dual-tasking uses simple stimulus-response tasks, whereas a language task appears to have many processes involved in the central stage (this is discussed further later). It seems that maybe less thought has been given to complex central stages and how they would interact between tasks. Therefore, while central cross-talk is predicted, the evidence for it is lacking.

However, the difference between multitasking and multiplexing has been made (Feng, Schwemmer, Gershman, & Cohen, 2014). Multiplexing refers to holding multiple representations in mind, whereas multitasking refers to the processing pathways used to act on those representations in parallel. Feng and colleagues argue that keeping multiple representations active simultaneously results in cross-talk between these representations. Their computational results suggest multiplexing has a much stronger effect on dual-tasking behaviour than multitasking. This work would suggest cross-talk in representations, not in processes.

Extensions of the theories

Sigman and Dehaene (2006) extended the RSB model by adding top down control into the task scheduling system. Their results (Sigman & Dehaene, 2005, 2006, 2008) suggest that top down control, which may reflect central executive processing time, contributes to dual-task costs. Additionally, active processes of planning task actions and disengaging from a task contribute to dual-task costs. In their model, these additional processes capture response times and distributions in dual-task behaviour.

Marti et al. (2015) additionally found, in an MEG study with a novel analysis technique, that neither the RSB theory nor the capacity-sharing theory could fully account for their data. In the first 200-350ms of task processing task 1 and task 2 were processed in parallel. After 350ms task 1 processes were shortened and task 2 processes slightly hindered, and after 450-500ms task 2 processes were fully delayed. This suggests that the tasks competed for capacity in order to be carried out, which is not in line with the RSB model where tasks passively queue for capacity, or in line with the capacity-sharing theory where tasks share capacity.

Dual-tasking with linguistic tasks

How do linguistic processes fit into the stages described by dual-task theories? One reason this question is difficult to answer is that the majority of dual-tasking research uses simple tasks, such as tone identification or numerical judgements. However, language processing involves multiple component processes. It is not only that a stimulus is perceived and directly mapped to a response; there are multiple linguistic stages in order to have a complete phonological and semantic representation of a word. One question which arises is whether the processes in production and comprehension are best ascribed to perceptual, central, or execution stages? And, how are these tasks carried out simultaneously if processes or representations are shared between tasks?

I now discuss the evidence determining which linguistic stages or processes require capacity. This is important as it allows us to make predictions about which processes could occur in parallel. To test whether a process requires capacity, a linguistic process is combined with a non-linguistic process in a PRP experiment. The process being tested is manipulated to make this process either easier or more difficult, and this typically results in longer response times. When this linguistic task is combined with a non-linguistic task, researchers investigate whether the manipulation is still measured in the response. If the manipulation is no longer measured, this indicates that the linguistic process can be carried out in parallel with another task, which would ascribe this process to the perceptual stage of dual-task theories. However, if the manipulation is measured, this indicates that the linguistic process was delayed until sufficient capacity was available, suggesting that the linguistic process is either ascribed to the central or execution stage of dual-tasking. The majority of experiments discussed below use this logic.

Language production

Conceptual selection

Research investigating whether conceptual selection requires capacity suggests that it does. A study by Mädebach, Jescheniak, et al. (2011; see also Wagner, Jescheniak, & Schriefers, 2010) tested processing ease in a series of picture-word interference experiments. Two images were presented simultaneously, with one to-be-named object and one distractor context object. Participants heard auditory distractors, which were phonologically related or unrelated to the context image. This allowed testing of whether the context object names were retrieved, as related distractors would speed picture naming. Processing ease was manipulated by varying how strongly degraded the images were.

Faster naming RTs (comparing related vs unrelated distractors) were only found when both images were non-degraded. This shows that difficulties in visual processing affected selection in the conceptual system, as when the images were harder to process, the phonological form of the distractor word was not activated. Importantly, this study demonstrates that very early in the production system resources are needed.

Lemma selection

In a seminal dual-tasking study which investigated linguistic capacity constraints, V. S. Ferreira and Pashler (2002) tested whether lemma selection, word form selection and phoneme selection require capacity in two experiments. All manipulations are described here but the word form and phoneme manipulation results are discussed in later sections. In Experiment 1, participants carried out the dual-task of picture naming after reading cloze sentences (task 1) and tone identification (task 2), with SOAs of 50, 150 and 900ms. To manipulate the duration of lemma selection, high or low cloze constraint sentences were presented. To manipulate the duration of word form selection, high and low frequency pictures were displayed.

V. S. Ferreira and Pashler (2002) found effects of the lemma selection manipulation. High cloze pictures were named faster than low cloze pictures, and importantly this effect propagated through to task 2 RTs. This suggests lemma selection requires capacity, as if no capacity were required then the longer lemma selection process in task 1 would be carried out in parallel with processing in task 2, and would not affect any central task 2 stages.

In Experiment 2, the same design was used with a different set of manipulations. To manipulate lemma selection participants named images while written distractor words were displayed. Distractors appeared simultaneously with the image or 1000ms after. The distractors were either conceptually related (to manipulate the duration of lemma selection), phonologically related (to manipulate the duration of phoneme selection), or unrelated to the picture name (as in Damian & Martin, 1999). Ferriera and Pashler again found longer naming latencies with conceptually-related distractors compared to unrelated distractors, and this effect propagated to task 2 RTs. This again demonstrates capacity demands of lemma selection.

In a different series of studies, lemma selection has also been shown to require capacity (Ayora, Janssen, Dell'Acqua, & Alario, 2009; Kleinman, 2013; Piai et al., 2014; Schnur & Martin, 2012). Piai et al. (2014) carried out six experiments testing a dual-task of tone identification (task 1) and picture naming with visual distractors (task 2), with SOAs of 0ms and 1000ms. The distractors were conceptually related or unrelated to the images. The authors were interested in whether semantic interference (i.e. slower naming RTs to pictures with conceptually related versus unrelated distractors) would be found at both SOAs. They found semantic interference in all experiments, even while the experiments differed on various dimensions, such as the SOAs, number of named images, types of stimuli, and tone frequency and duration. These results provide compelling evidence that lemma selection requires capacity. However, other studies have found no effects of semantic interference at short SOAs using similar paradigms (Ayora et al., 2011; Dell'Acqua, Job, Peressotti, & Pascali, 2007). This suggests that under certain conditions, lemma selection can occur concurrently with another task.

Phonological word form

Studies testing whether word form selection, including phonemic manipulations, have found mixed evidence about whether this stage requires capacity (e.g., Fargier & Laganaro, 2016; Klaus, Mädebach, Oppermann, & Jescheniak, 2017; Roelofs, 2008a; Sasisekaran & Donohue, 2016).

As previously described, V. S. Ferreira and Pashler (2002) carried out two dual task studies. In Experiment 1 they manipulated the level of word form selection by presenting high and low frequency images. They found that task 1 naming RTs were slower to low frequency than high frequency images, and that this effect percolated to task 2 tone identification RTs. This indicates that word form selection requires capacity. In Experiment 2 they manipulated phoneme selection by presenting images with phonologically related or unrelated written distractor words. In this experiment, only task 1 RTs showed an effect of the phonological manipulation, as naming RTs were faster to phonologically related versus unrelated images. This effect did not carry over to task 2 RTs. This suggests that phoneme selection – selecting phonemes based on the phonological word form – is not subject to capacity limitations.

Alternate results were found by Cook and Meyer (2008). In a series of dual-task experiments, they tested picture naming (task 1) and tone identification (task 2), presented with SOAs of 50, 150 and 900ms. In Experiment 1, participants named pictures (drawn in green) with pictorial distractors (drawn in red), which were phonologically related, unrelated, or identical to the picture. Cook and Meyer found that task 1 RTs were faster with identical and phonologically related distractors (compared to unrelated distractors), and this effect percolated to task 2 RTs. This suggests that phoneme selection demands capacity, contrary to the results of V. S. Ferreira and Pashler (2002).

In Experiment 2, the pictorial distractors were replaced with written word distractors, to more closely approximate the study by Ferreira and Pashler. The same pattern of results as Ferreira and Pashler was found: faster task 1 RTs to phonologically related versus unrelated distractors in task 1, and no propagation to task 2 RTs. In Experiment 3, Cook and Meyer tested why only pictorial distractors showed propagation effects. They hypothesised that distractor words slowed down speech monitoring processes, but distractor images did not. Therefore, pictorial and written distractors which were related to the image had the same facilitatory effect on phoneme selection, but written distractors additionally slowed speech monitoring processes, which would cancel out the measurement of propagation. Therefore, in Experiment 3 the same paradigm was used but with masked written primes. Task 1 RTs were faster when primes were phonologically related versus unrelated to the image, and importantly this effect percolated through to task 2 RTs. This finding is important as it demonstrates that phoneme selection is capacity demanding (see also Ayora et al., 2011).

A further study by Dent, Johnston, and Humphreys (2008) tested whether the effects of age of acquisition and word frequency occur at stages requiring capacity. In two experiments participants named images which varied on only age of acquisition or word frequency (task 1), and carried out tone identification (task 2), presented with SOAs of 50, 150 and 900ms. Dent and colleagues found that only frequency effects percolated to task 2 RTs. This suggests that processes corresponding to age of acquisition effects do not require capacity, whereas processes corresponding to frequency effects do. Dent et al. (2008) place the age of acquisition effect at lemma retrieval (providing diverging evidence from that discussed in the section *Lemma Selection*), and the word frequency effect at word form retrieval (in line with evidence from Jescheniak & Levelt, 1994).

Phonetic encoding

There is little research investigating whether phonetic encoding requires capacity, but available evidence suggests that it does. While previous research suggested that language processes including lemma selection and word form selection require capacity (for review see Roelofs & Piai, 2011), Jongman, Roelofs, and Meyer (2015) demonstrated that sustained attention was most important after word form selection. This suggests that these later stages also require capacity. In Experiment 1, participants named pictures, either with a determiner-noun phrase or with a longer determiner-adjective-noun phrase (task 1), and carried out an arrow discrimination task (task 2). Task stimuli were presented simultaneously, and RTs and eye movements were measured. Jongman et al. (2015) exploited the fact than participants tend to look at an image to be named until the word form is encoded, and then shift their gaze shortly before speech onset to the next experimental item (A. S. Meyer & Van der Meulen, 2000). Participants' sustained attention ability was also measured. Jongman et al. (2015) found that sustained attention ability correlated with task 1 naming RTs, but not with gaze durations. The fact that sustained attention ability correlated with RTs suggests that sustained attention is required during the entire timecourse of naming, and not only up to word form encoding, as a correlation between naming RTs and gaze durations would have suggested. This pattern of results suggests that sustained attention is especially important for phonetic encoding, which occurs after word form encoding.

In Experiment 2, Jongman et al. (2015) tested whether sustained attention is required in typical picture naming or only on more taxing dual-task trials. Participants named single images, and carried out a dual-task of picture naming (task 1) and arrow identification (task 2) with simultaneous presentation. Sustained attention abilities were again measured. Jongman et al. (2015) found that sustained attention ability correlated with naming RTs in both tasks, but with a higher correlation in the dual-task. This suggests that sustained attention is important in general for phonetic encoding, but is relied on more heavily in a dual-task.

Language comprehension

Phoneme selection

Gaskell et al. (2008) investigated whether phoneme selection in spoken word recognition requires capacity. Three experiments were carried out, with the dual-task of colour categorisation (task 1), and phonemic categorisation (task 2), presented with SOAs of 100, 200 and 1000ms. For the phonemic categorisation task, participants would hear a spoken word and determine whether the final consonant was one of two letters displayed on screen. In all experiments a difficulty manipulation was applied to task 2, and Gaskell and colleagues investigated whether this difficulty manipulation would be present in task 2 RTs.

In all experiments, no phonemic difficulty manipulations were measured in task 2 RTs. These results indicate that difficulty in phonemic processing was resolved concurrently with task 1 processing, and thus phoneme selection does not require capacity. This suggests that phoneme selection in comprehension can proceed at the same time as processes in other tasks.

Word form selection

There are conflicting results on whether word form selection in comprehension requires capacity. Cleland, Gaskell, Quinlan, and Tamminen (2006) tested a dual-task of colour categorisation (task 1) and lexical decision (task 2), presented with SOAs of 100, 200 and 800ms. The frequency of the task 2 words was manipulated as a measure of word form selection. Task 2 was an auditory (Experiment 1) or visual (Experiment 2) lexical decision task.

In both experiments no frequency effect was measured in task 2 RTs with short SOAs, but was present at long SOAs. This suggests that word form selection, manipulated here using frequency, does not require capacity.

However, Lien, Ruthruff, Cornett, Goodin, and Allen (2008) found a different pattern of results, albeit with a different experimental design. They tested dual-tasking in an EEG study and asked whether the P3 component would differ between low and high frequency words. The P3 component has previously been found to be sensitive to word frequency, such that frequent words have a larger P3 amplitude and shorter P3 latency (Polich & Donchin, 1988). Participants in Lien et al. (2008) carried out a dual-task of tone identification (task 1) and visual lexical decision, where the words varied on frequency (task 2), presented with SOAs of 100, 300 and 900ms. Lien and colleagues tested whether the P3 amplitude and latency would be different to high versus low frequency words in the lexical decision task.

The behavioural results showed a difference between high and low frequency word responses, regardless of SOA. However, with shorter SOAs the effect was smaller. These behavioural results are not entirely consistent with those found by Cleland et al. (2006). The ERP effects also showed a similar pattern. With more task overlap, the difference in amplitude between the high and low frequency words was greatly reduced, compared to with less overlap. This indicates that some processing of word forms requires capacity, but other processes can proceed in parallel with task 1.

In a further study, Cleland et al. (2012) tested whether word form selection requires capacity using a different paradigm. Participants listened to a spoken word through headphones (task 1; note this task required no response), and carried out visual lexical decision (task 2), with SOAs of 0 and 100ms. Participants were told that there would be a memory test for the task 1 words. The spoken words were manipulated by uniqueness point; some words had early uniqueness points and some late. Cleland and colleagues tested whether the effect of uniqueness point (i.e. the point in the word when that word becomes distinguishable from all other words) would be present in task 2 RTs. Indeed, they found longer task 2 RTs when the task 1 words had later uniqueness points.

In a second experiment, Cleland et al. (2012) demonstrated that this effect was not due to intentional encoding of the spoken words for the memory test, and tested whether similarity between tasks would have an effect. Participants carried out two dual-tasking conditions: listening to a spoken word through headphones (task 1) with visual lexical decision (task 2), or listening to a spoken word (task 1) with colour judgement (task 2). The SOAs tested were 0 and 100ms. A uniqueness point effect was found in task 2 RTs, confirming that this effect is not only attributable to intentionally remembering the spoken word. Additionally, in words with late uniqueness points, the uniqueness point effect was larger in lexical decision RTs compared to colour judgement RTs. A third experiment found the same pattern. These results suggest that word form selection requires capacity, and they hint that similarity in tasks (with two linguistic tasks) may result in larger effects.

In general, these results suggest that at least some aspects of word form selection in comprehension can proceed in parallel with another task. However, percolation of effects also shows that some aspects of word form selection require capacity. This suggests that word form selection is composed of multiple processes, some of which are resource-demanding.

Semantic activation

Hohlfeld, Sangals, and Sommer (2004) tested whether semantic activation requires capacity by investigating, in an EEG study, whether the N400 component was attenuated during dual-tasking. Participants heard prime words. They then carried out letter judgement (task 1) and heard a target word (task 2, but note no response is given), presented with SOAs of 100, 400 and 700ms. The prime-target pair were either synonymous or non-synonymous. A smaller N400 component was expected to the target word in a synonymous pair compared to a non-synonymous pair. They tested whether the amplitude and latency of the N400 would be affected by overlap between task 1 and presentation of the target word.

Hohlfeld et al. (2004) found that the N400 effect was delayed when task 1 and the target word overlapped, but the amplitude of the effect was not affected by task overlap. These results suggest that semantic activation requires capacity, and was delayed during dual-tasking, but as with word form selection some semantic processing occurred in parallel with task 1 processing.

In Experiment 2, Hohlfeld et al. (2004) tested two different task 1 tasks: a spatial response task or letter judgement. In all other respects the paradigm was the same as Experiment 1. This experiment was designed to rule out that the delay in the N400 in Experiment 1 was due to dual-tasking with two linguistic tasks. They found consistent effects: the N400 effect was delayed when the tasks overlapped, but the delay was larger with two linguistic tasks compared to one linguistic and one non-linguistic task. Again, the amplitude did not differ. This suggests that semantic activation requires capacity, but the delay in semantic access is greater when the two tasks are linguistic. However, some semantic access occurred in parallel with the first task.

Other studies using very similar paradigms have found the same patterns of results; namely that the N400 is attenuated by dual-tasking, indicating that semantic access requires capacity but does not demand full capacity (Hohlfeld & Sommer, 2005; Lien et al., 2008; Relander et al., 2009; Vachon & Jolicœur, 2012). This suggests that semantic activation in comprehension requires a graded amount of capacity.

Using a different paradigm, Logan and Schulkind (2000) tested whether semantic information in two separate tasks can be retrieved in parallel. Both tasks were visual lexical decision tasks and the words for each task were presented simultaneously. They found evidence that semantic information for both words was retrieved in parallel, because both responses were faster if they were congruent. These results provide corroborating evidence that semantic activation is possible in parallel with another task, and thus that semantic access does not require full capacity.

Conceptual selection

Very little research has investigated whether conceptual selection in comprehension requires capacity. Halin, Marsh, and Sörqvist (2015) carried out a dual-tasking study using an n-back task (n = 1 or 2; task 1), which manipulated cognitive load. Task 2 was to listen to an auditory story. Participants were told to ignore the story, but a surprise memory test involving questions about the concepts in the story was given afterwards. Halin et al. (2015) found that memory task performance was worse in participants who had the 2-back task compared to participants who had the 1-back task. This suggests that while engaged in a task with high cognitive load requirements, conceptual selection is affected, indicating that conceptual selection requires some capacity.

Drawing together production and comprehension capacity demands

Capacity is not required to the same extent in production and comprehension. In production, phonemic and phonetic encoding require capacity. However, in comprehension, these processes do not place strong demands on capacity, evidenced by phoneme selection in comprehension being carried out in parallel with a concurrent task. The processes of lemma selection and phonological word form selection in production require capacity. In comprehension, these processes do not require full capacity; effects of frequency and lexical selection require some capacity, but some processing is also carried out in parallel with another task. Accessing semantic information in comprehension similarly requires a graded amount of capacity, such that some semantic processing is serial and some parallel. In both production and comprehension lexical concept selection requires capacity. Thus, in general, production processes are capacity demanding and comprehension processes require no or graded capacity.

Why are capacity demands smaller in comprehension than in production? This is especially puzzling considering some processes, such as lexical selection, are assumed to be shared between tasks (see section *Shared and distinct representations in production and comprehension*).

One possibility is that comprehension may be able to cope with less capacity allocation than production without being noticeably hindered. One theory of comprehension – the good enough theory (F. Ferreira, Bailey, & Ferraro, 2002; Karimi & Ferreira, 2016) – argues that comprehension only needs to be good enough for the task at hand. Compensatory mechanisms in the comprehension system may compensate for any errors. In contrast, production must be accurate so that the speaker can say the words they want to say, in the correct order and with the sounds in their correct place. This may place greater demands on capacity. A second speculative possibility is that the difference in capacity demands reflects different underlying processes in production and comprehension.

While the above-reviewed studies show that production and comprehension have different capacity demands, they do not address the question of how two linguistic tasks would be carried out concurrently. Aside from some dual-task studies which used two linguistic tasks in comprehension (such as Cleland et al., 2012; Hohlfeld et al., 2004, though note these studies did not specifically test dual-tasking with two linguistic tasks; their studies were designed to rule out linguistic reasons for their effects), there is scant evidence for how production and comprehension are carried out in parallel. This is the main question underlying the temporal coordination of speech planning and comprehension, and thus I ask how processing works when the two tasks are combined. Here I review two studies which have combined word production with word comprehension[3].

Fargier and Laganaro, 2016

Fargier and Laganaro (2016) carried out a dual-tasking EEG study with picture naming (task 1) and tone or syllable identification (task 2). Participants saw a picture and after a fixed SOA of 300ms heard one of five tones or syllables. A go/no-go task was used; participants responded to only one tone/syllable. Participants always named the picture. Only no-go responses were analysed (80% data). Additionally, in a passive condition participants ignored the sound, and in a single naming condition, no auditory stimuli were presented. Fargier and Laganaro

[3]Note that some studies have combined two linguistic tasks testing at the level of sentence processing (Chipunza & Mandeya, 2005; Klaus et al., 2017; Moisala et al., 2015). As we focus on word production and comprehension we do not review these here, but we direct the interested reader to these studies to learn how people dual-task with higher linguistic units.

tested whether neural modulations would be different in an active dual-task of two linguistic tasks versus one linguistic and one non-linguistic.

Behaviourally, naming RTs were longer when syllables were the secondary task (953ms) compared to tones (887ms) in the active dual-task. The EEG results showed a modulation of the EEG signal starting approximately 400ms after trial onset in the syllable condition, which was not present in the tone condition. Fargier and Laganaro (2016) argue that this is a phonological level effect, because this effect was present only when the secondary task was linguistic, and the time course is consistent with the phonological encoding stage in production (Indefrey & Levelt, 2004; Indefrey, 2011). This suggests that only the syllable interfered with phonological encoding of the picture name. Interestingly, there were also modulations in the EEG signal in both tone and syllable conditions between 200 and 300ms, compared to single naming. At this time, the auditory distractor had not been presented, so any modulations could not be due to the auditory stimulus. The authors suggest that this effect is due to the dual-task: participants know a sound will soon be presented and dedicate some attention to processing the incoming sound. The timing of this effect is consistent with an effect at the level of lexical selection (Indefrey & Levelt, 2004). Therefore, if attention is reserved for secondary task processing, it may be that lexical access in picture naming is affected.

In general, these results demonstrate that a concurrent syllable (which is also an existing word in the native language of the participants) affects phonological encoding of the picture name more than a concurrent tone. Carrying out two linguistic tasks almost simultaneously results in interference in at least one task (picture naming; no syllable response was measured).

Paucke, Oppermann, Koch and Jescheniak, 2015

Paucke, Oppermann, Koch, and Jescheniak (2015) carried out a dual task involving two different picture stimuli, with a different task to be performed on each picture. The two tasks were either both linguistic, or one linguistic and one non-linguistic. Participants named the first picture, and carried out either a phoneme detection task (*does the name of this picture contain a X?*; linguistic task) or a size decision task (*does this fit in a shoebox?*; non-linguistic task) on the second picture. In Experiment 1 pictures were presented simultaneously, and in Experiment 2 they were presented with SOAs of 0, 150, 300 and 1200ms. The first picture also varied by frequency and recognisability (by degrading the image), to investigate if these effects propagated to task 2 RTs. Paucke et al. (2015) asked whether task similarity would increase dual-task interference, and whether there would be cross-talk between tasks.

Paucke and colleagues found that naming RTs to the first picture were longer when the tasks were similar (task 2 phoneme decision) compared to less similar. Additionally, the frequency and degradation effects propagated through to task 2 RTs with simultaneous task presentation, but the frequency effect was larger when the two tasks were more similar. These results show cross-talk between similar tasks and suggest that shared resources were involved.

In general, these results show that concurrent processing of two linguistic tasks results in cross-talk between tasks. However, both tasks tested are 'production' tasks; both responses are based on production-like behaviour. The same pattern of results may not generalise to a production-comprehension dual-task. Despite this, these results are highly informative because, if production and comprehension are tightly linked, cross-talk is predicted.

A working model of temporal overlap in production and comprehension

Research by Fargier and Laganaro (2016) and Paucke et al. (2015) show us that when carrying out two linguistic tasks, there is greater interference than when carrying out one linguistic and one non-linguistic task. Additionally, two linguistic tasks affect one another more than one linguistic and one non-linguistic task. Taken together with the evidence from capacity limitation studies, a proposal for how to reconcile word production and comprehension theories with dual-task theories is described. The core question here is how to map the stages in word production and comprehension to the stages in dual-task theories. More concretely, I ask: which stages in production and comprehension are central stages in dual-tasking?

Evidence suggests that each level of production is a central stage, as each stage is shown to require capacity. This would imply that in a picture naming task, only visual perception of the image and motor execution of the response are capacity-free processes; every process in between requires capacity. Specifically based on the WEAVER++/ARC model (Roelofs, 2014), this suggests that lexical concepts, lemmas, output forms and output phonemes are all central stages in a dual-task framework.

In comprehension, only some processes require full capacity. Taking the model of WEAVER++/ARC, the stage of input phonemes is not central. Input forms require some capacity, and thus can be considered a central stage. Lemma selection (lexical selection) and lexical concepts require capacity, and thus are central processes. However, there is a caveat that these two final stages can run a little in parallel with other stages. This would suggest that the final two stages are made up of some processes which require capacity and some which do not.

Based on the WEAVER++/ARC model, the working model also suggests some linguistic levels are shared between production and comprehension, and some levels are distinct. Taken with the capacity demand evidence, the working model argues that lexical concept and lemma selection cannot be carried out concurrently in production and comprehension, but form and phoneme selection can. I argue that selection of representations at these levels is capacity-demanding, and not necessarily the activation of representations. Figure 2.3 shows a box and arrow depiction of this working model. This provides constraints on the temporal overlap of word production and comprehension, which addresses a central question of this review.

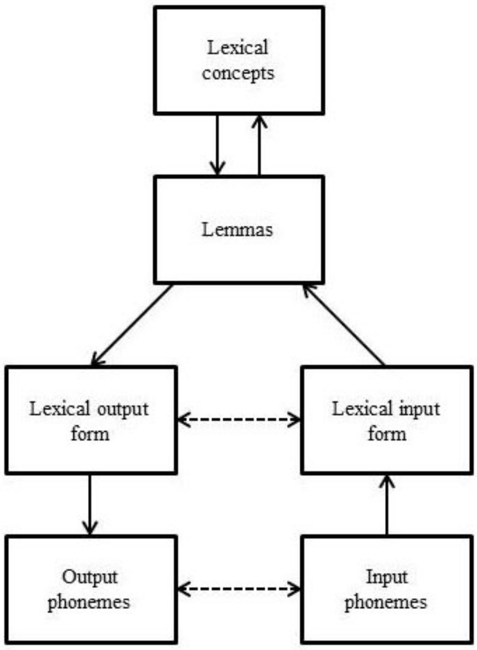

Figure 2.3: Diagram of the working model. Lexical concepts and lemmas are shared between production and comprehension. Input and output forms and phonemes are distinct but linked. Note that input and output forms also have bidirectional links, along with input and output phonemes.

Before discussing the working model in more detail, I address how this model fits with dual-task theories. The working model fits most strongly with capacity-sharing theories, where stages which require capacity share this capacity. However, the model makes the prediction that two stages - lexical concept selection and lemma selection - cannot be carried out concurrently, even with capacity-sharing. Thus, one prediction from this model is that these stages require a large amount of capacity to be carried out, and there is not enough capacity in reserve for the other task. This may be because a stage requires a minimum amount of capacity to run. Alternatively, it may be that with such a small amount of capacity left for the second task, central stages of the second task are either not run, or have such a small amount of capacity dedicated to them that this allocation is negligible. In contrast, the stages of form and phoneme selection can run in parallel sharing capacity. Predictions from the working model are not in line with the RSB theory because this theory predicts that capacity-demanding stages of different tasks should run sequentially. Figure 2.4 shows hypothesised capacity requirements of word production, illustrating the capacity needed for the stages. The capacity demands of lexical concept and lemma selection are almost at ceiling, leaving very little remaining capacity for another task. Form and phoneme selection require smaller amounts of capacity, leaving capacity available for a second task to be combined with these stages.

Altogether, this suggests that dual-task theories are not adequate to apply directly to language processing (or likely other complex tasks with multiple central stages), and adjustments to predictions made by the theories is most appropriate when thinking about the temporal overlap of production and comprehension. It is not as simple as taking any processes which require capacity and mapping these processes to a dual-tasking central stage.

I now discuss the working model of the temporal overlap of word production and comprehension in more detail, describe how the different processes may or may not run in parallel, and derive predictions from this model.

As demonstrated in Figure 2.3, I posit that lexical concept selection and lemma selection in production and comprehension cannot run in parallel; the processes at these levels must run serially. In contrast, form and phoneme selection in production and comprehension can run in parallel. However, due to the capacity demands which arise at these stages (especially in production), these stages do not run as efficiently when running in parallel compared to when running serially. This is partially due to capacity limitations, and partially due to the links between the stages; the representations can 'talk' to one another. From

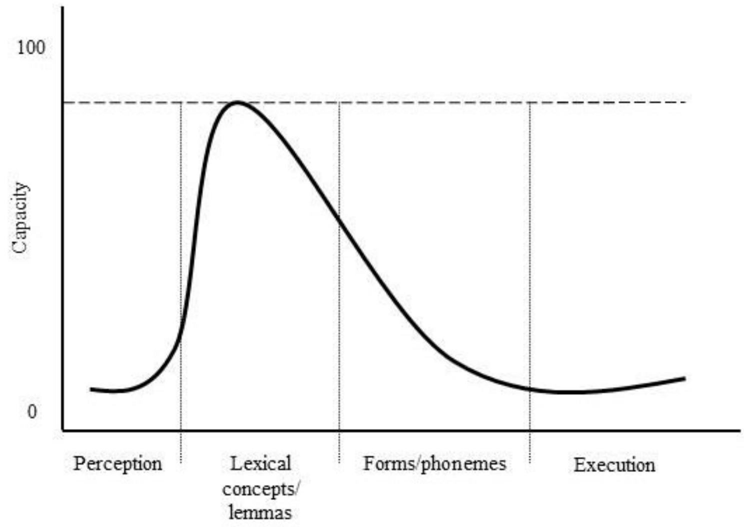

Figure 2.4: Hypothetical depiction of capacity demands across processing stages. The y axis shows capacity with an arbitrary hypothesised span of 0 to 100 (with 100 being maximum capacity). The x axis shows the processing stages of word production according to hypothesised capacity demands. Lexical concept and lemma selection require large amounts of capacity, rendering them almost uncombinable with another task. In contrast, form and phoneme selection require less capacity and can be combined with a separate task.

this model, it follows that different amounts of overlap will result in different amounts of interference between tasks.

If we take as an example a word to be produced, such as *caterpillar*, and a word to be comprehended, such as *watermelon*, then we can walk through the overlap in processes. If a speaker initiates planning of the word *caterpillar* then this word must go through the stages of lexical concept selection and lemma selection. While these processes are acting on the *caterpillar* representation, they cannot be used by the *watermelon* representation because capacity demands are too high. Thus the *watermelon* representation must wait until these stages are available. After these stages, *caterpillar* goes through form selection and phoneme selection. These stages can run in parallel with other stages, such that form and phoneme selection of *caterpillar* or lemma selection or lexical concept selection of *watermelon* can run in parallel. The same pattern holds in reverse; if *watermelon* is accessing lexical concept and lemma selection processes then *caterpillar* must wait.

If we assume that production and comprehension processes begin at the same time on *caterpillar* and *watermelon* then there may be little occasion for overlap of shared stages. While initial lexical concept selection and lemma selection processes are acting on *caterpillar*, phonemic and form selection is acting on *watermelon*. *Watermelon* likely activates a cohort of words (e.g., Gaskell & Marslen-Wilson, 1997, 2002), but this process can be accounted for under phoneme and form activation processes. Depending on the length of these processes, lexical concept selection and lemma selection may be free stages when *watermelon* requires them, as the *caterpillar* representation may already be 'finished' with these stages.

There is some support for this proposal based on the timings of component stages. In single picture naming, conceptual activation is estimated to take around 200ms, and lemma selection an additional 75ms, meaning both processes are accessed[4] by around 275ms after picture presentation (Indefrey & Levelt, 2004; Indefrey, 2011). An average picture naming latency is around 600ms. In comprehension, form, lexical and semantic information is assumed to be available after approximately 200ms from the uniqueness point of the word (Egorova, Shtyrov, & Pulvermuller, 2013; MacGregor, Pulvermüller, Van Casteren, & Shtyrov, 2012; Penolazzi, Hauk, & Pulvermüller, 2007; Pulvermüller, 1999; Pulvermüller, Shtyrov, & Ilmoniemi, 2005; Pulvermüller, Shtyrov, Ilmoniemi, & Marslen-Wilson, 2006). Thus, depending on the words (and surely other factors), simultaneous onsets of production and comprehension may be possible with minimal overlap of shared stages.

However, if production and comprehension do not occur with simultaneous onsets, certain processes must queue. If production processes begin before comprehension processes then there will be minimal overlap. If production processes begin after comprehension processes have started then there will be more overlap between stages. Specifically, if lemma selection and lexical concept selection are required for comprehending *watermelon*, then these processes cannot begin for producing *caterpillar*. This is a true bottleneck in the system. This bottleneck may be resolved in two ways: either *caterpillar* waits for access to these stages, or *caterpillar* has priority for these stages and *watermelon* must halt processing and resume after *caterpillar* is finished. Either possibility is

[4] I do not wish to imply that these stages are complete by this time point, as information can cascade through the production system such that downstream processes begin before upstream stages are complete (e.g., Dell, 1986; Dell & O'Seaghdha, 1992). In comprehension, access to different levels of linguistic information may be rapid but may also continue through time (Pulvermüller, 1999; Strijkers, Costa, & Pulvermüller, 2017). I merely take these timings as rough estimations of access to production and comprehension stages.

plausible, but I hypothesise that it is simplest for a representation already being processed at that stage to 'lock' the stage until it is finished, such that if lexical concept and/or lemma selection is in process for *watermelon*, *caterpillar* must wait. This could change depending on task demands placing heavier demands on aspects of executive control (Miyake et al., 2000). Executive control has been shown to be needed in object and action naming (Shao, Roelofs, & Meyer, 2012). For example, research investigating the role of inhibition in word production has found that speakers who are selectively able to inhibit competing words take less time to carry out lemma selection, evidenced by smaller semantic interference effects in a PWI paradigm (Shao, Meyer, & Roelofs, 2013). Additionally, speakers with greater non-selective inhibition ability, referring to the ability to suppress any unwanted response, were faster at naming images (Shao et al., 2013). This suggests that general inhibition skill may speed up some decision processes. These results suggest that the speed or scheduling of some stages in the working model could be affected by factors external to these linguistic stages.

Relationships between words may also affect how they are simultaneously processed. Research using PWI paradigms shows that categorically related words presented simultaneously with a picture interfere with naming, resulting in longer naming responses (e.g., Damian & Martin, 1999; Roelofs & Piai, 2017; Schriefers et al., 1990). However, associatively related words can speed naming responses (see Abdel Rahman & Melinger, 2009). Words phonologically related to a picture name can also speed naming responses (e.g., A. S. Meyer & Schriefers, 1991; Schriefers et al., 1990). These relationships will likely also hold when coordinating word production and comprehension. The facilitatory or inhibitory effects have their origin when selecting representations at these different stages. Therefore, categorically related words will slow down lemma selection, which will slow processing in both the planned word (e.g. *caterpillar*) and the comprehended word (e.g. *butterfly*). Associatively related words will speed lemma selection, leading to faster processing in the planned word (e.g. *caterpillar*) and the comprehended word (e.g. *metamorphosis*). Phonologically related words will speed form selection, leading to faster form selection for the planned word (e.g. *caterpillar*) and comprehended word (e.g. *catamaran*). Thus, representational similarity could help or hinder selection when coordinating word production and comprehension.

Considering that there is evidence that informs us about how relationships between words can affect naming, one may wonder why there is a need to discuss the overlap between production and comprehension at all. Why do the results from PWI studies not already address the temporal overlap question? One limitation with applying knowledge from PWI studies to coordinating production and comprehension is that typically only one output response is measured (the naming RT). With only one response, we do not know how the distractor word was affected by concurrent picture naming. Discovering how both the produced and comprehended words are affected in a dual-task is one step to determining how production and comprehension are carried out with temporal overlap.

What are the predictions that this model makes? Firstly, lexical concept selection and lemma selection are stages which cannot be carried out in parallel in the two tasks. Therefore, the model predicts that lemma selection and lexical concept selection would be delayed in either production or comprehension while the other task is occupying one of the stages. This would be measurable in RTs in these tasks; if one task has to wait for the bottleneck, that task will take longer. This would only be predicted when there is overlap in the stages, as without temporal overlap there is no queuing. Another test for this is to investigate whether neural signatures associated with lemma selection or lexical concept selection are delayed with increased overlap (cf. Fargier & Laganaro, 2016).

This model also predicts that form and phoneme selection stages can run in parallel, though they may 'talk' to one another. As stated above, the model predicts that with phonological overlap between production and comprehension, processing will be faster than if there is no overlap. This effect should be largest (i.e. fastest) when the stages are simultaneously active. There would be slower responses if the words are not phonologically related. However, due to capacity demands, simultaneous activation of stages will also result in slower responses than if the stages are not simultaneously active. This would be measured in RTs to the tasks, and again could be tested with an EEG study to determine whether neural signatures of phonological encoding are shifted in latency or show amplitude changes depending on overlap.

A further more general prediction is that it is possible to carry out form and phoneme stages in parallel. It is possible that there are harder restrictions on coordinating word production and comprehension, such that stages which require capacity are carried out serially. For instance, entirely serial processing would predict that all capacity demanding stages in producing *caterpillar* are carried out before any capacity demanding stages of comprehending *watermelon*, or vice versa. Specifically for the working model, this would predict that form and phoneme stages in production cannot be combined with other stages. The result would be that processing in either production or comprehension (whichever is deemed task 1) would be carried out in much the same way as when performed alone. A serial processing hypothesis predicts that there would be no effects of task overlap in task 1, and any effects would only be found in task 2. Therefore, testing two linguistic tasks in a dual-task situation will allow us to determine whether there is any parallel processing at all.

The current literature has found variable effects of whether task 1 is affected by an SOA manipulation (i.e. overlap between task 1 and task 2). In some studies task 1 RTs are affected by an SOA manipulation (Paucke et al., 2015; Roelofs, 2008a; Sasisekaran & Donohue, 2016), and in others there is no SOA effect in task 1 RTs (Cook & Meyer, 2008; Dent et al., 2008; V. S. Ferreira & Pashler, 2002; Logan & Schulkind, 2000). Further research should determine under which conditions the tasks are carried out in parallel.

The model described above only discusses the overlap of producing one word with comprehending one word. However, in a conversation, people tend to speak and listen to utterances composed of phrases. The working model can make some predictions for how this would be carried out, if it is scaled up. If a phrase is being planned, and this phrase requires lexical concept selection and lemma selection for the words in that phrase, then these stages will be occupied longer than with a single word. Again, the items to be comprehended will need to wait for the stages to be available. Alternatively, if the comprehended words access these stages first, then the phrase to be produced will have to wait for access. Form and phoneme stages can run in parallel, or with other stages, for both production and comprehension.

An aspect not covered by this model is how access to each stage is determined. It may be the first task which reaches the bottleneck (Pashler, 1994; Tombu & Jolicœur, 2003), tasks may compete for access (Marti et al., 2015), or it may be decided by a scheduler (D. E. Meyer & Kieras, 1997). I speculate that production-related processes have priority over comprehension-related processes when tasks are equally important. The logic supporting this is that production processes are potentially more important, because for a speaker to produce their target word, these processes must be carried out. In comparison, while the auditory speech signal is fleeting, a speaker could ask their interlocutor to repeat themselves if needed. Thus, I currently make the strong prediction that production-based processes will acquire access to stages before comprehension-based processes. In contexts where it is more important to listen to an interlocutor, or to be ready to speak, these priorities may change.

Further questions which arise are how relationships at different levels can affect one another. For instance, how is temporal overlap coordinated if words are both categorically and phonologically related? And how is simultaneous processing of these words affected by the amount of overlap between the tasks? A different related question is how different stages may interact with one another. For instance, the form selection process in production is capacity demanding, and thus should be unable to occur with lemma selection or lexical concept selection in comprehension. However, these stages are separate and are currently hypothesised to run in parallel. Predictions from this working model can thus be tested in future studies.

This working model is based on current well-developed theories of language production and comprehension (Dell, 1986; Gaskell & Marslen-Wilson, 2002; W. J. M. Levelt et al., 1999; McClelland & Elman, 1986; Roelofs, 1997, 2014) which allow us to make quite precise predictions and informed assumptions about how linguistic processes in production and comprehension are carried out. However, a different class of production and comprehension models propose simultaneous activation of semantic and phonological information in planning and comprehension (Pulvermüller, 1999; Pulvermüller et al., 2005; Strijkers et al., 2017). Under these theories, all linguistic information is initially activated and a process of reverberation causes later re-activation of different linguistic levels. Currently, these models are not specified in enough detail to determine exactly how simultaneous production and comprehension (as in our working model) would be carried out.

In conclusion, in this review I have discussed dual-tasking theories, linguistic theories, and evidence of capacity allocation in linguistic tasks. I have proposed a working model of how word production and comprehension may be carried out in close temporal overlap. This model sets some constraints on this process, such that lexical concept selection and lemma selection run as serial processes, but form and phoneme selection are separate, and thus can be carried out in parallel. Some predictions from the model are tested later in this thesis, but this model makes many further testable predictions. I adopt strong prediction positions arising from the model, with the full expectation that some, if not many, will not be borne out by the data. Only with these strong predictions will a theory of how production and comprehension are carried out in overlap be refined. This contributes to the wider dual-tasking literature by presenting a specific case of dual-tasking with similar and complex tasks, and contributes to the psycholinguistic literature by positing an outline for how this important part of language – coordinating dialogue – may be carried out.

3 | Interference in dual-tasking simple linguistic tasks may be due to linguistic interference or acoustic complexity

Abstract

Recent evidence suggests that the dual-task of speech planning and comprehension often overlap during dialogue (Bögels et al., 2015). However, how people carry out two linguistic tasks concurrently is unknown. We tested whether there was additional interference in a dual-task experiment involving two linguistic tasks compared to one involving one linguistic and one non-linguistic task. In Experiment 1, participants named pictures (task 1) and identified syllables or tones (task 2) with SOAs of 50ms, 300ms and 1800ms. Syllables and tones were equally difficult to identify. We found that at the 50ms and 300ms SOA task 1 naming RTs were longer in the syllable compared to the tone condition, indicating a larger dual-task effect. The effect at 50ms was larger than at 300ms. Task 2 identification RTs were longer in the syllable condition compared to the tone condition at all SOAs, indicating cross-talk between two linguistic tasks. In Experiment 2, the same design was used but participants heard sine-wave speech versions of the syllables and were told these sounds were distorted syllables or non-speech sounds. This experiment was designed to test whether participants' perception of the stimuli as linguistic or not would result in interference, while keeping the acoustic complexity of the auditory stimuli constant. This manipulation did not have a consistent effect on task 1 or task 2 RTs, and therefore did not provide evidence supporting a linguistic basis to the cross-talk effect in Experiment 1. Together, our results suggest that two linguistic tasks interfere with one another more than one linguistic and one non-linguistic task. This effect may be due to linguistic interference or acoustic complexity of the auditory stimuli.

Introduction

During a conversation, people must plan what to say and understand what their interlocutor is saying. Despite the fact that conversation is a core niche for language use (Levinson & Torreira, 2015), very little is known about how speech planning and comprehension occur together in a conversation. In the current series of experiments, we sought to find out whether people could plan a word and concurrently identify an incoming syllable or tone, and asked whether there were any costs associated with carrying out two concurrent linguistic tasks, compared to when carrying out a dual-task with one linguistic and one non-linguistic task.

When planning whilst comprehending, participants are carrying out a dual-task, where planning is one task and comprehension is the other. Multiple dual-task studies have shown that carrying out two tasks together results in a cost compared to when carrying out the tasks individually (see Fischer and Plessow (2015) for a review). Two main theories of dual-tasking explain these costs: the response selection bottleneck theory (Pashler, 1994, 1998), and the capacity-sharing theory (Navon & Miller, 2002; Tombu & Jolicœur, 2003). Both theories agree that when carrying out a task, there are three stages: the perceptual stage (where the stimulus is perceived), the central stage (where the stimulus is processed), and the execution stage (where a response is given to the stimulus). In a dual-task, the perceptual and execution stages can run in parallel with any other stage. However, the theories diverge on whether or not this is also the case for the central stage.

The response selection bottleneck theory states that central processes cannot run in parallel because there is a bottleneck at the central processing stage; the central processes of the second task are delayed until those of the first task are finished. In contrast, the capacity-sharing theory states that central processes can run in parallel, but they must share limited capacity, meaning that central stages task longer. If full capacity is allocated to task 1, then performance will mimic that of the response selection bottleneck theory. Both of these theories assume a bottleneck at the central stage, which is either structural or based on capacity limitations.

Dual-tasking research has shown that several components of production and comprehension processes are capacity-demanding. Language production theories largely agree that when producing a word, there is first activation of a lexical concept, followed by lexical selection, phonological form retrieval, and articulation (Dell, 1986; Dell & O'Seaghdha, 1992; W. J. M. Levelt et al., 1999). There is evidence that all levels are subject to dual-task interference, including the flow of information from the conceptual level to the lexical level (Mädebach, Jescheniak, et al., 2011), lexical selection (V. S. Ferreira & Pashler, 2002; Kleinman, 2013; Piai, Roelofs, & Schriefers, 2011; Piai et al., 2014; Roelofs & Piai, 2011; but see Ayora et al., 2011; Dell'Acqua et al., 2007), word-form encoding (Cook & Meyer, 2008; V. S. Ferreira & Pashler, 2002), and converting the phonemic form to articulatory gestures (Jongman et al., 2015). This indicates each process in language production requires capacity, and as such, would be considered part of the central processing stage of dual-task theories.

Less research has investigated capacity-demanding aspects of comprehension, but the available research indicates that lexical selection requires some capacity (Cleland et al., 2012; Hohlfeld et al., 2004; but see Cleland et al., 2006). However, phonemic access (Gaskell et al., 2008) may not. Thus, comprehension requires capacity but potentially to a lesser extent than production.

The dual-task studies mentioned above all investigated production or comprehension with a non-linguistic task. However, the dual-task of production and comprehension is a dual linguistic task. If resources are shared between production and comprehension (Wickens, 2008), then the two tasks could interfere with one another and we may find evidence of cross-talk. Cross-talk arises when aspects of the two tasks interfere with one another, because the tasks involved share representations or processes (Hommel, 1998; Navon & Miller, 2002; Wickens, 2008). Cross-talk has mostly been found when there is overlap in the input or output responses or modalities, such as congruent visual stimuli locations or congruent button press responses (Alards-Tomalin et al., 2017; Arrington, Altmann, & Carr, 2003; Hazeltine et al., 2006; Hommel, 1998; Janczyk, 2016). Cross-talk could occur in two linguistic tasks if the tasks are processed in parallel, and if the task representations can affect one another. This would mean that the representation(s) of to-be-spoken words would affect, and/or be affected by the representation(s) of to-be-understood words.

Very little research has investigated whether shared processes or representations in central stages of tasks cause cross-talk. One recent study (Fargier & Laganaro, 2016) investigated whether two linguistic tasks caused more interference than a linguistic task and a non-linguistic task[1]. Fargier and Laganaro (2016) carried out a dual-task ERP experiment, specifically investigating whether there

[1]Paucke et al. (2015) carried out a dual-task involving two production tasks, but it is unclear how the effects from a production-production dual task would transfer to a production-comprehension dual task

were any differences in word production if the secondary task (a go/no-go task) was linguistic (syllables) compared to non-linguistic (tones). In this experiment, participants saw pictures to be named and after 300ms heard a tone or syllable. Participants were instructed to press a button if they heard a pre-specified tone or syllable (one out of five). Only no-go trials were analysed. There were three conditions: a single picture naming condition (with no secondary task), a passive naming condition (where the tone/syllable was heard but ignored), and an active naming condition (where participants responded to the picture and the tone/syllable).

The behavioural results demonstrated that with tones as distractors, naming was slower in the active and passive dual-tasks compared to the single task, and there was no difference in naming RTs between the active and passive conditions. The same pattern of results was found in the syllable condition. When naming RTs from both conditions were analysed together, there was a significant task by distractor interaction, driven by the fact that in the active condition, naming RTs in the syllable task were much longer than in the tone task, whereas in the passive condition the difference between conditions was smaller. The spatio-temporal ERP analyses showed differences between the tone and syllable task. Specifically, ERP modulations approximately 350ms before speech onset were found only in the active condition of the syllable task compared to the active tone task.

The results of Fargier and Laganaro (2016) point to a linguistic dual-task effect which occurs only in the syllable task, and not in the non-linguistic tone task. Based on the time course of this effect, the authors argue this effect likely arises at the phonological level (Indefrey & Levelt, 2004; Indefrey, 2011), evidencing that the phonological level in picture naming is affected by a concurrent linguistic task.

However, there are limitations to this study. Dual-tasking was tested at only one SOA (300ms), leaving open the question of whether earlier or later processes would be affected by a dual-task. Real words were used as syllables, which may have caused both lexical and phonological interference. In the studies presented in this chapter, we tested dual-tasking at multiple SOAs with non-word syllables.

In Experiment 1, we investigated whether participants suffered greater interference between two linguistic tasks compared to one linguistic and one non-linguistic task. As in Fargier and Laganaro (2016), participants carried out a dual-task of picture naming (task 1) and tone or syllable identification (task 2). The syllables were non-words in Dutch, and therefore had no lexical representations. Tasks were separated by SOAs of 50ms, 300ms, or 1800ms. These SOAs were chosen to allow overlap at almost all processing levels between tasks (at 50ms), overlap in phonological processing (at 300ms; similar to Fargier & Laganaro, 2016) and no overlap (at 1800ms).

We first present predictions for picture naming + tone identification. For task 1 picture naming RTs, dual-task theories make diverging predictions. If participants carry out task 1 and task 2 serially then we predicted no effects of SOA in picture naming RTs. If participants carry out task 1 and task 2 in parallel then we predicted an SOA effect in picture naming RTs, such that picture naming RTs would be slower with increased overlap between tasks. In task 2 identification RTs, we predicted an SOA effect such that tone identification RTs would decrease as SOA increased.

For picture naming + syllable identification, we made the same predictions. In task 1 picture naming RTs, we predicted no effect of SOA if the two tasks are carried out serially. If they are carried out in parallel then we predicted an SOA effect in picture naming RTs, with longer RTs with increasing task overlap. For task 2 identification RTs, we predicted an SOA effect, with smaller RTs as SOA increased.

If there is more cross-talk between picture naming and syllable identification than between picture naming and tone identification, we predicted an additional effect of identification task on RTs. If the two tasks are carried out serially then this additional effect was predicted to be present in task 2 identification RTs, such that RTs in the syllable identification task would be longer than those in the tone identification task, especially at early SOAs. If the two tasks are carried out in parallel, then we also predicted an identification task effect in task 1 picture naming RTs, such that at early SOAs picture naming RTs would be longer in the picture naming task with a secondary task of syllable identification compared to tone identification, in line with Fargier and Laganaro (2016).

In Experiment 2 we carried out the same dual-task experiment of picture naming and auditory identification, but we controlled the acoustic complexity of the stimuli and varied linguistic instruction. This manipulation was inspired by McQueen, Eisner, Burgering, and Vroomen (in press), where participants were taught new object-word associations. The words were sine-wave speech (SWS) versions of non-words (Remez, Rubin, Pisoni, & Carrell, 1981). Critically, half of the participants were told these SWS stimuli were distorted non-words, and the other half that they were computer-generated sounds. Eisner and colleagues found that participants were faster to learn associations and remembered more words if they believed the SWS stimuli were distorted words compared to computer-generated sounds. Thus, the participants' conscious perception of the sounds as linguistic or non-linguistic affected their performance. We explored whether a similar effect would be obtained in a dual-task paradigm. Would a sound interfere more with picture naming if that sound was perceived as a linguistic compared to a non-linguistic sound? If so, this would indicate that any interference effect in Experiment 1 would arise from mapping the sounds onto linguistic representations. Importantly, this also meant that the acoustic complexity of the auditory sounds was kept constant.

Experiment 1

In Experiment 1 we tested whether participants suffered greater interference measured in RTs when carrying out two concurrent linguistic tasks (picture naming and syllable identification) compared to when carrying out a concurrent linguistic and non-linguistic task (picture naming and tone identification). We tested this at three SOAs: 50ms, 300ms, and 1800ms.

Methods

Participants

41 native Dutch-speaking participants (mean age = 21 years, SD = 2.37 years, 7 males) were recruited from the Max Planck Institute for Psycholinguistics participant database. Participants received €10 for participation. Participants had normal hearing, normal or corrected-to-normal vision, and did not suffer from dyslexia, language-related disorders or neurological disorders according to self report.

Apparatus

The experiment was presented on a Iiyama HM703UT screen connected to an HP Z400 workstation running Windows 7 using the software Presentation (version 16.5, Presentation Neurobs). A Sennheiser ME64 microphone recorded participants' speech (to task 1) and a custom made button box reported button presses (to task 2). Participants listened to the auditory stimuli through Sennheiser HD 280-13 headphones.

Materials

For the picture naming task, 10 coloured photographs were selected from the set of items used in Belke (2013), which belong to carefully controlled semantic categories. We selected one image per category. Three images (*jas* (coat), *ring* (ring) & *ui* (onion)) were replaced with new images taken from a google image search due to the image being pixelated (*ring* and *ui*) or unclear (*jas*). All target names began with a different consonant with the exception of two images, *bus* (bus) and *bal* (ball). No target name contained the vowels or consonant used in the syllables.

Two sine wave pure tones were generated using Audacity (Audacity(R), 2014). The low tone was 300Hz and the high tone 800Hz. The two syllables were [a:k] and [i:k], referred to as 'aak' and 'iek'. The syllables were of the form VVC so that the syllables were maximally discriminable from their onset and the vowels roughly matched the tones in height identification: /aa/ (phonetically [a:]) is a 'low' vowel and /ie/ (phonetically [i:]) is a 'high' vowel in the vowel trapezium. This was designed to overcome any inherent stimulus-response mappings which would be present for the tone responses (i.e. left is low, right is high) which might not exist for the syllables. Participants were explicitly told about this height mapping for tones and syllables. Syllables and tones were 460ms long, with the syllabic vowel length of 'aak' being 263ms and the vowel length of 'iek' being 264ms. Both 'aak' and 'iek' were 222Hz over the entire syllable. Tones and syllables had an equal volume of 70dB.

Design

The variables of stimulus type (tone or syllable identification) and SOA (50ms, 300ms or 1800ms) were manipulated within-participant. Participants carried out 360 experimental trials (180 per stimulus type). These 180 trials were composed of 60 trials at each SOA. Within these 60 trials, each image was repeated 6 times; 3 times per tone/syllable token. Stimulus type was blocked, SOA varied within block, and block order (tone identification block followed by syllable identification block, or vice versa) was counterbalanced across participant.

Each participant received a unique input list generated using the Mix program (van Casteren & Davis, 2006) for each condition. For each input list, the following criteria were followed: a) repetitions of images were separated by at least two other images, b) for images which began with the same phoneme (*bus* and *bal*), at least two other images separated their presentation, c) the same SOA value repeated maximally 3 times, and d) the same tone/syllable repeated a maximum of five times. Participants also carried out 30 practice dual-task trials prior to the experimental trials, ordered by the same criteria.

Additionally, participants carried out 60 trials each of single tone and syllable identification. Unique input lists were created for these tasks, with the criteria that no more than 5 repetitions of the same tone or syllable were presented.

Procedure

Participants were tested individually in a sound-shielded booth. The experiment consisted of 5 parts. In part 1, participants were familiarised with the 10 images. Each image was presented one by one on screen with the name of the image printed underneath. Participants were told to first silently read the image name and then to name the image aloud. Participants were informed that these names should be used throughout the experiment. The experimenter controlled the presentation of the next image. After one cycle of the 10 images, the images were presented again without the target name written underneath. Participants named the image and the experimenter corrected the participant if they used an incorrect name (no participant needed correcting), before displaying the next image.

In part 2, participants carried out single tone or syllable identification. Each trial began with a fixation cross for 500ms. Then a tone or syllable was presented and participants pressed the left or right button on the button box to indicate which tone/syllable they heard. The left button was the response for the low tone/'aak' syllable and the right button the response for the high tone/'iek' syllable. After responding, or 1500ms in the case of no response, a blank screen was displayed for 500ms before the following trial began.

Part 3 consisted of the practice dual-task trials, followed by the critical dual-task trials with either tone or syllable identification. Each trial began with a fixation cross for 500ms before the image was displayed for 1000ms. If a button press response was recorded within the 1000ms, the picture was replaced with a blank screen. At the 50ms SOA, the tone/syllable was presented 50ms after image onset. At the 300ms SOA, the tone/syllable was presented 300ms after image onset. At the 1800ms SOA, the tone/syllable was presented 1800ms after image onset (after participants had named the image). Participants were instructed to first name the image and then respond to the tone/syllable. A blank screen was displayed for 2000ms after tone/syllable presentation before the following trial began.

Parts 4 and 5 were identical to parts 2 and 3, but with the other stimulus type (tone or syllable).

Participants took a break between each part of the experiment, and after 90 trials of the dual-task parts (parts 3 and 5). Participants were offered sweets or chocolate in the breaks. At the end of the experiment participants were fully debriefed. The entire testing session took approximately 50 minutes.

Analysis

Eight participants were excluded from the analysis (1 due to experimenter error; 2 due to carrying out tasks in the opposite order for more than 20% trials; 4 due to carrying out tasks in the opposite order for more than 20% trials in one condition; 1 due to RTs greater than 2 SDs above the group mean). Data from 33 participants remained for analysis.

For single task RTs, the first 30 trials of each block were removed as practice trials, and incorrect responses were excluded from analysis. Any RTs lower than 200ms or greater than 3000ms were excluded. A total of 2006 trials (98% data) remained for analysis.

For dual-task trials, only trials with correct responses for both task 1 (picture naming RTs) and task 2 (tone/syllable identification RTs) were analysed. Picture naming RTs were measured from picture onset, semi-automatically calculated using Praat (Boersma, 2002), and manually checked for hesitations, disfluencies, incorrect names and any other disturbances. Any production which deviated in this way from the target name was classified as incorrect and removed from analysis. In addition, trials in which the participant responded to the tone or syllable before naming the image were also removed from analysis. Identification RTs were measured from tone/syllable onset and were recorded by the presentation software. Trials where participants pressed the wrong button, or no button was recorded, were removed from analysis. The first two trials after each break were also removed. Any RTs smaller than 200ms or greater than 3000ms were removed. A total of 10832 trials (91% data) remained for dual-task analysis.

Inter-response intervals (IRI) were also calculated. This measurement is the difference between task 1 and task 2 RTs. If this difference is close to zero then this is can be evidence of response grouping (Hazeltine et al., 2006). Response grouping refers to when participants strategically withhold their response to one task to execute responses to both tasks together. The closer the IRI is to 0, the more likely it is that response grouping occurred. Response grouping causes a problem in the interpretation of dual-task data because if the IRI is very close to 0, it is unclear whether participants carried out task 1 before task 2, carried out both tasks in parallel, or even carried out task 2, followed by task 1, before executing their responses.

All RT data were analysed with linear mixed effects models using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) as implemented in R (R Core Team, 2017). The maximal random effects structure supported by the data was retained in each model; random effect structures for each model are reported in the corresponding results sections below. RTs were log transformed prior to analysis to control for skewed distributions. The continuous control variable trial was centred. The control variable block order (of dual-task blocks) and the experimental variable stimulus type (syllable or tone) were sum-to-zero contrast coded, and the experimental variable SOA was helmert contrast coded with two separate comparison contrasts. The first contrast was coded as (-0.25, -0.25, 0.5), corresponding to comparing SOAs of 50ms and 300ms with SOA 1800ms. The second contrast was coded as (-0.5, 0.5, 0), comparing SOA 50ms with SOA 300ms while ignoring SOA 1800ms. In line with the literature, we take $|t|$ greater than 2 as significant. CIs are reported and were generated using the confint function using the profile method in lme4. For post-hoc comparisons, the lsmeans package (Lenth, 2016) was used with the p value corrected for multiple comparisons using Tukey's HSD. Error data were not analysed because of the low proportion (<%5) of analysable errors (non-analysable errors included responding to task 2 before task 1, outliers, and noises in the speech recordings such as coughs and sneezes).

Results

Single identification task

In the single identification task, participants performed tone or syllable identification. The mean RTs for syllables were 444.4ms (SD = 178), and for tones were 443.7 (SD = 157). The data were analysed with a linear mixed effects model, including fixed effects of trial, stimulus type (tone or syllable) and the interaction between stimulus type and trial. The random effects structure contained random intercepts by participant and a random slope of stimulus type by participant. None of the main effects nor interaction were significant predictors, as shown in Table 3.1. We interpret this lack of significance as indicating that the processing time of tone and syllable identification is very similar, and one task is not inherently more difficult than the other[2].

Table 3.1: Experiment 1: Mixed effect model output for the single identification task.

Fixed effects	*Estimate*	*Std. Error*	*t value*	*CI*
Intercept	2.6	0.015	17.23	2.59, 2.65
Trial	-0.00025	0.00022	-1.14	-0.00069, 0.00018
Stimulus type	0.005	0.011	0.48	-0.016, 0.026
Trial*Stimulus type	-0.00019	0.00022	-0.86	-0.00063, 0.00025

Dual-task performance: Task 1 RTs

Figure 3.1 illustrates the task 1 naming RTs by stimulus type as a function of SOA. At the 50ms SOA participants were 67ms slower to name the picture with the secondary task of syllable identification than tone identification; at the 300ms SOA they were 28ms slower; at the 1800ms SOA they were 7ms slower.

A linear mixed effects model was used to analyse the data with log-transformed naming RT as the dependent variable. The fixed effect predictors were trial, block order, stimulus type (tone vs syllable), SOA (50 + 300 vs 1800; 50 vs 300), and a stimulus type by SOA interaction. The random effects structure contained random intercepts by participant and item, and random slopes for stimulus type and SOA by participant and item. The model output is displayed in Table 3.2.

There were no main effects of the control variables order or trial. Turning to the experimental variables, there was a main effect of stimulus type, meaning that RTs in the syllable condition were significantly longer than in the tone condition. RTs were longer at the 50ms and 300ms SOAs compared to the 1800ms SOA, and longer at 50ms than 300ms. The interaction between stimulus type and SOA was significant for both SOA comparisons, indicating that the effect of stimulus type was significantly larger at the 50ms and 300ms SOA than at 1800ms, and larger at 50ms than 300ms.

[2]We also carried out a pre-test to test whether the tones and syllables had equal RTs, and we found no difference between RTs in the pre-test (mixed model analysis of 16 participants; effect of stimulus type was t = -0.72).

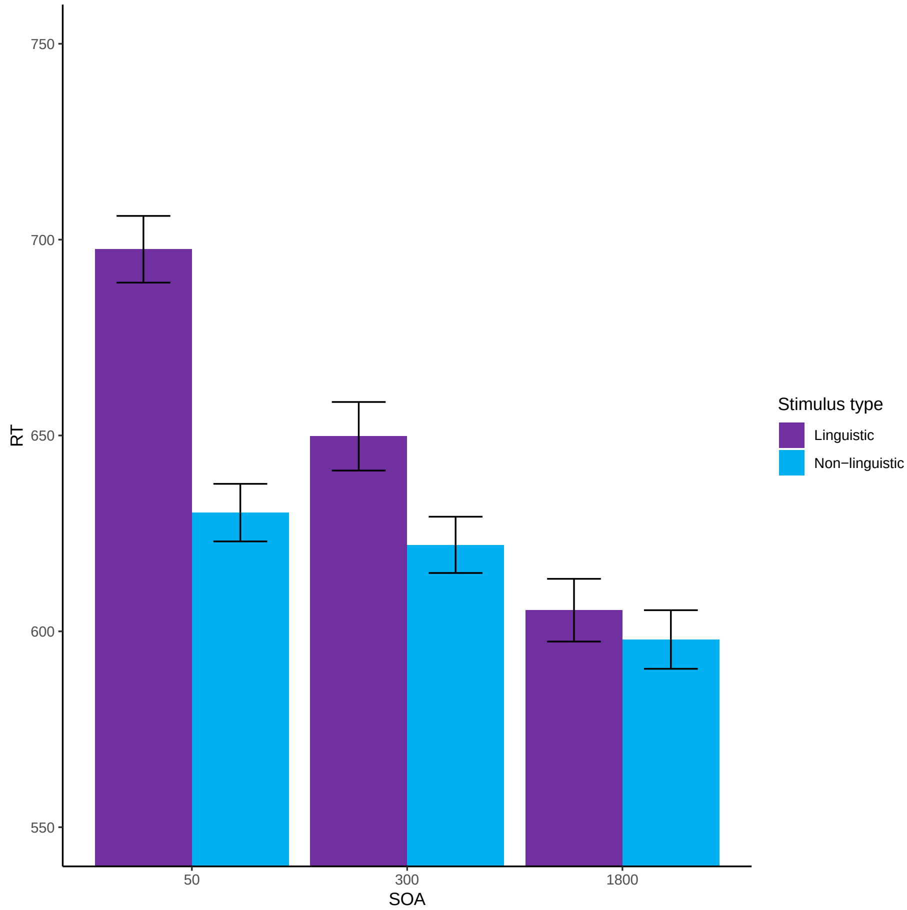

Figure 3.1: Experiment 1: RTs to task 1 (picture naming) in ms by SOA as a function of stimulus type. Error bars are within-participant 95% confidence intervals.

Dual-task performance: Task 2 RTs

Figure 3.2 illustrates the task 2 RTs by SOA as a function of stimulus type. At 50ms, participants were 116ms slower to respond to the syllables than the tones, and 77ms slower at 300ms. At the 1800ms SOA, participants responded to syllables 34ms slower than to tones.

A linear mixed effects model was used to analyse the data. The dependent variable was log-transformed RT. The fixed effect predictors were trial, block order, log-transformed task 1 RTs, stimulus type, SOA, and a stimulus type by SOA interaction. The random effects structure contained random intercepts by participant and item, random slopes for stimulus type by participant and item, and a random slope of SOA and interaction between stimulus type and SOA by participant. The model output is displayed in Table 3.3. We included the control variable log-transformed task 1 RTs as participants were instructed to respond to

Table 3.2: Experiment 1: Mixed effects model output of task 1 RTs. Shown are fixed effect estimates, SEs, t values and CIs. SOA1 is the comparison 50ms + 300ms vs 1800ms. SOA2 is the comparison 50ms vs 300ms.

Fixed effects	*Estimate*	*Std. Error*	*t value*	*CI*
Intercept	2.79	0.01	213.39	2.76, 2.81
Order	2.52e-03	2.87e-03	0.88	-2.25e-03, 9.58e-03
Trial	1.36e-06	1.54e-05	0.09	-2.89e-05, 3.16e-05
Stimulus type	1.03e-02	3.32e-03	3.12	3.82e-03, 1.68e-02
SOA1	-4.39e-02	5.16e-03	-8.51	-5.42e-02, -3.36e-02
SOA2	-2.18e-02	6.3e-03	-3.46	-3.43e-02, -9.31e-03
Stimulus type*SOA1	-1.75e-02	2.23e-03	-7.87	-2.19e-02, -1.32e-02
Stimulus type*SOA2	-1.32e-02	1.97e-03	-6.72	-1.71e-02, -9.37e-03

task 1 before task 2. Therefore, effects originating in task 1 may be found in task 2 due to response ordering. Modelling this control predictor helps to partial out some of this variance.

The control variable order was not significant. Log-transformed task 1 RT was significant, meaning that the speed of performing task 1 significantly influenced the speed of task 2. This is to be expected, as participants were instructed to respond to task 1 before task 2. Trial was also significant, indicating that participants sped up over the course of a block. There was a significant effect of stimulus type, meaning that participants responded slower to syllables than tones. RTs at the 50ms and 300ms SOAs were significantly longer than at the 1800ms SOA, and were significantly longer at the 50ms SOA than 300ms SOA. The stimulus type by SOA interaction was not significant, indicating that syllables were responded to slower than tones at all SOAs.

Dual-task performance: Inter-response intervals (IRI)

The amount of time between the response to task 1 and task 2 (IRI) was calculated for each trial separately and then averaged across conditions. Table 3.4 shows the mean IRIs by stimulus type and SOA. At 50ms and 300ms, IRIs with the syllable stimulus type were 49ms larger than with tones. At 1800ms, IRIs with the syllable stimulus type were 27ms larger than with tones. No IRI was close to 0, indicating no response grouping in this data set.

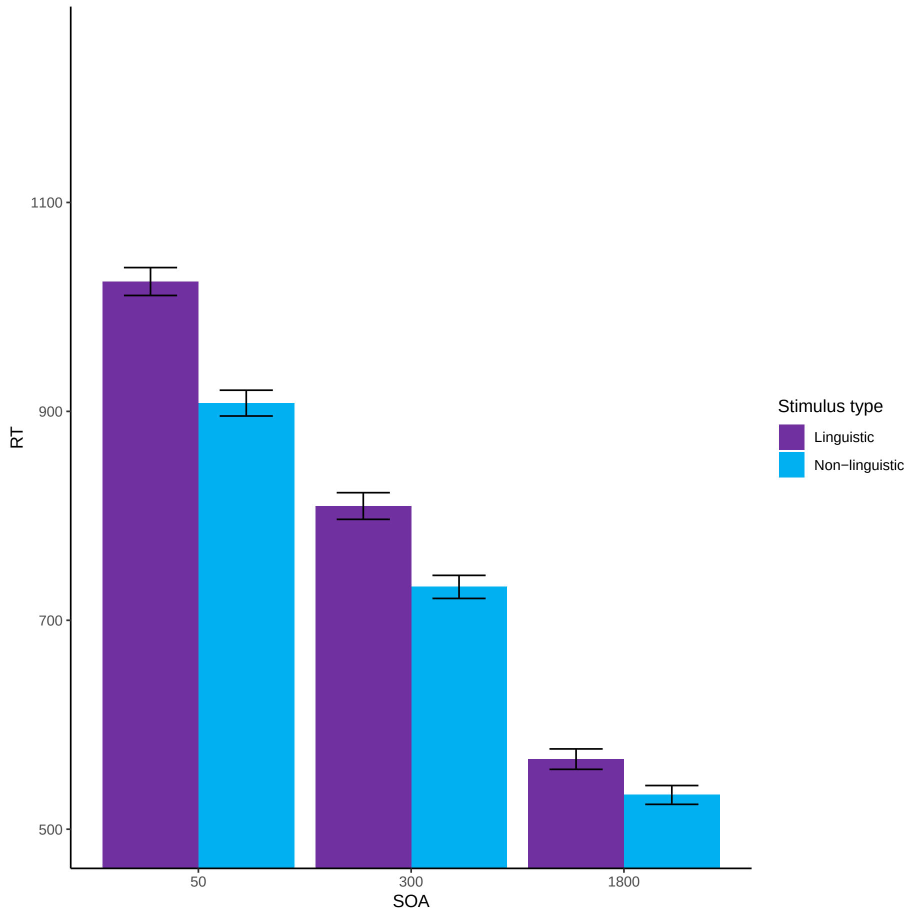

Figure 3.2: Experiment 1: RTs to task 2 (identification) in ms by SOA as a function of stimulus type. Error bars are within-participant 95% confidence intervals.

Table 3.4: Experiment 1: Mean RTs (in ms) for IRI by stimulus type and SOA.

SOA	*Stimulus type*	*Mean*	*SD*
50	Syllable	376.78	230.61
	Tone	327.61	214.85
300	Syllable	459.61	224.4
	Tone	409.9	206.32
1800	Syllable	1761.7	252.67
	Tone	1734.89	232.33

Again, a linear mixed effects model was run to determine any effects of stimulus type or SOA on the size of the IRI. The dependent variable was the log RT difference between task 1 and task 2 RTs. The fixed effect variables were trial, block order, stimulus type, SOA, and a stimulus type by SOA interaction. The

Table 3.3: Experiment 1: Mixed effects model output of task 2 RTs. Shown are fixed effect estimates, SEs, t values and CIs. SOA1 is the comparison 50ms + 300ms vs 1800ms. SOA2 is the comparison 50ms vs 300ms

Fixed effects	*Estimate*	*Std. Error*	*t value*	*CI*
Intercept	1.42	0.033	43	1.36, 1.49
Order	4.38e-03	3.37e-03	1.3	-2.79e-03, 1.15e-02
Log T1 RT	5.1e-01	1.13e-02	45.24	4.88e-01, 5.33e-01
Trial	-4.29e-05	1.82e-05	-2.36	-7.86e-05, -7.24e-06
Stimulus type	1.62e-02	3.87e-03	4.19	8.63e-03, 2.38e-02
SOA1	-2.34e-01	1.2e-02	-19.56	-2.58e-01, -2.11e-01
SOA2	-9.41e-02	5.4e-03	-17.43	-1.05e-01, -8.34e-02
Stimulus type*SOA1	-7.75e-03	5.35e-03	-1.45	-1.84e-02, 2.89e-03
Stimulus type*SOA2	1.8e-03	2.87e-03	0.63	-3.88e-03, 7.54e-03

random effects structure contained random intercepts by participant and item, and random slopes of stimulus type and SOA by participant and by item. The model results are displayed in Table 3.5. Main effects of SOA are not discussed as they are meaningless in this analysis.

We found that the control variables order and trial were not significant. There was a significant effect of stimulus type, meaning that in general IRIs were longer in the syllable condition compared to the tone condition. The interaction between stimulus type and the SOA contrast of 50ms and 300ms compared to 1800ms was significant, indicating that the IRIs were larger in the syllable condition at 50ms and 300ms than the tone condition, with a smaller difference at 1800ms. The interaction between stimulus type and SOA, where the SOA contrasts 50ms and 300ms, was not significant, as the syllable IRIs were equally larger than the tone IRIs at both SOAs.

Discussion

We found SOA effects in both task 1 naming RTs and task 2 identification RTs, where RTs were longer in both tasks with shorter SOAs. This indicates parallel processing of the two tasks, such that participants distributed their capacity across tasks. More importantly, we found main effects of stimulus type (syllable or tone) in both task 1 and task 2 RTs. Responses were faster in both tasks when the secondary task was tone identification compared to syllable identification. This indicates cross-talk between the tasks which was stronger for two linguistic

Table 3.5: Experiment 1: Mixed effects model output of IRIs. Shown are fixed effect estimates, SEs, t values and CIs. SOA1 is the comparison 50ms + 300ms vs 1800ms. SOA2 is the comparison 50ms vs 300ms.

Predictor	*Estimate*	*Std. Error*	*t value*	*CI*
Intercept	2.75	1.98e-02	139.04	2.71, 2.79
Order	9.1e-03	6.31e-03	1.44	-0.0037, 0.022
Trial	-3.58e-05	3.85e-05	-0.93	-0.00011, 3.97e-05
Stimulus type	2.18e-02	6.71e-03	3.26	0.0087, 3.5e-02
SOA1	9.84e-01	3.68e-02	26.74	0.91, 1.06
SOA2	1.5e-01	1.68e-02	8.94	0.12, 0.18
Stimulus type*SOA1	-3.86e-02	5.56e-03	-6.94	-0.049, -2.77e-02
Stimulus type*SOA2	-8.03e-03	4.92e-03	-1.63	-0.018, 1.61e-03

tasks. For task 1 picture naming RTs, there was also an interaction between SOA and stimulus type, indicating that the effect of cross-talk was strongest at earlier SOAs (i.e. with greater task overlap). This pattern of results suggests that concurrently identifying a syllable interferes more with picture naming than concurrently identifying a tone.

In this experiment, tones and syllables did not only vary in their linguistic nature but also in their acoustic complexity. Therefore, a possible non-linguistic explanation for the results of Experiment 1 is that the different acoustic properties of tones and syllables drove this effect. Specifically, a syllable is more acoustically complex than a tone. Listening to irrelevant sounds disrupts memory recall for to-be-remembered items (Divin, Coyle, & James, 2001; Elliott, 2002; Jones, Madden, & Miles, 1992), and memory recall is disrupted to a greater extent for changing state sounds, such as syllables, than for steady state sounds, such as tones (Gabriel et al., 2012; Klatte, Lachmann, Schlittmeier, & Hellbrück, 2010). Experiment 2 therefore explored whether dual-task interference would depend on participants' perception of the auditory stimuli as being linguistic or non-linguistic, while keeping the acoustic complexity of the sounds constant.

Experiment 2

In Experiment 2 we used the same paradigm as Experiment 1 but created sine-wave speech (SWS) versions of the syllables. SWS stimuli are created by tracking the first three (or more) formants of the speech sound, and generating these formants with sine waves. The speech typically sounds like whistles, or robot noises. A large body of research into SWS has demonstrated that it can be perceived as speech or non-speech depending on the knowledge of the person listening (Dehaene-Lambertz et al., 2005; Liebenthal, Binder, Piorkowski, & Remez, 2003; Möttönen et al., 2006; Remez et al., 1981; Remez, Pardo, Piorkowski, & Rubin, 2001; Tremblay, Nicholls, Alford, & Jones, 2000). When participants are told which language sound the SWS came from, they can hear that linguistic token, but tend to perceive SWS as noise otherwise.

We created SWS versions of the syllables used in Experiment 1 ('aak' and 'iek'). One group of participants were told the SWS sounds were computer-generated sounds, and the other group were told that the SWS sounds were distorted forms of the syllables 'aak' and 'iek'. With this manipulation we sought to determine whether knowledge of the linguistic nature of the auditory stimulus would create the same pattern of effects as in Experiment 1, whilst keeping the acoustic complexity of the auditory stimulus constant (as the same sounds were presented to each group of participants).

Methods

Participants

65 native Dutch-speaking participants (mean age = 22.4 years, SD = 2.67 years, 15 males) were recruited from the Max Planck Institute for Psycholinguistics participant database. Participants received €8 for participation. Participants had normal hearing, normal or corrected-to-normal vision, and did not suffer from dyslexia, any language-related disorders or neurological disorders according to self report.

Apparatus

The experiment was presented on a Benq monitor using the software Presentation (version 16.5, Presentation Neurobs). A Sennheiser microphone recorded participants' speech (to task 1) and a custom made quiet button box (created at the MPI) recorded button presses (to task 2). Participants listened to the auditory stimuli through Sennheiser HD437 headphones.

Materials

For the picture naming task, the same materials were used as in Experiment 1.

SWS versions of the syllables 'aak' and 'iek' (used in Experiment 1) were created in Praat (using the script from Darwin (2005) with three tracked formants). Pilot testing of the two generated SWS sounds found that the 'iek' sound was ambiguous, but the 'aak' sound was too different from the original 'aak' syllable and was unambiguously *not* language. This was because the automatic SWS generation of 'aak' had raised the pitch of the sound. In order to create two SWS stimuli which were equally ambiguous between computer-sound and speech, we generated 16 SWS versions of 'aak' and 'iek' with manipulated pitch[3]. The pitch range of these sounds overlapped.

Each of the versions were presented to 11 native Dutch speakers with a 7-point rating scale (ranging from 1: 'this sounds nothing like language' to 7: 'this sounds exactly like aak/iek'), and participants rated each sound. The range of values for the SWS versions of 'aak' was 2.18 - 3.36. The range of values for the SWS versions of 'iek' was 3.54 - 4.5. Because the values for the 'iek' versions were generally higher than the 'aak' versions, we took the item from each type with the most similar average rating: 3.36 for 'aak' and 3.54 for 'iek'. A paired samples t-test of the ratings for these values found no significant difference between them ($t(10) = 0.28$, $p = 0.78$). Thus, these two SWS tokens were used in the experiment.

Design

The variable instruction (whether participants were in the computer-generated sound or distorted syllable group) was manipulated between-participant. SOA (50ms, 300ms and 1800ms) was manipulated within-participant. Participants carried out 180 experimental trials, composed exactly as in Experiment 1. Each participant received a unique input list with the same criteria as Experiment 1. Participants also carried out 30 practice dual-task trials prior to the experimental trials.

Participants carried out 80 trials each of the single task. Unique input lists were created for these tasks, with the same criteria as Experiment 1.

Procedure

The testing procedure was identical to Experiment 1, except that instruction was tested between-participants so participants only carried out one dual-task block. Depending on the instruction condition, the sounds were always referred to as either computer-generated sounds or the distorted syllables 'aak' and 'iek'. Par-

[3]We thank Joe Rodd for his assistance with this.

ticipants were also reminded before beginning the dual-task to name the picture before pressing the button, to minimise data loss.

At the end of the experiment, participants filled in a questionnaire to check that the instruction manipulation was successful, before being fully debriefed. The entire testing session took approximately 40 minutes.

Analysis

Six participants were removed (1 due to more than 20% inaccurate trials in the single identification task; 1 due to RTs more than 2 SDs above the group mean for task 1; 1 due to equipment error; 1 due to responding to task 2 before task 1 in more than 20% trials; 2 due to having more than 20% analysable trials removed after removal of all errors). Data from 59 participants remained for analysis after exclusion; 29 in the computer-generated condition (non-linguistic) and 30 in the distorted syllable condition (linguistic).

For single task RTs, the first 20 trials of each block were removed as practice trials. The same error criteria as Experiment 1 were used. A total of 3478 trials (98% data) remained for analysis. For dual-task trials, the same removal criteria were followed as in Experiment 1. 9850 trials (95% data) remained for analysis. All data were analysed in the same way as in Experiment 1. Errors were not analysed due to the low proportion of analysable errors ($< 5\%$).

Results

Single identification task

In the single identification task, participants discriminated between the sounds after being told they were computer-generated sounds (non-linguistic) or distorted syllables (linguistic). The mean RTs to distorted syllables were 518ms (SD = 134), and to computer-generated sounds were 478ms (SD = 135). The data were analysed with linear mixed effects models, with fixed effects of trial, instruction (distorted syllables or computer-generated sounds) and a trial by instruction interaction. The random effects structure contained random intercepts by participant. Model results are reported in Table 3.6.

There was a significant effect of trial, such that participants sped up through the experiment, and a significant trial by instruction interaction, as participants sped up earlier when identifying computer-generated sounds compared to distorted syllables. There was no main effect of instruction, providing evidence that

RTs were equal with linguistic or non-linguistic instruction (though this effect was marginal).

Table 3.6: Experiment 2: Mixed effects model analysis of single task identification RTs.

Fixed effects	*Estimate*	*Std. Error*	*t value*	*CI*
Intercept	2.68	9.30e-03	288.27	2.66, 2.7
Trial	-6.67e-04	8.14e-05	-8.19	-8.26e-04, -0.00051
Instruction	1.79e-02	9.3e-03	1.92	-3.34e-04, 0.036
Trial*Instruction	2.31e-04	8.14e-05	2.84	7.14e-05, 0.00039

Dual-task performance: Task 1 RTs

Figure 3.3 displays the task 1 naming RTs. Participants were numerically faster to name pictures with a secondary task of distorted syllable identification than computer-generated sound identification: 22ms faster at the 50ms SOA, 7ms faster at the 300ms SOA, and 18ms faster at the 1800ms SOA. These numerical differences were smaller than and in the opposite direction to task 1 RTs in Experiment 1.

A linear mixed effects model was used to analyse the data. The dependent variable was log-transformed RTs. The fixed predictors were trial, instruction (linguistic vs non-linguistic), SOA, and an instruction by SOA interaction. The random effects structure contained random intercepts by participant and item, and random slopes for SOA by participant and item. Model results are reported in Table 3.7.

There was a significant effect of trial such that participants sped up across blocks. There was no effect of instruction, indicating that picture naming RTs did not vary by whether participants were told the secondary task was linguistic or not. There was a main effect of SOA, showing that task overlap at the 50ms and 300ms SOAs lead to slower RTs than no task overlap at the 1800ms. There was no difference between RTs at the 50ms and 300ms SOAs. Instruction and SOA did not interact, meaning RTs were the same regardless of instruction at each SOA.

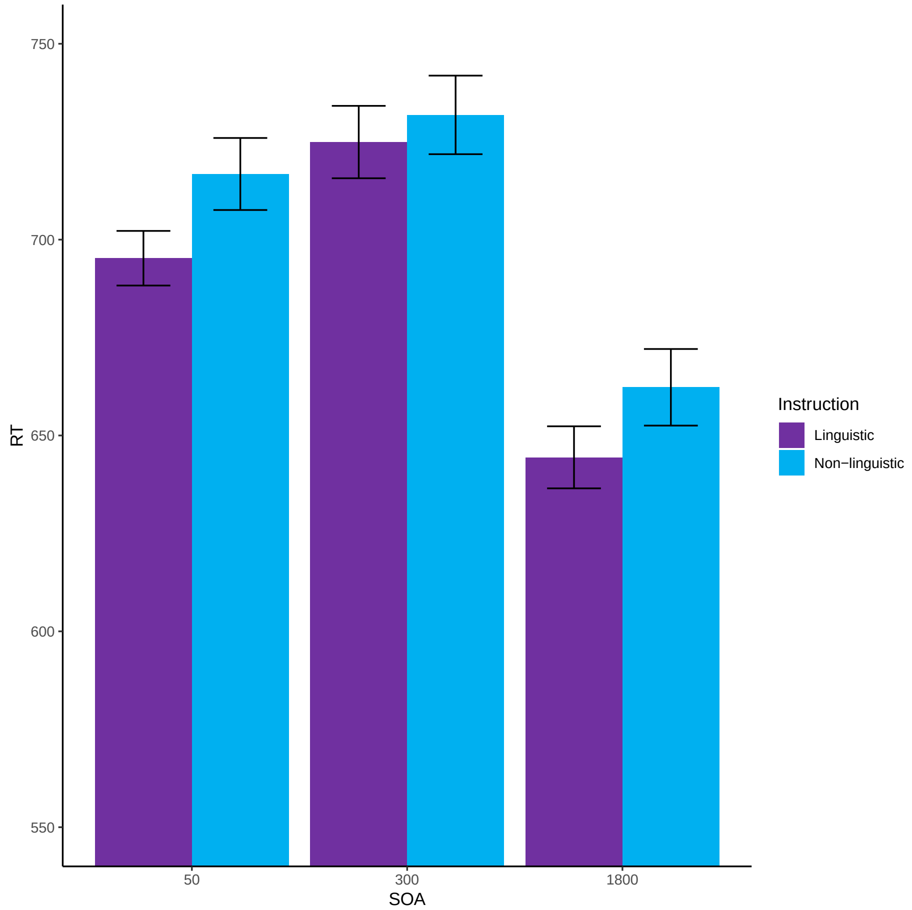

Figure 3.3: Experiment 2: Task 1 RTs in ms by SOA as a function of instruction. Error bars are between-participant 95% confidence intervals.

Dual-task performance: Task 2 RTs

Figure 3.4 displays the task 2 RTs. Participants were numerically slower with linguistic instruction compared to non-linguistic instruction at SOAs of 50ms and 300ms: 57ms slower at 50ms, and 65ms slower at 300ms. At 1800ms, participants were 9ms faster to respond with linguistic instruction than non-linguistic instruction.

The linear mixed effects model had log-transformed RTs as the dependent variable. The fixed predictors were trial, log-transformed task 1 RTs, instruction, SOA, and an instruction by SOA interaction. The random effects structure contained random intercepts by participant and item, and a random slope for SOA by participant. Results from the model are displayed in Table 3.8.

The control variable log-transformed task 1 RTs was significant, meaning that the speed that people responded to task 1 significantly predicted their task 2 RTs. This is unsurprising as participants were instructed to respond to task 1 before

Table 3.7: Experiment 2: Mixed effects model output of task 1 RTs. Shown are fixed effect estimates, SEs, t values and CIs. SOA1 is the comparison 50ms + 300ms vs 1800ms. SOA2 is the comparison 50ms vs 300ms.

Fixed effects	*Estimate*	*Std. Error*	*t value*	*CI*
Intercept	2.83	8.68e-03	326.11	2.81, 2.85
Trial	-1.17e-04	1.55e-05	-7.54	-0.00015, -0.00009
Instruction	-4.09e-03	7.6e-03	-0.54	-0.019, 0.011
SOA1	-5.48e-02	5.27e-03	-10.41	-0.065, -0.044
SOA2	8.97e-03	4.99e-03	1.8	-0.00084, 0.019
Instruction*SOA1	-5.17e-04	4.67e-03	-0.11	-0.0097, 0.0086
Instruction*SOA2	4.13e-03	4.51e-03	0.92	-0.0047, 0.013

responding to task 2. The control variable trial was also significant, indicating that participants slowed down through a block.

There was no main effect of instruction, meaning that RTs with linguistic or non-linguistic instruction were similar. There were main effects of SOA, meaning that RTs at 50ms and 300ms were significantly slower than RTs at 1800ms, and RTs at 50ms were significantly slower than RTs at 300ms. The interaction between instruction and SOA was not significant, indicating that identification RTs were similar in each task at each SOA.

Table 3.8: Experiment 2: Mixed effects model output of task 2 (identification RTs). Shown are fixed effect estimates, SEs, t values and CIs. SOA1 is the comparison 50ms + 300ms vs 1800ms. SOA2 is the comparison 50ms vs 300ms.

Fixed effects	*Estimate*	*Std. Error*	*t value*	*CI*
Intercept	1.67	3.84e-02	43.41	1.59, 1.74
Trial	5e-05	2.04e-05	2.45	9.96e-06, 9.01e-05
Log T1 RT	4.57e-01	1.32e-02	34.76	4.32e-01, 4.83e-01
Instruction	1.15e-02	8.75e-03	1.32	-5.63e-03, 2.87e-02
SOA1	-2.27e-01	8.58e-03	-26.52	-2.44e-01, -2.11e-01
SOA2	-6.81e-02	3.41e-03	-19.99	-7.48e-02, -6.14e-02
Instruction*SOA1	-1.6e-02	8.55e-03	-1.87	-3.27e-02, 7.8e-04
Instruction*SOA2	1.5e-03	3.41e-03	0.44	-5.17e-03, 8.16e-03

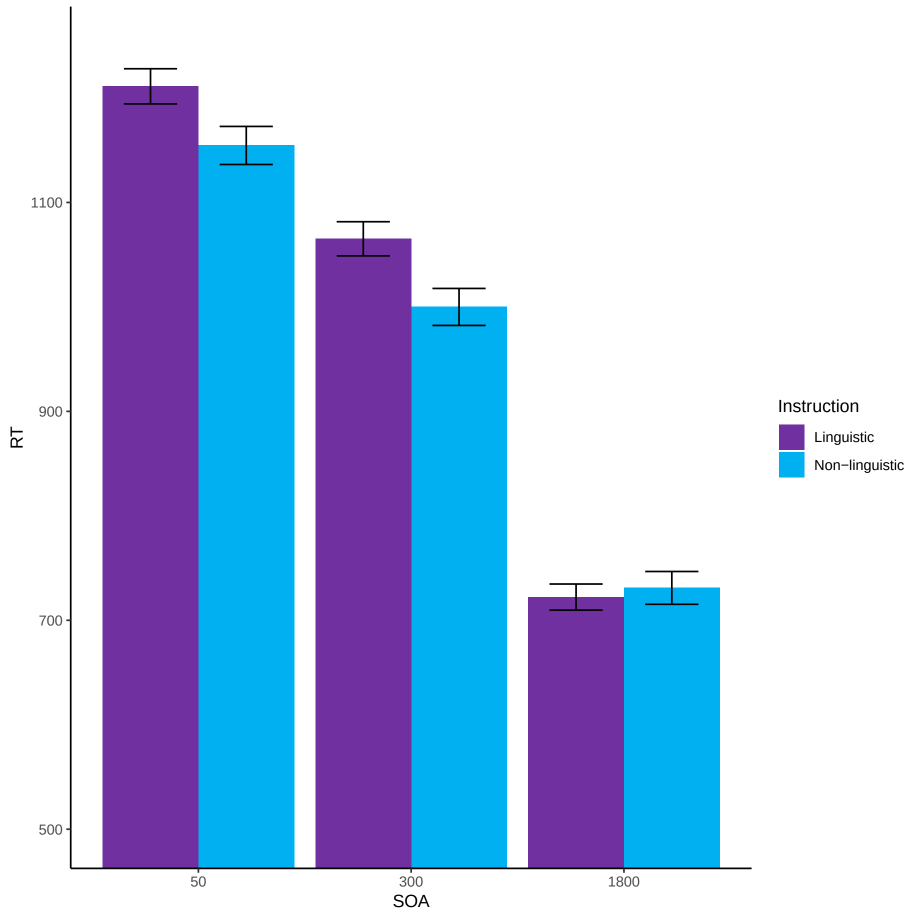

Figure 3.4: Experiment 2: Task 2 RTs in ms by SOA as a function of instruction. Error bars are between-participant 95% confidence intervals.

Dual-task performance: Inter-response intervals (IRI)

As with Experiment 1, we also analysed the IRIs to see if we found evidence of response grouping in either condition, and to determine whether there was an instruction effect in the IRIs. Table 3.9 shows the mean IRIs by instruction and SOA. IRIs in the linguistic instruction condition were larger at all SOAs than in the non-linguistic instruction condition: 78ms at the 50ms, 72ms at the 300ms SOA, and 9ms at the 1800ms SOA.

Table 3.9: Experiment 2: Mean IRIs (in ms) by instruction and SOA.

SOA	*Instruction*	*Mean*	*SD*
50	Linguistic	565.92	311.24
	Non-linguistic	487.80	294.26
300	Linguistic	640.29	299.50
	Non-linguistic	568.14	294.29
1800	Linguistic	1877.80	278.50
	Non-linguistic	1868.71	332.36

A linear mixed effects model used log-transformed IRIs as the dependent variable. The fixed effect predictors were trial, instruction, SOA, and an instruction by SOA interaction. The random effects structure contained random intercepts by participant and item, and a random slope for SOA by participant. The model output is displayed in Table 3.10. As with Experiment 1, we have not discussed SOA effects.

There was a significant effect of trial such that IRIs increased within blocks. There was a main effect of instruction, indicating that IRIs were larger when the secondary task was linguistic compared to non-linguistic. The interaction between instruction and SOA was significant, indicating that the IRI was larger with linguistic instruction than non-linguistic instruction at 50ms and 300ms, but not at 1800ms. However, there was no difference of instruction in IRIs between the 50ms and 300ms SOAs.

Table 3.10: Experiment 2: Mixed effects model output of IRIs. Shown are fixed effect estimates, SEs, t values and CIs. SOA1 is the comparison 50ms + 300ms vs 1800ms. SOA2 is the comparison 50ms vs 300ms.

Fixed effects	*Estimate*	*Std. Error*	*t value*	*CI*
Intercept	2.88	1.4e-02	205.48	2.85, 2.91
Trial	1.47e-04	3.51e-05	4.18	7.79e-05, 0.00022
Instruction	2.83e-02	1.23e-02	2.3	4.19e-03, 0.052
SOA1	7.77e-01	2.05e-02	38	7.37e-01, 0.82
SOA2	8.20e-02	8.62e-03	9.52	6.51e-02, 0.099
Instruction*SOA1	-5.28e-02	2.05e-02	-2.58	-9.28e-02, -0.013
Instruction*SOA2	-1.38e-02	8.62e-03	-1.6	-3.07e-02, 0.0031

Discussion

We found SOA effects in both task 1 and task 2 RTs, indicating parallel processing of the two tasks. This indicates, as in Experiment 1, that participants distributed their capacity across both tasks. We found null results of instruction in both task 1 naming RTs and task 2 identification RTs. This indicates that whether or not the SWS sounds were perceived as speech or non-speech did not affect their processing or how strongly they interfered with picture naming.

One account for this pattern is that our manipulation did not work and participants did not differ in how they perceived the SWS sounds. However, post-test questionnaire data argue against this hypothesis. The post-test questionnaire indicated that when participants were told the sounds were computer-generated, most of them did not spontaneously hear language sounds. In fact, when these participants were informed that the sounds were generated from syllables many were surprised. Participants in the opposite group - the distorted syllable group - indicated that they could hear that the sounds came from the syllables 'aak' and 'iek'. Some of these participants, when told that in the other condition participants were told the sounds were computer-generated, remarked that they thought people would hear the syllables. Additionally, in an analysis of data from participants who confidently believed the manipulation, we found the same pattern of results as when analysing all participants. This suggests that our participants did perceive the auditory stimuli as instructed.

We also found an effect of instruction in the IRI responses, where the IRI was larger with linguistic instruction than non-linguistic instruction. A longer IRI could indicate that it was more challenging to disengage from task 1 processing and engage in task 2 processing. As this effect is greater when the two tasks are instructed as linguistic, this would suggest that approaching a dual linguistic task causes some interference that cannot be attributed to acoustic complexity.

However, there was a large amount of variation in the Experiment 2 data. For example, in task 2 RTs we found numerical differences of 57ms and 65ms between instruction conditions at the 50ms and 300ms SOAs. These effects were not significant, though numerically large when regarded alongside the effects in Experiment 1. This is likely because Experiment 2 involved a between-participant manipulation instead of a within-participant manipulation, as in Experiment 1. However, this variation could have masked true linguistic effects in the data. Future research should test the effect of instruction in a within-participant design to increase power.

General discussion

In this study, we tested whether participants would experience greater interference when carrying out a dual-task composed of two linguistic tasks versus a dual-task composed of one linguistic task and one non-linguistic task. In Experiment 1 we found a main effect of stimulus type (syllable or tone auditory stimuli) in both task 1 naming RTs and task 2 identification RTs. We additionally found an SOA by stimulus type interaction in task 1 RTs, such that syllable identification interfered more with picture naming than tone identification at earlier SOAs. These results suggested a linguistic interference account, where two linguistic tasks interfere with one another and affect responses in both tasks, which is greater than typical dual-task interference.

In Experiment 2 we tested whether the effect found in Experiment 1 was due to the linguistic nature of the auditory sounds, while keeping their acoustic complexity constant. We found null effects of linguistic instruction (whether participants were told that the auditory stimuli were linguistic or not) in task 1 and task 2 RTs. These results cast doubt on the linguistic interference hypothesis proposed to account for the interference effect in Experiment 1. We first discuss the results of Experiment 1 in more depth and then turn to the results of Experiment 2.

In Experiment 1 we found a main effect of stimulus type in both task 1 picture naming RTs and task 2 identification RTs, such that responses were longer in both tasks if the secondary task was syllable identification compared to tone identification. This indicates cross-talk between the tasks that increased when the two tasks were linguistic compared to when one was linguistic and the other a non-linguistic task (Wickens, 2008). This effect was qualified by an interaction between SOA and stimulus type in task 1 RTs, such that at early SOAs picture naming RTs were longer when the secondary task was syllable identification compared to tone identification. This effect was greatest at the 50ms SOA, smaller but still significant at the 300ms SOA, and smallest at the 1800ms SOA. These results demonstrate greater cross-talk with maximal overlap, and indicate that both task responses were affected when the tasks were more similar.

Taken alone, the results of Experiment 1 suggest interference was greater between two linguistic tasks than one linguistic and one non-linguistic task (Fargier & Laganaro, 2016). Since our syllables were non-words, any specific interference that arose was likely at a phonological level. Fargier and Laganaro (2016) observed a strong linguistic cross-talk effect at an SOA of 300ms. Our strongest effect was at the 50ms SOA (a 67ms effect); at the 300ms SOA our effect was smaller at only 28ms. Since Fargier and Laganaro (2016) tested only one SOA, we cannot compare the time course of effects between studies. However, we note that the naming RTs were faster in our study (579ms at the 1800ms SOA) compared to Fargier & Laganaro (831ms in the single naming condition), potentially due to differences in the visual stimuli (coloured photographs vs black and white line drawings). Therefore, phonological encoding of the picture names may have begun earlier than in Fargier and Laganaro (2016). This would explain why our strongest effect was seen earlier than expected based on Fargier and Laganaro (2016).

However, the linguistic interference account of the data is thrown into doubt by the results of Experiment 2. In Experiment 1, the linguistic nature and acoustic complexity of the stimuli were confounded. We could not specify which of the two was responsible for the cross-talk effect in Experiment 1. The linguistic hypothesis seems more plausible than acoustic complexity, as we can specify how this effect would arise. Specifically, phonologically encoding the syllable would interfere with phonological encoding of the picture name, giving rise to an interference effect (cf. Fargier & Laganaro, 2016). In contrast, it is difficult to say how acoustic complexity would impact planning of the picture name, especially as the tones and syllables did not appear to differ in difficulty (as shown by the pre-test and the single identification task results from Experiment 1).

In Experiment 2 we kept the acoustic complexity of the auditory stimuli constant by generating SWS versions of the syllables 'aak' and 'iek', and presented them in a between-participant condition of linguistic instruction. One group of participants were told the stimuli were computer-generated sounds, and hence non-linguistic. The other group of participants were told the sounds were distorted versions of the syllables 'aak' and 'iek', and hence linguistic. We tested whether participants' perception of the stimuli as linguistic or non-linguistic would affect linguistic interference. We found no significant effects of instruction in task 1 RTs or task 2 RTs.

Therefore, the results from Experiment 2 do not provide additional evidence for the linguistic interference hypothesis as there were no clear differences between participants in either instruction group. The null result does not allow us to draw strong conclusions about linguistic interference or acoustic complexity.In sum, when keeping acoustic complexity constant, we found no additional effects of linguistic interference in this paradigm.

In both experiments we found SOA effects in task 1 RTs. This indicates that the tasks were processed in parallel in both experiments. This finding supports a capacity-sharing account of dual-tasking (Navon & Miller, 2002; Tombu & Jolicœur, 2003). Importantly, an SOA effect shows that the different pattern of results between experiments cannot be explained by participants processing the tasks serially in one experiment and in parallel in the other.

In conclusion, we can only conclude that syllables interfered more with picture naming than tones when presented with overlapping SOAs between tasks. Similarly, syllable identification was hindered more by concurrent naming than tone identification. Whether this pattern was due to cross-talk between linguistic representations in two tasks, or due to acoustic differences between syllables and tones, remains to be seen.

4 No linguistic interference in a linguistic dual-task when acoustic complexity is controlled

Abstract

Evidence from Chapter 3 suggests that when dual-tasking with two linguistic tasks, there is additional linguistic interference additive to general dual-task interference. However, the linguistic nature of the auditory stimuli was confounded with acoustic complexity, where the linguistic auditory stimuli (syllables) were more acoustically complex than the non-linguistic auditory stimuli (tones). In the current experiment, participants carried out a similar dual-tasking experiment as in Chapter 3, but importantly, the acoustic complexity of the auditory stimuli was kept constant by presenting participants with acoustically matched linguistic and non-linguistic sounds (vowels and musical rain) for the identification task. No additional linguistic interference was found in the dual-task results in either task 1 (picture naming) or task 2 (sound identification). These results demonstrate that, in this paradigm, there is no additional linguistic interference between two linguistic tasks when the acoustic complexity of the task 2 sounds is controlled. We make suggestions for future studies to directly test acoustic complexity, and speculate on the possible ways acoustic complexity may affect picture naming.

Introduction

There is recent evidence that while people comprehend speech, they can also plan their own speech (e.g., Bögels et al., 2015; Barthel et al., 2016). While this suggests that speech planning and comprehension proceed concurrently, there may be constraints on this process. Here, we tested whether people can carry out two linguistic tasks concurrently, and whether the interference found when carrying out two linguistic tasks was greater than in a dual-task involving one linguistic and one non-linguistic task. Importantly, we controlled the acoustic complexity of the secondary task, such that the linguistic and non-linguistic auditory stimuli for task 2 had equal acoustic complexity. These experiments follow on from those in Chapter 3.

In Chapter 3 we described a dual-task experiment involving picture naming (as task 1) and tone or syllable identification (as task 2). We found that picture naming RTs were longer when dual-tasking with syllables than tones at SOAs resulting in task overlap (50ms and 300ms). Identification RTs to syllables in task 2 were also longer at all SOAs (50ms, 300ms and 1800ms) than those to tones. Importantly, the tones and syllables were pretested to ensure equal difficulty in the task, meaning that this effect could not be due to different levels of difficulty in processing syllables compared to tones. These findings suggested that there was linguistic interference between the tasks that was additive to typical dual-task interference. Given the nature of the stimuli, we speculated that this interference arose at the phonological level (cf. Fargier & Laganaro, 2016).

However, an alternative interpretation of these results is that the difference in acoustic complexity of the sounds resulted in greater interference in the dual-task. In Experiment 1 in Chapter 3, acoustic complexity and linguistic nature were confounded. In Experiment 2 in Chapter 3, these factors were no longer confounded. Participants were presented with sine-wave speech versions of the syllables used in Experiment 1. One group of participants was told the sounds were computer-generated, and the other group were told the sounds were distorted syllables. Thus, the linguistic manipulation was tested between-participants. Participants carried out the same dual-task of picture naming and sound identification. While the results were suggestive of the linguistic interference effect disappearing when acoustic complexity was controlled, there was a large amount of variation in RTs. Therefore, whether linguistic interference is found when acoustic complexity is controlled for is still unclear.

In the current study, participants were tested in two sessions. Acoustic complexity was kept constant across the experiment sessions. In session 1, participants carried out a size judgement task on pictures as task 1, and a sound identification task, where one of two sounds was played on each trial, as task 2. In session 2, participants carried out picture naming with the same pictures as task 1, and the same sound identification task as task 2. In both sessions we tested dual-tasking with two SOAs, at 50ms and 1800ms, resulting in task overlap or task switching.

The size judgement task is a non-linguistic task (e.g., Paucke et al., 2015). This task requires access to conceptual information to judge the size of an image, but no access to lexical or phonological information of a picture name (Jescheniak, Schriefers, Garrett, & Friederici, 2002). The picture naming task is, by definition, linguistic, and involves conceptual, lexical, phonological and articulatory processing to produce a picture name. Therefore, in session 1 task 1 was a non-linguistic task, and in session 2 task 1 was a linguistic task.

The sound identification task had a linguistic condition: vowel identification, and a non-linguistic condition: 'musical rain' identification (Uppenkamp et al., 2006). Musical rain is created by taking a sound (such as a vowel) and randomising the formant values of that sound, resulting in a sound with a musical quality. Because the vowels in this experiment were synthesised [1] and the musical rain sounds were created from these synthesised vowels, the sounds were acoustically matched. By being acoustically matched, the vowel and musical rain sounds shared (1) the same temporal characteristics, (2) the same energy profiles, and (3) the same long-term average spectrum. The sound types only differed in their short-term temporal make up, with the musical rain sounds having random variation in the carrier frequency of each of the four damped sinusoids at the start of each cycle. With this definition of acoustic complexity, we argue that the syllables used in Experiment 1 of Chapter 3 differed from the tones in acoustic complexity because the tones were made of one single spectral component (a single frequency pure sine wave) with stable temporal dynamics. We chose a musical rain acoustic manipulation because other typical ways of manipulating speech, such as noise-vocoded or rotated speech (Scott, Blank, Rosen, & Wise, 2000), were not effective on the single syllables (the syllables still sounded like linguistic stimuli). We replaced the syllable identification task as used in Chapter 3 with

[1]The vowels are not strictly vowels as they were not recorded by a speaker, but were generated by sine waves. When stacked these sine waves sounded vowel-like.

vowel identification because it was not possible to create acoustically matched musical rain versions of the syllables.

Results from the two sessions allowed us to assess whether the interference we found previously in naming latencies was due to two linguistic tasks being carried out concurrently, or whether the effect would disappear if the acoustic complexity of the auditory stimuli were matched. In these two sessions we thus manipulated whether task 1 was linguistic (across session) or whether task 2 was linguistic (within session). Importantly, the structure of both sessions was almost identical. The same images were used in the task 1 size judgement task (session 1) and the task 1 picture naming task (session 2). The same vowels and musical rain sounds were used in both sessions. The trial structure of both experiments was the same. The only difference was the task to be performed on the task 1 pictures; in session 1, participants pressed buttons in response to a pre-defined question of 'Does this item fit into a shoebox?', and in session 2 participants named the images aloud.

The order of the sessions was fixed such that participants always carried out the size judgement task before the naming task, to minimise the likelihood that participants would retrieve the picture name during the size judgement task, rendering it a linguistic task. During session 1, the experimenter was careful to never name the images. This means that session 1 combined a non-linguistic task 1 with a linguistic and non-linguistic task 2, and session 2 combined a linguistic task 1 with a non-linguistic and linguistic task 2.

We predicted different patterns of results for the two sessions. In session 1, linguistic tasks were not combined. Therefore, in session 1, we predicted no main effect of the auditory stimulus type (vowels or musical rain), and no interaction between stimulus type and any other variable. However, for session 2, two linguistic tasks were combined: picture naming and vowel identification. Therefore, in session 2 we expected to find the same pattern of results as in Experiment 1 in Chapter 3: an effect of stimulus type on picture naming RTs (an interaction between stimulus type and SOA in task 1 RTs), and longer identification RTs to vowels compared to musical rain (a main effect of stimulus type in task 2 RTs). However, if no effect of stimulus type was found in session 2 responses, this would suggest that the stimulus type effect found in Experiment 1 of Chapter 3 may have been related to the acoustic complexity of the auditory stimuli rather than their linguistic nature.

Experiment

Participants took part in two sessions. In session 1, they carried out the dual-task of a size judgement task on an image (*Does this item fit into a shoebox?*) as task 1, and vowel or musical rain (MR) identification as task 2. In session 2, they carried out picture naming as task 1, and vowel or MR identification as task 2. Importantly, across sessions participants carried out the size judgement task and naming task on the same pictures, and heard the same sounds. We sought to replicate the previous finding of increased interference in the dual linguistic condition (picture naming + vowel identification) compared to the dual-non-linguistic condition (picture naming + MR identification). We contrasted this with interference in the size judgement + identification conditions, where we predicted no difference in dual-task interference when dual-tasking with the size judgement task and either identification task.

Participants

50 right-handed native Dutch speakers (mean age = 22, SD = 2.6, 40 female) were recruited from the Max Planck Institute for Psycholinguistics participant database. All participants reported no language, hearing or psychiatric disorders. No participants took part in the pretest or previous studies. Participants were paid 16 euro.

Apparatus

The experiment was programmed in the software Presentation (version 20.0, Presentation Neurobs). The experiment was run on an HP Z400 workstation running Windows 7 and displayed on an Iiyama Prolite E1980SD screen. Participants gave button press responses on a custom made button box. This button box contained four buttons arranged horizontally. Verbal responses were recorded by a Sennheiser ME64 microphone, with recordings at 16-bit and 44100Hz. Participants were presented with the sounds through Sennheiser HD280-13 headphones.

Materials

For task 1 size judgement/picture naming, the same 10 coloured photographs were used as in Chapter 3. The images are from Belke (2013) and all are from different semantic categories. Images were sized 300 by 300 pixels and presented in the centre of the screen. For the session 1 size judgement task, five of the images (*bus* 'bus', *bal* 'ball', *jas* 'coat', *deur* 'door', *stoel* 'chair') were too big to fit in a shoebox (answering 'no' to the size judgement question) and the remaining five images were small enough to fit in a shoebox (*glas* 'glass', *hand* 'hand', *ui* 'onion', *pet* 'cap', *ring* 'ring'; answering 'yes' to the size judgement question). For the session 2 picture naming task, participants named the images.

For the sound identification task, two vowels and two musical rain (MR) sounds were created in Praat (Boersma, 2002) following the procedure in Uppenkamp et al. (2006). The vowels /a/ and /i/ were synthesised by generating four sine waves with differing frequencies for each sound. For the /a/ vowel, the four sine wave carrier frequencies were 912Hz, 1572Hz, 2852Hz and 2690Hz. For the /i/ vowel, the four sine wave carrier frequencies were 294Hz, 2524Hz, 2911Hz and 4097Hz. For both vowels, the first three sine wave frequencies were taken as the F1, F2 and F3 values listed by Adank, Van Hout, and Smits (2004) for average frequencies for these vowels by native female Dutch speakers. The fourth frequency for each sound was taken by estimating the F4 values of the vowels in the syllables used in the previous experiment ('aak' and 'iek'), which were recorded by a female native Dutch speaker. To mimic naturally occurring spectral tilt (lower power as frequency increases), each individual sine wave was multiplied by $1/((f/500)*3)$, with f denoting the original carrier frequency. Fade in and out was added to the first and last 1ms of each pulse, and the resulting sine waves were then summed to create the vowels.

The two vowels were not only differentiated in vowel quality (/a/ vs. /i/ by varying sine wave frequencies), but also in pitch (222 Hz for /a/; 333 Hz for /i/) and intensity (70.5 dB SPL for /a/; 50.5 dB SPL for /i/). Each MR sound was created from one of the two vowel sounds by jittering the frequencies at each pulse. To keep the MR sounds discriminable from one another for the identification task, the frequencies were jittered in a constrained way (i.e., minimally 10% and maximally 150% of the original frequency). As each MR sound was generated from the corresponding vowel sound, the MR sounds also contained the pitch and intensity information to aid in identification. We found that MR sounds without pitch and intensity information resulted in sounds which were indiscriminable from one another, and thus the pitch and intensity manipulations aided discriminability between the two MR sounds. Vowels retained the pitch and intensity information to keep the acoustic complexity constant. All sounds were matched in duration (460ms) and sampling frequency (44100Hz). Figure 4.1 shows the synthesised vowels /a/ and /i/ and the musical rain sounds used in this experiment.

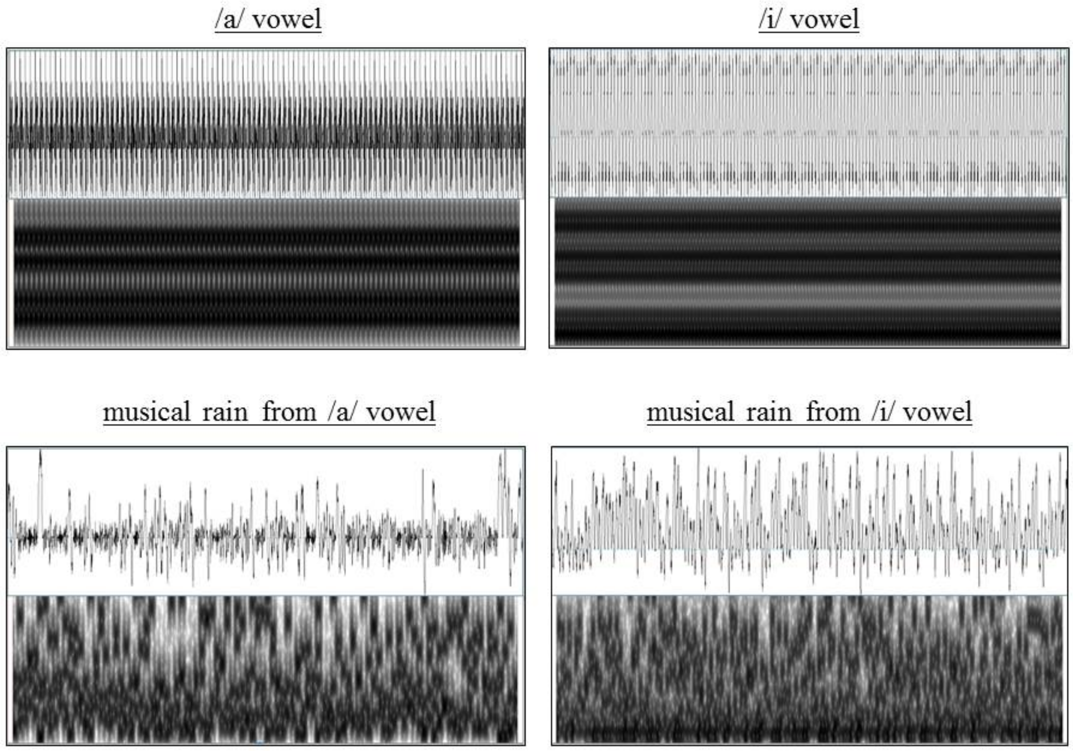

Figure 4.1: Waveforms and spectrograms of the /a/ and /i/ vowels (top figures) and corresponding musical rain sounds (bottom figures).

The vowels and MR sounds were pretested to ensure they were equally difficult to identify, by testing whether there were significant differences in RTs. We found no significant differences between RTs in vowel identification or MR identification (17 participants, mixed models analysis, $|t| < 1$). Because of the intensity manipulation, we additionally analysed whether participants had different RTs to loud and quiet sounds, and whether this varied by stimulus type. Again, no significant differences were found ($|t|$s < 1). Thus, we concluded that vowels and MR sounds were equally easy to identify.

Design

The variables stimulus type (vowel or MR), SOA and experiment session were manipulated within participants. In each session, participants carried out one block of task 1 with vowel identification, and one block of task 1 with MR identification. The order of these blocks was counterbalanced between participant. SOA was variable within block, with SOAs of 50ms and 1800ms. Session 1 (task 1: size judgement + task 2: vowel/MR identification) was always carried out before session 2 (task 1: picture naming + task 2: vowel/MR identification). Session 1 and session 2 were separated by 5-10 days (mean = 7 days).

Participants carried out 100 single identification trials in each session to learn the sound-response mapping. For the first 30 trials, participants received written feedback on screen. 'Correct' (correct) was written if the correct button was pressed for the sound heard, and 'fout' (incorrect) was written if the incorrect button was pressed. No feedback was given for the remaining 70 trials. Trials were presented in a pseudo-randomised order per participant (van Casteren & Davis, 2006) with the constraint that a maximum of five trials containing the same sound could follow one another.

There were 240 experimental dual-tasking trials in total, 120 with vowel identification and 120 with MR identification. Of these 120 trials, 60 trials were at SOA 50ms and 60 trials at SOA 1800ms. Of each group of 60 trials, 30 trials were with the quiet vowel/MR sound and 30 with the loud vowel/MR sound. Each image repeated 3 times in these 30 trials. Trials were presented in a pseudo-randomised order (van Casteren & Davis, 2006) with specific constraints. For session 1, the constraints were: maximum repetition of 3 of the same SOA, the same picture had at least 2 intervening different pictures, the same size judgement could only repeat a maximum of 5 times, and the same auditory stimulus could only repeat a maximum of 5 times. For session 2, the constraints were: maximum repetition of 3 of the same SOA, pictures beginning with the same phoneme had a minimum of 3 intervening trials, and the same auditory stimulus could repeat a maximum of 5 times.

Procedure

In session 1, participants were familiarised with the images and their expected size judgements on paper. Participants were then seated in a sound-shielded testing booth. Participants first saw the 10 images one by one on screen and gave a size judgement via button press. A fixation cross was centred on screen for 500ms. The image was then presented centred on screen and the word 'ja' (yes; items fits into a shoebox) or 'nee' (no; item does not fit into a shoebox) was written underneath the image in white Arial font, size 20. The image stayed on screen until participants gave a response, by pressing the 'ja' or 'nee' button on the button box. Then a blank screen was presented for 1000ms. The first button (i.e. leftmost button in the horizontal row) on the button box was always a 'nee' response and the second button (i.e. middle left) was always a 'ja' response. Participants responded using these two buttons with their left hand, using their middle finger for the first button and their index finger for the second button. Participants then saw all images again without the word 'ja' or 'nee', and they responded to the size judgement by button press. Immediately after the response, the word 'correct' or 'fout' was displayed for 500ms in the middle of the screen, in green or red size 36 capitalised Arial font, respectively. The experimenter controlled presentation of the next image.

Participants then carried out the identification trials. Participants were first familiarised with the vowels/MR sounds five times per sound. Participants saw a fixation cross for 500ms before the vowel/MR sound was played. Participants responded within 1500ms or the trial was terminated. After response or time out, a blank screen was presented for 500ms before the next trial began. The /a/ vowel/'low' MR sound was always responded to by the third button (i.e. middle right) on the button box, and the /i/ vowel/'high' MR sound by the fourth button (i.e. rightmost button). Participants used their right hand to respond to the sounds, using their index finger for the third button and middle finger for the fourth button. For the first 30 trials, participants also saw feedback on their response ('correct' or 'fout') for 500ms in the centre of the screen.

Following this, participants carried out 20 practice dual-task trials. A fixation cross was displayed for 500ms. At SOA 50ms, the image was presented and the vowel/MR sound was presented 50ms later. The image remained on screen for 1000ms. A blank screen was then displayed for 2000ms. At SOA 1800ms, the image was presented for 1000ms, followed by a blank screen for 800ms. Then (i.e. 1800ms after image onset) the sound was presented. A blank screen was then displayed for 2000ms. Participants were informed to first respond to the image and then respond to the sound. The experimental dual-task trials followed the same structure as the practice trials. There were breaks between each part of the experiment.

Session 2 was almost identical to session 1, except that participants were familiarised with the picture names and task 1 was picture naming. Participants were familiarised on paper with the images with their names written alongside them. In the picture familiarisation part of the experiment, participants first saw each image on screen with the name of the image written underneath, and read this aloud. The experimenter controlled the display of the next image. After seeing all pictures once, participants saw the same images without names written underneath. Participants again named the images and the experimenter controlled the presentation of the following image. Each image was thus named twice. Additionally, during the dual-task trials, participants were instructed to name the image aloud before responding to the sound. The sounds were responded to with the index and middle finger of the right hand, using the same buttons as in session 1. After completing session 2, participants were fully debriefed about the experiment. Each session lasted approximately 45 minutes.

Analysis

For session 1, data from 12 participants were excluded due to experiment software crashing (N = 1), experimenter error (N = 1), more than 30% errors in the size judgement task (N = 3), or more than 30% errors in the identification task (N = 7). Data from 38 participants remained for analysis.

For session 2, only 45 participants were tested (5 participants did not return for the second session). Data from 10 further participants were excluded due to experimenter error (N = 1) or more than 30% errors in response order (N = 9). Data from 35 participants remained for analysis. 28 participants contributed data to both sessions.

For the single identification task, errors were defined as either RTs which were incorrect, faster than 200ms or longer than 3000ms, or when no response was measured. Errors accounted for 3.5% of the single identification task data in session 1 and 1.4% in session 2. For the dual-task, errors were defined as either task 1 or task 2 responses which were incorrect, RTs which were faster than 200ms or slower than 3000ms in either task, no responses, and responses given in the wrong order (i.e. task 2 response before task 1). Errors accounted for 11.3% of the dual-task data in session 1 and 12% of the dual-task data in session 2. Error data were not analysed.

We additionally investigated the inter-response interval (IRI), which is the time between the response to task 1 and task 2. Values close to zero suggest response grouping in the data, which makes it difficult to determine which task participants carried out first (Hazeltine et al., 2006). The IRI can also indicate if disengagement from task 1 and engagement with task 2 is more difficult in one condition than another (i.e. with longer IRIs).

The data from both sessions were combined for analysis. Task 1 RTs were measured from task 1 stimulus onset, and task 2 RTs were measured from task 2 stimulus onset. All data were analysed with linear mixed effects models using the lme4 package (Bates et al., 2015) in R (R Core Team, 2017). The maximal random effects structure supported by the data was retained, after increasing iterations and checking for full correlations between random effects. RT data were log-transformed prior to analysis to reduce skew. Control variables included trial number and stimulus type order (whether the vowel or MR block was carried out first). Trial was centred, and stimulus type order was sum-to-zero contrast coded. The experimental variables were stimulus type (vowel or MR), SOA, and session. These were sum-to-zero contrast coded. We took |t| greater than 2 as a marker of significance. Confidence intervals were calculated with the confint.merMod function in lme4 using the profile method.

Results

Single identification task

In session 1, RTs for the MR sounds were 73ms slower than for vowels. In session 2, RTs for the MR sounds were 50ms slower than for the vowels. Figure 4.2 displays the single task data in both sessions.

A linear mixed effects model was run on log-transformed RTs. The fixed effect predictors were order, trial, stimulus type, session, and an interaction between stimulus type and session. Random intercepts by participant and random slopes of stimulus type and session by participant were estimated. Fixed effect model results are shown in Table 4.1.

Table 4.1: Fixed effects from the mixed effects model for the single identification task.

Fixed effects	*Estimate*	*Std. Error*	*t value*	*CI*
Intercept	2.59	0.011	239.1	2.57, 2.61
Order	-0.0006	0.01	-0.06	-0.02, 0.023
Trial	0.001	0.001	1.02	-0.0009, 0.003
Stimulus type	0.028	0.005	5.27	0.018, 0.039
Session	-0.011	0.004	-2.48	-0.02, -0.002
Stimulus type*Session	0.007	0.001	6.55	0.005, 0.009

There were no effects of order or trial. There was a significant effect of stimulus type and a significant effect of session. The interaction was also significant. This indicates that participants responded slower to MR sounds overall, and responded slower in session 2 than in session 1, but the difference in RTs to MR sounds and vowels was smaller in session 2 than in session 1. An analysis with

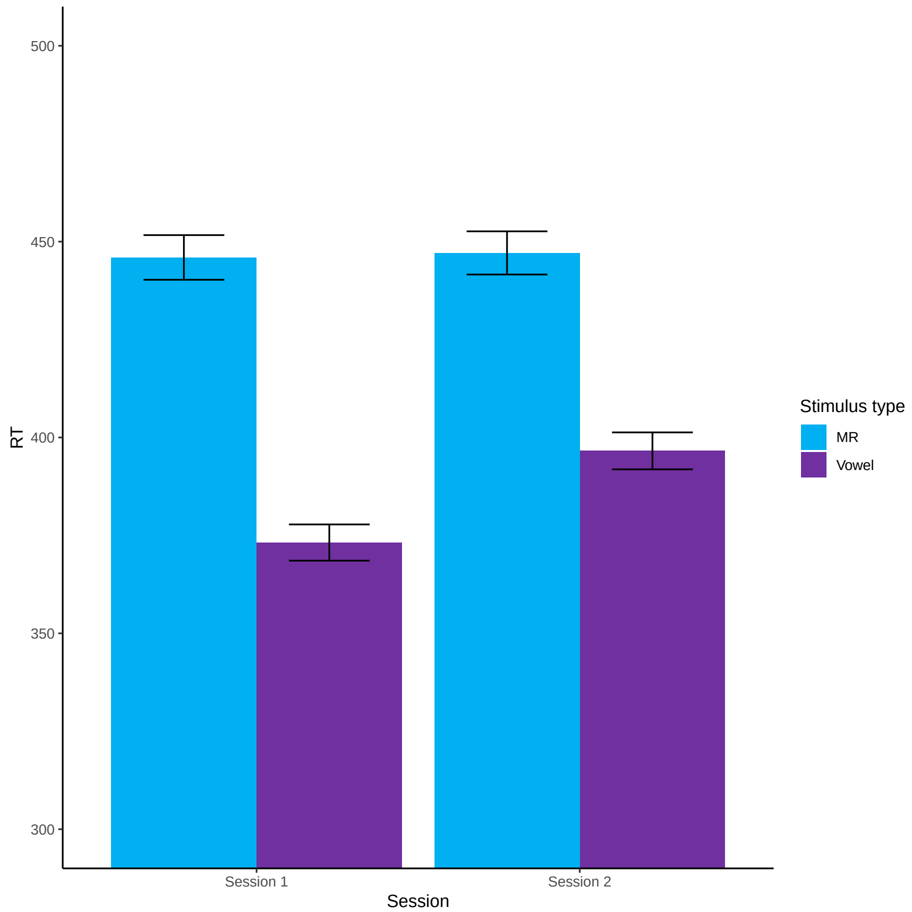

Figure 4.2: Single identification task RTs to vowel and musical rain in ms by session. Error bars are within-participant 95% confidence intervals.

only the participants with data for both sessions showed the same pattern of results (stimulus type, session, and the interaction |t| values were all greater than 2).

Dual-task: Task 1 size judgement/picture naming RTs

In session 1, participants pressed a button to indicate whether the depicted object fit into a shoebox. In session 2, participants named pictures aloud. RTs are visualised in Figure 4.3.

A linear mixed effects model was run on log-transformed task 1 RTs. The fixed effect predictors were order, trial, stimulus type, SOA, session, and interactions between stimulus type, SOA, and session. Random intercepts for participant, picture stimulus, and auditory stimulus were estimated. Random slopes of stimulus type, SOA, session, and a stimulus type by SOA interaction were modelled

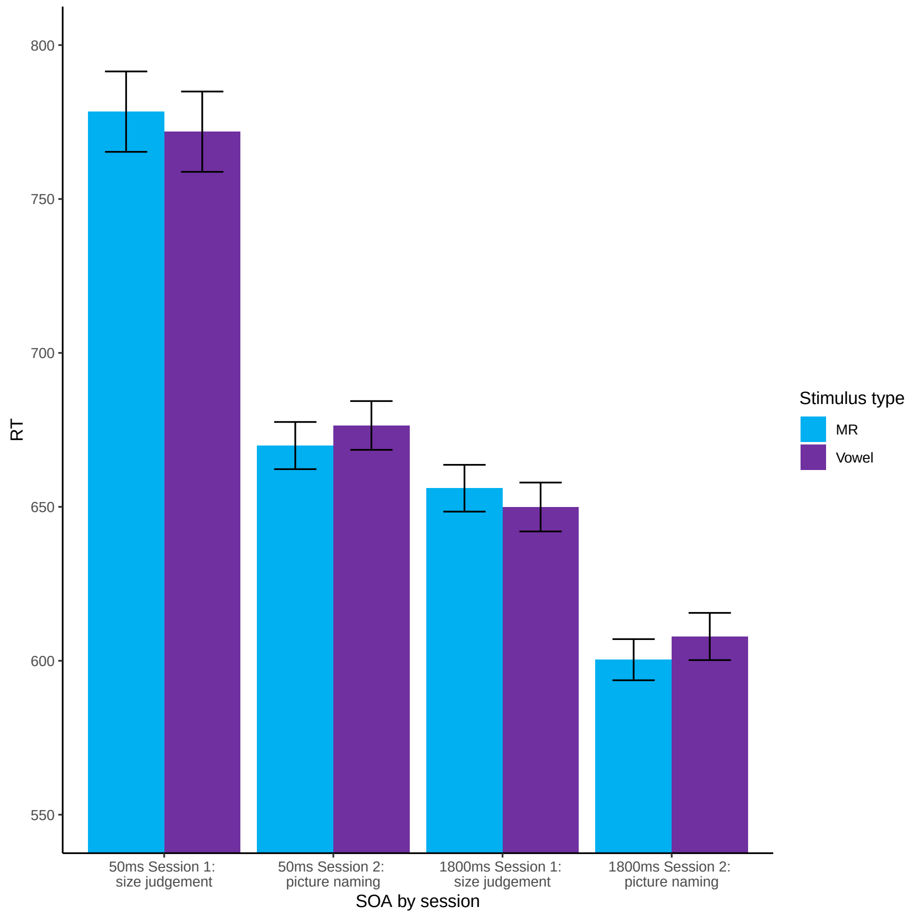

Figure 4.3: Task 1 RTs for size judgement (session 1) and picture naming (session 2) in ms, by SOA as a function of task 2 stimulus type. Error bars are within-participant 95% confidence intervals.

by participant. A random slope of stimulus type was modelled by picture stimulus. Fixed effects from the model output are shown in Table 4.2.

The results suggest no significant effect of stimulus type, indicating that task 1 RTs did not vary by whether the secondary task was vowel or MR identification. Stimulus type did not interact with SOA or session, and the three-way interaction between stimulus type, SOA and session was not significant either. These results demonstrate that there was no effect in task 1 RTs of the type of dual-task (linguistic or non-linguistic).

There was a significant effect of SOA, as task 1 RTs were slower at SOA 50ms compared to SOA 1800ms. There was a significant effect of session, as task 1 RTs were were slower in the size judgement task (session 1) compared to picture naming (session 2). There was also a significant interaction between SOA and session. as the difference in task 1 RTs by SOA was larger in session 1 compared to

Table 4.2: Fixed effects from the mixed effects model for the dual-task task 1 RTs.

Fixed effects	*Estimate*	*Std. Error*	*t value*	*CI*
Intercept	2.82	8.93e-03	315.49	2.8, 2.84
Order	1.64e-02	5.88e-03	2.79	0.003, 0.03
Trial	5e-03	7.51e-04	6.66	0.004, 0.006
Stimulus type	3.08e-04	2.05e-03	0.15	-0.004, 0.005
SOA	2.98e-02	2.91e-03	10.21	0.024, 0.036
Session	2.15e-02	4.07e-03	5.27	0.013, 0.03
Stimulus type*SOA	1.06e-05	1.34e-03	0.008	-0.0026, 0.0027
Stimulus type*Session	8.06e-04	8.17e-04	0.99	-0.0008, 0.0024
SOA*Session	3.81e-03	8.53e-04	4.47	0.002, 0.005
Stimulus type*SOA*Session	-6.33e-04	7.98e-04	-0.79	-0.0022, 0.0009

session 2. Analysis of participants who contributed data to both sessions showed the same pattern of results.

These results show that there was no effect of linguistic dual-tasking additive to non-linguistic dual-task interference when the stimuli were controlled for complexity. Task 1 RTs were not longer in the linguistic dual-task, contrary to the prediction of the linguistic interference hypothesis. Importantly the effect of SOA demonstrates that participants carried out the task as a dual-task in both sessions.

Dual-task: Task 2 vowel/MR identification RTs

In both sessions, participants pressed a button to identify the vowel or MR heard on each trial. RTs are visualised in Figure 4.4.

A linear mixed effects model was run on log-transformed task 2 RTs. The fixed effect predictors included log-transformed task 1 RTs, order, trial, stimulus type, SOA, session, and interactions involving stimulus type, SOA, and session. Random intercepts by participant, picture stimulus, and auditory stimulus were estimated. Random slopes of stimulus type, SOA, session, and a stimulus type by SOA interaction were modelled by participant. Random slopes of stimulus type and session were also modelled by picture stimulus. Fixed effect from the model are displayed in Table 4.3.

We found no significant effect of stimulus type, but stimulus type did interact with SOA and with session. The interaction of stimulus type and SOA was driven by a larger difference between MR and vowel RTs at the 1800ms SOA (with MR

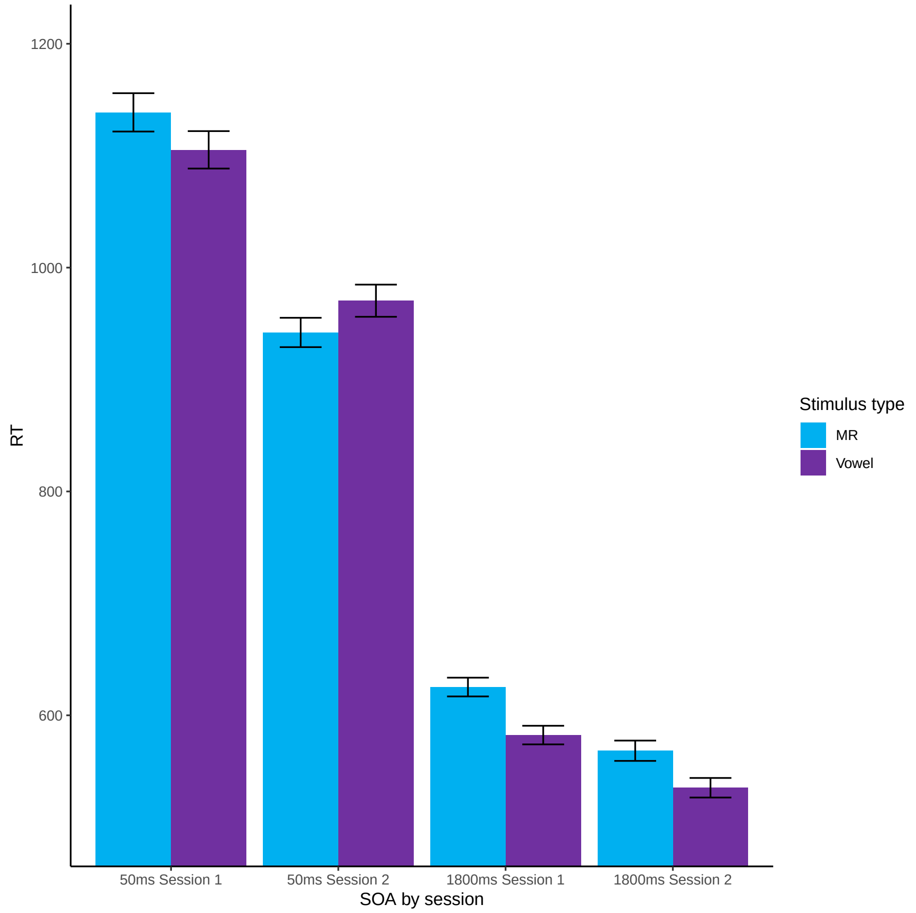

Figure 4.4: Task 2 RTs to vowel and musical rain identification in ms by SOA as a function of session. Error bars are within-participant 95% confidence intervals.

RTs being longer) than at the 50ms SOA. This is expected based on the single identification task results which suggest that MR identification was more difficult than vowel identification, leading to longer RTs. This difficulty may have been absorbed into general dual-task processing at the 50ms SOA, leading to no stimulus type difference at the early SOA. The interaction of stimulus type and session is due to a larger difference between vowel and MR RTs in session 1 (38ms) than in session 2 (2ms). The three-way interaction just reached significance.

There was a significant effect of SOA, as task 2 RTs were longer at 50ms than at 1800m. There was a significant effect of session, as RTs were longer in session 1 than in session 2. There was also a significant interaction between SOA and session, as RTs were longer in session 1 at 50ms than session 2, with a smaller RT difference between SOAs at 1800ms. We expected an SOA effect in task 2 RTs as typical evidence that participants carried out a dual-task. The longer re-

Table 4.3: Fixed effects from the mixed effects model for the dual-task task 2 RTs.

Fixed effects	*Estimate*	*Std. Error*	*t value*	*CI*
Intercept	1.59	0.026	60.03	1.54, 1.64
Task 1 RT	0.45	0.009	51.26	0.44, 0.47
Order	0.006	0.007	0.84	-0.01, 0.02
Trial	0.003	0.0008	3.31	0.001, 0.004
Stimulus type	0.007	0.005	1.31	-0.004, 0.018
SOA	0.12	0.004	31.05	0.11, 0.12
Session	0.022	0.003	6.74	0.015, 0.029
Stimulus type*SOA	-0.008	0.002	-4.82	-0.011, -0.004
Stimulus type*Session	0.003	0.0009	3.04	0.001, 0.005
SOA*Session	0.005	0.0009	5.38	0.0043, 0.007
Stimulus type*SOA*Session	0.002	0.0009	2.01	0.00004, 0.003

sponses in session 1 may be due to the fact that participants gave two manual responses in session 1, compared to a verbal and manual response in session 2. When analysing data from participants who contributed to both sessions, the same pattern of results was found, except that the three-way interaction did not reach significance ($t < 2$).

These results do not provide support for the linguistic interference hypothesis, as we do not find evidence that RTs in vowel identification were longer than in MR identification in session 2.

Dual-task: Inter-response intervals (IRI)

Mean IRIs are reported in Table 4.4. No IRIs are close to zero, indicating no response grouping in this data.

Table 4.4: Mean IRI RTs (in ms) by SOA, session and stimulus type.

SOA	*Session*	*Stimulus type*	*Mean*	*SD*
50	Session 1	MR	410.3	281
		Vowel	383.4	258
	Session 2	MR	322.1	232
		Vowel	344	239
1800	Session 1	MR	1769.2	253
		Vowel	1732.4	260
	Session 2	MR	1768	241
		Vowel	1727.4	247

A linear mixed effects model was fitted to the data with log-transformed IRIs as the dependent variable. Fixed effect predictors included order, trial, stimulus type, SOA, and session, and interactions including stimulus type, SOA, and session. Random intercepts of participant, picture stimulus, and auditory stimulus were modelled. Random slopes of stimulus type, SOA, and session were modelled by participant, and random slopes of stimulus type and session were modelled by picture stimulus. Main effects of SOA are not discussed as they are meaningless in this analysis. Fixed effects from the model are displayed in Table 4.5.

Table 4.5: Fixed effects from the mixed effect model for IRIs.

Fixed effects	*Estimate*	*Std. Error*	*t value*	*CI*
Intercept	2.86	0.013	211.57	2.83, 2.88
Order	0.0005	0.003	0.17	-0.006, 0.007
Trial	-0.004	0.001	-3.1	-0.007, -0.001
Stimulus type	-0.00037	0.009	-0.04	-0.017, 0.016
SOA	-0.39	0.009	-43.95	-0.4, -0.37
Session	0.034	0.006	5.41	0.022, 0.047
Stimulus type*SOA	-0.005	0.001	-3.62	-0.0073, -0.0022
Stimulus type*Session	0.006	0.001	3.92	0.0028, 0.0085
SOA*Session	0.033	0.001	22.76	0.031, 0.036
Stimulus type*SOA*Session	0.006	0.001	4.43	0.0032, 0.0083

There was no effect of stimulus type in IRIs, indicating that broadly the IRIs did not differ depending on whether the auditory stimulus was a vowel or MR sound.

There was an effect of session, with significantly larger IRIs in session 1 than session 2. The interactions between stimulus type and SOA, and stimulus type and session, were significant. The stimulus type by SOA interaction was driven by larger IRIs in the MR condition compared to the vowel condition at the 1800ms SOA (39ms larger) compared to at 50ms (5ms larger). The stimulus type by session interaction was driven by a larger difference in IRIs in session 1 between the vowel compared to the MR condition (32ms difference), compared to in session 2 (9ms). The three way interaction was also significant. Post-hoc pairwise comparisons with multiple comparisons corrected by Tukey's HSD using the lsmeans package (Lenth, 2016) in R specifically investigated whether the longer IRIs with the vowel stimulus type compared to the MR stimulus type at 50ms in session 2 was significant. This specific contrast was investigated because the three-way interaction was significant, and longer IRIs here would suggest a linguistic effect in the data. This comparison was not significant ($t(3.11) = 0.7$), providing no support for a linguistic effect.

A similar pattern of results was found when analysing data only from participants who contributed to both sessions. However, the stimulus type by SOA interaction was not significant with this smaller number of participants ($|t| < 2$), though the pattern was in the same direction as seen with the larger data set (MR IRIs were 40ms longer at the 1800ms SOA, and 10ms longer at the 50ms SOA, than vowel IRIs).

The stimulus type by SOA effect is likely driven by the difficulty of the MR task compared to the vowel task, as the effect was largest at the 1800ms SOA where there is no task overlap. The interaction between stimulus type and session is also likely driven by the higher difficulty of MR responses, especially when two manual responses are given on a trial. Importantly, there are no significant effects which suggest that IRIs are longer when vowel identification is task 2 compared to MR identification as task 2. These results do not provide evidence for the linguistic interference hypothesis.

Bayesian analysis of dual-task RTs

Because null results of stimulus type were found in both task 1 and task 2 responses, we analysed the data with Bayesian paired t-tests. We compared RTs in the vowel and MR conditions in task 1 and task 2, with the data split by SOA and session. Bayesian analyses are useful because they allow us to test whether the lack of a significant stimulus type effect is meaningful and can be interpreted as a true lack of an effect.

Bayesian t-tests were run using the BEST package (Kruschke & Meredith, 2018) in R. For all comparisons the log-transformed RT data were compared to facilitate comparison with the linear mixed effects model analysis. We tested whether the credible intervals indicated by the results of the Bayesian t-tests would cross zero. If the credible intervals cross zero then there is support for a null effect in the data i.e. the conditions are not significantly different. We specified priors for each comparison which are reported below. Note that all tests were run separately with non-informative priors and the same pattern of results was found. All models converged with Rhat values of 1; if models did not converge initially then the number of thinSteps was increased (this increases the number of steps of the MCMC chain, and is typically used when data are autocorrelated).

Task 1 size judgement/picture naming RTs

In session 1 participants carried out a size judgement task. We predicted no differences in size judgement RTs by stimulus type at either SOA under the linguistic interference account. Therefore, we specified the prior to be a difference of 10ms with a SD of 3.3ms at the 50ms and 1800ms SOAs. We took a difference of 10ms to be an effect we would not consider significant. In session 2, participants carried out a picture naming task. Here, at the 50ms SOA, the linguistic interference hypothesis predicted a difference in picture naming RTs by stimulus type, with longer RTs when the stimulus type was a vowel compared to MR. We specified a prior with a difference of 50ms between RTs, and a SD of 16.67ms. We chose this value as in Experiment 1 Chapter 3 at the 50ms SOA there was a 67ms difference between the syllable and tone condition picture naming RTs. At 1800ms SOA, no difference was predicted between responses, so we again specified a prior of a 10ms difference with a SD of 3.3ms. Table 4.6 shows the task 1 RT Bayesian analysis results.

In both session 1 and session 2 there is support for a null effect at both SOAs, because the credible intervals of all tests include zero. These results are not in line with the linguistic interference hypothesis.

Task 2 vowel/identification RTs

In both sessions participants carried out vowel and MR identification. Under the linguistic interference hypothesis, and based on the results of Experiment 1

Table 4.6: Bayesian analysis: task 1 size judgement/picture naming RTs. Mu = difference in means between the RTs in the size judgement and picture naming tasks.

Task 1	*SOA*	*Mu*	*SD*	*Credible Interval*
Size judgement	50	0.0018	0.0084	-0.015, 0.018
Size judgement	1800	0.0057	0.0047	-0.0036, 0.015
Picture naming	50	-0.0025	0.003	-0.0089, 0.0038
Picture naming	1800	-0.0014	0.004	-0.0085, 0.0057

in Chapter 3, longer RTs in vowel identification compared to MR identification were predicted in session 2, with no difference predicted in session 1. For session 1 tests, we specified priors for both SOA comparisons as effects of 10ms with SDs of 3.3ms. In session 2 at the 50ms SOA we specified a prior effect of 100ms with a SD of 33ms, because in Experiment 1 in Chapter 3 we found that syllable identification RTs were 116ms longer at the 50ms SOA than tone identification RTs. At the 1800ms SOA we specified a prior of a 10ms difference with a SD of 3.3ms. Table 4.7 shows the task 2 RT Bayesian analysis results.

Table 4.7: Bayesian analysis: task 2 vowel/MR identification RTs. Mu = difference in means between RTs to vowel and MR identification

Session	*SOA*	*Mu*	*SD*	*Credible Interval*
Session 1	50	0.011	0.008	-0.0057, 0.027
Session 1	1800	0.034	0.008	0.018, 0.051
Session 2	50	0.01	0.005	-0.021, 0.00034
Session 2	1800	0.027	0.006	0.014, 0.039

In both sessions results are consistent at the 50ms SOA with a null effect in the data. In both sessions at the 1800ms SOA, there is evidence for the presence of an effect. These effects are driven by longer RTs in MR identification in both sessions. This was not predicted, and as previously stated, is likely due to the fact that MR identification was more difficult. These analyses provide no support for the linguistic interference hypothesis.

General discussion

In this experiment we aimed to test whether linguistic interference would be found when carrying out a dual-task composed of two linguistic tasks, if the acoustic complexity of the second task was kept constant. Participants carried out two different dual-tasks in two sessions. In session 1 they carried out a non-linguistic size judgement task with a linguistic (vowel) or non-linguistic (MR) identification task. In session 2, participants carried out a linguistic picture naming task with the same vowel and MR identification tasks. The linguistic interference hypothesis predicted that naming RTs in session 2 would be longer with the secondary task of vowel identification compared to MR identification at the 50ms SOA. Additionally, the linguistic interference hypothesis predicted that vowel identification RTs would be significantly longer than MR identification RTs in session 2 (where two linguistic tasks were combined).

We found no evidence supporting the linguistic interference hypothesis. This indicates that when controlling for acoustic complexity in this task, there was no additional linguistic interference. Task 1 RTs were equal in picture naming regardless of whether task 2 was linguistic (vowel identification) or non-linguistic (MR identification). Bayesian analyses support that this null effect is a true null effect. Vowel identification RTs were not longer than MR identification RTs in session 2 either; in fact, MR RTs were longer. There were also no linguistic effects in the IRI data, indicating that the interval between task 1 and task 2 responses did not differ by the linguistic relationship between the tasks. Therefore, the main predictions from the linguistic interference hypothesis were not borne out by this data.

Importantly, we found a significant effect of SOA in task 1 RTs. This shows that task processing in task 1 and task 2 was carried out simultaneously at the 50ms SOA, in line with capacity-sharing theories of dual-tasking (Tombu & Jolicœur, 2003). This also means that the lack of linguistic interference effects cannot be driven by sequential processing of both tasks.

We also found that task 2 RTs in the MR identification task were longer than in the vowel identification task. We believe that this is due to the difficulty of the MR identification task. Despite preserving some pitch information and adding a loudness difference between the two MR sounds, mapping the MR sounds to their button press responses appeared to be more difficult for the participants in this experiment compared to the pretest. We acknowledge that the identification tasks did not appear to be equally difficult for our participants and future research should ensure that the auditory stimuli are more similar in difficulty.

Thus, the results from this experiment provide no support for the linguistic interference hypothesis when acoustic complexity is taken into account. This suggests that, in a dual-task composed of two simple linguistic tasks, if acoustic complexity is controlled then both tasks can be coordinated such that there is no additional interference due to the linguistic nature of the tasks.

The results of this experiment call into question the interpretation of the results of Experiment 1 in Chapter 3. These results were interpreted as providing some support for a linguistic interference hypothesis, where two linguistic tasks carried out concurrently caused mutual interference in both tasks additional to general dual-tasking interference. However, the linguistic nature of the auditory stimuli and acoustic complexity were confounded. The results of the current experiment lead to caution in interpreting the results of Experiment 1 in Chapter 3 as a linguistic interference effect, as it may have been the difference in the acoustic complexity of syllables and tones which led to this result. Caution is also advised for other studies interpreting a purely linguistic locus of interference in a linguistic dual-task (cf. Fargier & Laganaro, 2016).

These results are in line with the results from Experiment 2 in Chapter 3, where we also found no additional linguistic interference when acoustic complexity was held constant. However, we stress that in this study, and in Experiment 2 of Chapter 3, we did not manipulate acoustic complexity. Therefore, while the results are suggestive of acoustic complexity playing a role in dual-task interference, this was not specifically tested. Future studies should investigate whether dual-task interference varies by acoustic complexity while the linguistic nature of the auditory stimuli is controlled. This would provide evidence for acoustic complexity having a causal role in linguistic dual-task interference.

This experiment tested picture naming with a very simple auditory linguistic task, namely vowel identification. It may be that with more complex linguistic tasks, such as picture naming combined with lexical decision, linguistic interference would be found. Thus, the lack of a linguistic interference effect in these data may be due to the fact that the vowel identification task is not 'linguistic enough'.

Despite the fact that acoustic complexity was not directly manipulated in this experiment, these results suggest that an acoustically complex secondary stimulus can affect picture naming. To the best of our knowledge, acoustic complexity of the auditory stimulus has not been tested in any dual-task experiments, or experiments involving picture naming. However, research investigating serial memory recall has found that acoustic complexity plays a role. In general, irrelevant sounds disrupt serial memory recall (e.g., Jones et al., 1992). While some evidence suggests that speech disrupts serial memory recall to a greater extent than other sounds (for review, see Jones, 1995) other research has found that the amount of disruption to serial recall is similar if the sounds are acoustically equated, regardless of whether they are speech-related (Tremblay et al., 2000). Therefore, in a different domain of cognition, acoustic complexity has been shown to play a role in disrupting some cognitive processes.

We can speculate about how an acoustic complexity effect may arise. One possibility is that a more acoustically complex sound requires attention during perceptual processing, which draws attention away from picture naming. This suggests that attention being drawn to aid in auditory perceptual processing is taken away from initial stages of linguistic encoding. Evidence has suggested that allocating focal perceptual attention to a visual stimulus may require central attention, such that visual perceptual processing is hampered by a concurrent task (e.g., Brisson & Jolicœur, 2007; Han, 2017, but note that perceptual processing requires central attention only in specific situations in these studies). A recent study testing whether background noise varying in acoustic complexity affects the production effect - where words which are read aloud are remembered better than words which are read silently - found that certain background noise types resulted in the production effect disappearing (Mama, Fostick, & Icht, 2018). With fluctuating background noise, which was either a mix of eight voices speaking different sentences, or band-passed energetic noise which was shaped similarly to the mixture of voices, the production effect disappeared in comparison to steady-state noise (Mama et al., 2018). The authors interpret their finding as an attentional effect, where attention is diverted to processing the noise and thus less attention is applied to the production task (see also Little, Martin, & Thomson, 2010, for evidence of different attentional effects relating to different auditory stimuli). Altogether, we speculate that some auditory perceptual processes may require attention also required by linguistic processing in the current dual task.

An alternative hypothesis is that maintaining a representation in memory which is more acoustically complex may draw resources away from planning processes. This would entail that resources needed throughout the planning process may be affected, not only at the onset. Therefore, both proposals result in attention being allocated to the auditory stimulus, but differ in their timing of when that attention is allocated.

We note however that speech sounds are acoustically complex by our definition. The majority of dialogue involves streams of speech sounds following one another; people do not often listen to single vowels. Thus, during a conversation, a person may plan speech while listening to acoustically complex sounds, which happen to also be speech. Thus, whether interference is due to linguistic-specific interference, acoustic complexity, or a combination of the two, it is still hypothesised to be present when planning and comprehending speech simultaneously.

In conclusion, our results show that when holding acoustic complexity constant in this dual-task, no additional effects of linguistic interference are found. This suggests that the results of Experiment 1 in Chapter 3 may have been driven by the acoustic complexity of the auditory stimuli, and not their linguistic nature. Overall, these results suggest that acoustic complexity may play a role in dual-task interference, but further testing is required to validate this hypothesis.

5 Dual-tasking with simple linguistic tasks: Evidence for serial processing[1]

Abstract

In contrast to the large amount of dual-task research investigating the coordination of a linguistic and a non-linguistic task, little research has investigated how two linguistic tasks are coordinated. However, such research would greatly contribute to our understanding of how interlocutors combine speech planning and listening in conversation. In three dual-task experiments we studied how participants coordinated the processing of an auditory stimulus (S1), which was either a syllable or a tone, with selecting a name for a picture (S2). Two SOAs, of 0ms and 1000ms, were used. To vary the time required for lexical selection and to determine when lexical selection took place, the pictures were presented with categorically related or unrelated distractor words. In Experiment 1 participants responded overtly to both stimuli. In Experiments 2 and 3, S1 was not responded to overtly, but determined how to respond to S2, by naming the picture or reading the distractor aloud. Experiment 1 yielded additive effects of SOA and distractor type on the picture naming latencies. The presence of semantic interference at both SOAs indicated that lexical selection occurred after response selection for S1. With respect to the coordination of S1 and S2 processing, Experiments 2 and 3 yielded inconclusive results. In all experiments, syllables interfered more with picture naming than tones. This is likely because the syllables activated phonological representations also implicated in picture naming. The theoretical and methodological implications of the findings are discussed.

[1]Fairs, A., Bögels, S., & Meyer, A. S. (2018). Dual-tasking with simple linguistic tasks: Evidence for serial processing. *Acta Psychologica, 191*, 131-148. doi:10.1016/j.actpsy.2018.09.006.

Introduction

A key issue in cognitive psychology is how different cognitive processes are coordinated with one another. This issue has often been investigated in dual-task paradigms, where on each trial participants are asked to respond to two stimuli presented in quick succession. Many dual-task studies have investigated combinations of a linguistic and a non-linguistic task (e.g. Ayora et al., 2011; Cleland et al., 2012; Cook & Meyer, 2008; V. S. Ferreira & Pashler, 2002). There is much less research concerning combinations of two linguistic tasks. Such research is, however, of great importance for psycholinguistics. This is because language is most often used in conversation, where upcoming speakers can begin to plan their utterances while they are still listening to their interlocutor (Barthel et al., 2016; Bögels et al., 2015; Levinson & Torreira, 2015; Sjerps & Meyer, 2015). While such linguistic dual-tasking is often seen as essential for holding a conversation, the underlying skills are still poorly understood. For instance, it is currently unknown how utterance comprehension is affected by concurrent speech planning, or how speech planning is affected by concurrent comprehension. Evidence concerning these important issues can come from dual-task studies with two linguistic tasks. In the present study, we used dual-task paradigms to examine how the processing of a syllable or a tone was combined with picture naming. This research had two goals: (1) to explore the usefulness of dual-task paradigms for research on the coordination of speaking and listening; specifically to determine whether previous key findings of studies using non-linguistic stimuli could be replicated with linguistic stimuli, and (2) to contribute to psycholinguistic theories of conversation; specifically to examine how a key component of speech planning, lexical selection, could be combined with the processing of a spoken syllable.

We used two paradigms, the psychological refractory period (PRP) paradigm (Pashler, 1994), and the task choice (TC) paradigm (Besner & Care, 2003). Both paradigms used the same stimuli, namely one of two tones or syllables (stimulus 1, S1) and a picture with a written distractor word (stimulus 2, S2), but they differed in the tasks. In the PRP experiment (Experiment 1) two overt responses were required: identification of the tone or syllable and naming of the picture. In the TC experiments (Experiments 2 and 3) no overt response was required for S1. Instead, S1 instructed the participant in how to respond to S2, by naming the picture or by reading aloud the distractor. Earlier PRP experiments (Piai et al., 2014; Schnur & Martin, 2012) using non-linguistic S1 found that participants strongly preferred to postpone lexical selection until after response selection of S1. In contrast, earlier TC experiments have shown that the initial processing of non-linguistic S1 can occur in parallel with lexical selection. Our aim was to determine whether we would replicate these patterns with both non-linguistic and linguistic S1. One hypothesis is that syllables and tones should be processed in the same way. An alternative is that syllables, being linguistic stimuli, may automatically activate associated linguistic representations and consequently interfere more with lexical selection than tones, and/or that the processing of a linguistic S1 may be hampered more by concurrent picture naming than processing of a non-linguistic S1. Because of such cross-talk participants may adopt more sequential processing strategies when syllables rather than tones are used as S1. In the remainder of this Introduction we focus on the predictions for the PRP experiment (Experiment 1). The predictions for the TC paradigm are laid out later (Experiments 2 and 3).

Experiment 1 was a near-replication and extension of Experiment 4 conducted by Piai et al. (2014), which we describe in some detail. On each trial of Piai et al.'s study, participants carried out a response to a tone (S1) and named a picture (S2). The stimulus onset asynchrony (SOA) between tone and picture onset was 0ms or 1000ms. Piai et al. (2014) manipulated the difficulty of lexical selection by combining the pictures with written distractor words that were categorically related to the picture names (as in "deer-rabbit") or unrelated. Numerous studies have demonstrated that categorically related distractors slow down picture naming compared to unrelated ones (e.g., Damian & Martin, 1999; Roelofs, 2003; Schriefers et al., 1990). The difference in naming latencies between the related and unrelated distractor conditions is termed the semantic interference effect (Glaser & Düngelhoff, 1984; Schriefers et al., 1990) and is attributed to competition between distractor and picture names arising during lexical selection (Roelofs, 1992, 2003). Briefly, when a picture-word compound is seen, the written distractor and picture activate their associated lexical representations in parallel. Due to mutual activation between categorically related lexical representations, lexical selection for the target name is hampered more by a related compared to an unrelated distractor, as it takes longer to resolve competition.

The main question addressed by Piai et al. (2014) was when lexical selection occurred relative to the selection of the response to the tone. Relevant evidence came from comparing the interference effects at the two SOAs. At the 1000ms SOA, response selection for the tone and lexical selection were most likely carried out in sequence. Consequently, the usual semantic interference effect should be observed. In contrast, when the tone and picture were presented simultaneously, response selection for the tone and lexical selection could be coordinated in different ways. Dual-task theories often distinguish three task stages: pre-selection, response selection, and post-selection (D. E. Meyer & Kieras, 1997; Pashler, 1994, 1998; Tombu & Jolicœur, 2003). Response selection constitutes a processing bottleneck; that is, only one response can be selected at a time (Pashler, 1984, 1994). This bottleneck has been assumed to be structural (Pashler, 1994) or strategic (D. E. Meyer & Kieras, 1997)[2].

To return to Piai et al.'s study, if lexical selection is part of pre-selection processes, it should occur in parallel with pre-selection and response selection processes for the tone. Any competition between target and distractor should be resolved during the "cognitive slack" (Pashler, 1994), i.e. the time that lexical selection waits until the response to the tone has been selected. Therefore at the 0ms SOA, the semantic interference effect should be absent or much reduced compared to the effect seen at the 1000ms SOA. In contrast, if lexical selection is part of response selection or post-selection processes (see Figure 5.1), there is no cognitive slack to absorb the semantic interference effect. Consequently, the effect should be as strong at the 0ms as at the 1000ms SOA. This is because in both cases, lexical selection occurs after response selection for the tone.

Piai et al. (2014)'s results supported the latter hypothesis. Participants were overall slower to name the pictures at the 0ms than at the 1000ms SOA, and slower in the related than in the unrelated distractor condition, and these effects were additive. In other words, the interference effect was not absorbed into cognitive slack at the 0ms SOA. This pattern of results is consistent with the pattern seen in a number of other studies using the same paradigm (Ayora et al., 2011; V. S. Ferreira & Pashler, 2002; Piai et al., 2011, 2014; Schnur & Martin, 2012; but see Dell'Acqua et al., 2007). It supports the view that the semantic interference effect does not arise prior to, but during or after response selection. It also implies that participants strongly preferred to select the responses to the tone and the picture in sequence.

The main question for Experiment 1 of the present study was whether we would observe the same pattern of results as Piai et al. (2014) when we combined picture naming with tone identification, as they had done, and with syllable identification. Thus, in addition to SOA (0ms and 1000ms) and relatedness

[2]Other theories assume no such bottleneck. Response selection of two tasks can be carried out in parallel, but posit a finite amount of capacity which is shared between tasks (Tombu & Jolicœur, 2003). In contrast to response selection, the pre-selection and post-selection processes for two tasks can run in parallel with any other stage.

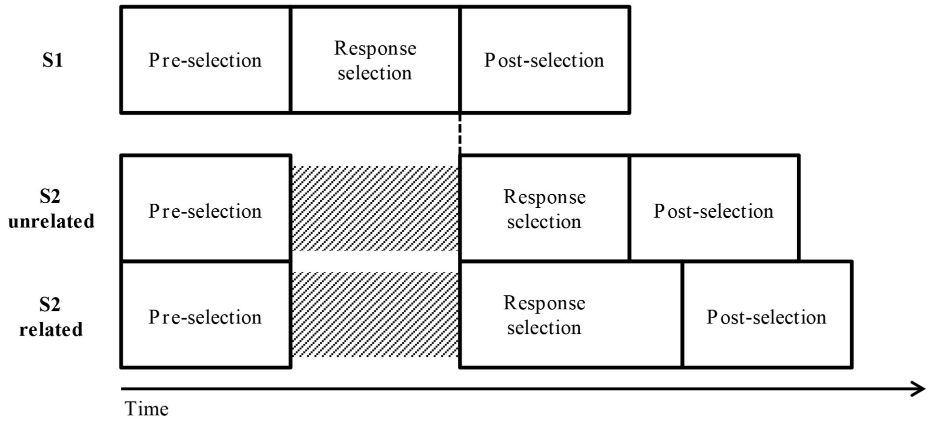

Figure 5.1: Diagram of tone/syllable identification (S1; task 1 - top bar) and picture naming with distractors (S2; task 2 - bottom two bars) at SOA 0ms. The pre-selection stages of the two tasks are carried out simultaneously. The response selection stage of S1 is carried out before the response selection stage of S2. Semantic interference occurs in the response selection stage, with a longer stage for related compared to unrelated stimuli. The greyed area shows the cognitive slack.

between target and distractor, we varied whether S1 was one of two syllables or one of two tones (S1 type). An obvious hypothesis is that the nature of S1 should not affect the pattern of interference. At the 1000ms SOA the response to S1 should be selected well before S2 appears, regardless of the type of S1. At the 0ms SOA the pre-selection processes for S1 and S2 should run in parallel but the responses should be selected sequentially, again regardless of the type of S1.

Alternatively, one might expect to find effects of S1 type. This is because the syllables should activate matching and similar phonological representations (Luce & Pisoni, 1998; Marslen-Wilson, 1987) and possibly word meanings (Gaskell & Marslen-Wilson, 2002). If, as recently has been proposed (Miozzo, Pulvermüller, & Hauk, 2015; Strijkers et al., 2017), pictures rapidly activate phonological information, interference may arise affecting the processing of the syllables, which may slow down syllable compared to tone processing. The same holds if written distractor words rapidly activate associated phonological forms (Brown, Joneleit, Robinson, & Brown, 2002). In other words, when speakers prepare to name pictures, concurrently presented syllables might be harder to identify than tones. In addition, there may be effects of S1 type on picture naming. Syllables should not interfere more than tones with the visual and conceptual processing of the pictures, but additional interference may arise during linguistic encoding. In particular, interference in phonological encoding of the picture names may arise (Schriefers et al., 1990). This would lead to longer picture naming latencies when syllables compared to tones are used as S1.

The prediction that syllables should interfere more with picture naming than tones is consistent with results of a recent PRP study by Fargier and Laganaro (2016), where S1 were tones or syllables and S2 pictures[3]. As task 1 participants pressed a button to respond to one of five tones or syllables (a go/no-go task; only no-go trials were analysed), and as task 2 they named the picture. Fargier and Laganaro (2016) found longer picture naming latencies in the syllable than the tone condition. ERP modulations in the syllable condition around 350ms before speech onset suggested that this effect arose during the phonological encoding of the picture names. The interpretation of these results is complicated by the fact that the syllables were existing words of the participants' native language, that only a single SOA (300ms) was used, and that the manual response latencies were not recorded. Nevertheless, the results point towards specific linguistic interference arising from syllable compared to tone processing on picture naming.

Experiment 1: PRP paradigm

Experiment 1 tested dual-tasking in a PRP paradigm. Participants carried out syllable or tone identification as task 1, with a button press response, and named pictures with distractors as task 2. Two SOAs were tested (0ms and 1000ms). The distractor words were categorically related or unrelated to the picture names. In this experiment we had two aims: 1) to replicate the finding of semantic interference at both SOAs with S1 as tones (e.g., Piai et al., 2014; Schnur & Martin, 2012); and 2) to investigate whether a similar pattern held with S1 syllables. Participants additionally carried out a single S1 identification task block before the dual-tasking trials to practice tone/syllable identification.

Methodology

Participants

36 participants (M = 23 years, SD = 3.5, 31 female) were recruited from the Max Planck Participant Database. All self-reported as right-handed, with no language,

[3]A study by Paucke et al. (2015) tested two production tasks concurrently, where two pictures were displayed side by side. Task 1 was picture naming of the first picture and task 2 was phoneme detection in the planned name of the second picture.

sight or hearing disorders. Participants were paid €12 for participation, and were given sweets to motivate them to stay on task. The experiment was granted ethical approval by the Radboud University Social Sciences ethics committee in accordance with the Declaration of Helsinki.

Materials

Two sine wave pure tones were generated using Audacity (Audacity team 2012) at 300Hz (low tone) and 800Hz (high tone). Two Dutch syllables, [a:k] and [i:k], referred to as 'aak' and 'iek', were recorded by a female native Dutch speaker. The syllables were of the form VVC so that they were maximally discriminable from the onset of the syllable, and so they roughly matched the tones in height discrimination ([a:] is 'low' and [i:] is 'high' in the vowel trapezium). Participants were made aware of the high/low mapping for both tones and syllables. All sounds were 460ms long. The vowel length of 'aak' was 263ms and the vowel length of 'iek' was 264ms. Both 'aak' and 'iek' were on average 222Hz over the entire syllable. Tones and syllables were equalised to 70dB.

The tones and syllables were pre-tested to ensure equal reaction times (RTs) as a proxy for difficulty. In a pre-test with 16 participants there was no significant difference between RTs to the tones and syllables (M_{Tone} = 497ms (SD = 216ms), $M_{Syllable}$ = 511ms (SD = 191ms), t = -0.72). The same stimuli were also used in a previous experiment (Fairs, Bögels, & Meyer, in preparation) where RTs in the single identification task were almost identical between conditions (M_{Tone} = 443.7ms (SD = 157ms), $M_{Syllable}$ = 444.4ms (SD = 178ms), t = 0.37, 33 participants). We concluded that the syllables and tones were matched in difficulty.

The picture materials were taken from Piai et al. (2014). The pictures were 32 white line drawings on a black background, and each picture was shown once with a related distractor and once with an unrelated distractor. Distractors were part of the response set. These 64 picture-distractor pairs were shown once at each SOA with each tone and each syllable. This led to 512 trials, 256 with each S1 (tone or syllable). A further 20 images with distractor words were selected for practice trials. These images were not used in the experiment proper. Of the practice images, only two distractors belonged to the response set (i.e. were also practice picture names). All images were sized 300 x 300 pixels and centred in the middle of the screen. The distractor words were printed in white size 36 Arial font in the centre of the picture.

Design

Participants carried out single tone/syllable identification for 60 trials before moving on to the dual-task trials. Each participant had a unique input file with a pseudo-random order for the single task, generated with the Mix programme (van Casteren & Davis, 2006), with the constraint that the same tone/syllable could repeat at most five times.

For the dual-task trials, task 1 was tone/syllable identification. Button press latencies were measured as the task 1 response. Task 2 was picture naming (with a written distractor). Picture naming latencies from picture onset were measured as the task 2 response. Stimulus 1 type (S1 type; tone or syllable) was a within-participant factor, and blocked, and this block order was counterbalanced between participants. Two SOAs were tested: 0ms and 1000ms. SOA and relatedness were within-participant factors, and were pseudo-randomly presented within each block. Each participant had a unique input file with the following constraints: a) maximum of three repetitions of SOA; b) maximum of three repetitions of relatedness; c) minimum distance of 20 pictures between each picture repetition; d) maximum of five repetitions of the same S1; e) minimum distance of two written distractors between each distractor repetition.

256 experimental trials were shown with each S1 type. An additional six warm-up trials (from the practice set) were added: two at the beginning of each block, and two after each break, which were removed prior to analysis. Participants practised the dual-task on a separate set of 24 practice trials, which (aside from the six warm-up trials) were not displayed during experimental trials.

Apparatus

The experiment was presented on a Benq monitor using the software Presentation (version 16.5, www.neurobs.com). Participants listened to the tones/syllables through Sennheiser HD437 headphones. A custom made quiet button box (created at the MPI, using small microphones rather than buttons) recorded button presses to task 1. A Sennheiser microphone recorded participants' speech to task 2 and the vocal response on each trial was recorded in an individual sound file of 3000ms by the Presentation software.

Procedure

Participants were first familiarised with each practice and experimental picture in the experiment. All pictures were presented in a random order. Each picture was displayed slightly above the centre of the screen with the name of the picture written underneath. Participants were instructed to look at the picture before reading the name out loud and to remember the picture name. Once the participant had read the name aloud, the experimenter displayed the next picture.

Participants were then familiarised with each tone/syllable. After this, participants carried out 60 single task trials. Each trial began with a fixation cross for 700ms followed immediately by the auditory stimulus. Participants pressed the left button for the low tone/aak syllable, and the right button for the high tone/iek syllable. The trial ended when a button was pressed or after 1500ms if there was no response. A blank screen was then presented for 500ms before the onset of the next trial. The response buttons were not counterbalanced across participants to avoid disrupting any inherent stimulus-response mappings (i.e. low is left, high is right).

Participants then practised the dual-task. For the first six trials of the practice block, participants saw only pictures with distractors, and were instructed to name the pictures. After this, each dual-task trial began with a fixation cross for 700ms. In trials with SOA 0ms, the auditory stimulus and visual stimulus (S1 and S2) were displayed at the same time. The S2 stimulus remained on screen for 500ms before being replaced with a blank screen for 1750ms. In trials with SOA 1000ms, the S1 stimulus was presented first and 1000ms after auditory onset the S2 stimulus was displayed for 500ms, with a blank screen displayed afterwards for 1750ms. Experimental trials were presented with exactly the same structure. Participants were instructed to respond to the auditory stimulus with a button press before naming the picture.

Participants took a break between each new task of the experiment, and were given two breaks (one after 90 trials and the other after 180 trials) during the dual-task trials. In the middle of the experiment, participants were encouraged to have a longer break and to leave the testing booth. The experimenter controlled when the participant would start the next block of the experiment. After the experiment, participants were fully debriefed. The entire testing session took approximately 75 minutes.

Results

Pre-processing and analysis

Data were checked for errors. A trial contained an error if the wrong button was pressed for the tone/syllable, if there were any hesitations or disfluencies in the speech recording, if the picture name was incorrect, or if participants named the picture before pressing the button. Any participant who made more than 20% errors in either the single or the dual-task was removed. 4 participants were removed from the dataset for having extremely long RTs in the single task (N = 1), not following task instructions (N = 1), or having more than 20% errors in picture naming trials (N = 2). This left analysable datasets from 32 participants.

The first 30 trials of the single task were removed as practice trials before error checking. Incorrect responses and RTs shorter than 200ms were removed (N = 45, 2.3%). For the dual-task, button press RTs were automatically measured by the experimental software. Speech latencies were manually measured using Praat (Boersma, 2002) for each trial. Before error checking, warm-up trials and any trials with an RT of less than 200ms were removed. 8.2% (N = 1386) of all dual-task trials were removed as errors.

Data were analysed with linear mixed effects models (lme4 package; Bates et al., 2015) using R (R Core Team, 2017). The maximal random structure that would lead to convergence (after adjusting simulation runs and the optimiser) is presented (Barr, Levy, Scheepers, & Tily, 2013). All RTs were log-transformed (base 10) prior to analysis to reduce skew, continuous variables were centred and categorical variables were sum-to-zero contrast coded. Error data were analysed with a binomial logit mixed model, and with the control variable trial centred and scaled. The same models were also run on the raw RT data to confirm that effects were not masked by the log-transformation. Results from these models are presented in the Supplementary Materials (section 7). All raw RT models found the same pattern of results. We took $|t|$ greater than 2 to be our marker of significance. 95% confidence intervals are reported, calculated using the 'profile' option in the confint.merMod function in lme4. All post-hoc tests were carried out with the lsmeans package (Lenth, 2016).

Experiment 1: Single identification task

For the single identification task, mean RTs to the tones were 410ms (SD = 125ms), and mean RTs to the syllables were 450ms (SD = 137ms). The difference between conditions was not predicted as explained in Materials (section 2.1.2).

A linear mixed effect model with log RT as the dependent variable, S1 type as an experimental fixed effect, and fixed control effects of block order (was the tone or syllable block carried out first), trial, and a block order by S1 type interaction was modelled. The random effects structure included a random intercept by participant and random slope of S1 type by participant.

For the fixed control effects, we found an effect of trial (estimate = -0.0006, SE = 0.0002, t = -3.2, CI [-0.001 -0.0003]) as participants sped up through the block. The main effect of block order was not significant ($t < 1$) but the block order by S1 type interaction was significant (estimate = 0.017, SE = 0.007, t = 2.58, CI [0.004 0.03]). This was driven by the fact that participants who carried out the syllable identification block before the tone identification block had a 78ms difference between single task RTs ($M_{syllable}$ = 462ms (SD = 125ms), M_{tone} = 384ms (SD = 108ms)), whereas participants with the reverse order had only a 7ms difference between RTs (M_{tone} = 433ms (SD = 155ms), $M_{syllable}$ = 440ms (SD = 135ms)). We also found a main effect of S1 type (estimate = 0.023, SE = 0.007, t = 3.44, CI [0.01 0.036]), indicating that responses to syllables were slower than to tones, even when controlling for the block order effect.

A linear mixed effects model of the error data (2.3% of the data; 1.7% in syllable identification and 3.1% in tone identification) was conducted with the same model structure as for the RT data. We found no significant effects (all z's < 2, all p's $> .1$).

Experiment 1: Task 1 RTs – tone/syllable identification

Figure 5.2 shows the mean RTs for task 1 by S1 type and SOA in the dual-task. At SOA 0ms, syllables were responded to 54ms slower than tones. At SOA 1000ms, syllables were responded to 38ms slower than tones.

A linear mixed effects model run on the task 1 data included log transformed RT as the dependent variable. Fixed control predictors included trial, block order, and a block order by S1 type interaction. The fixed experimental predictors were SOA, S1 type, relatedness (related or unrelated distractor word), and a S1 type by SOA interaction. The random effects structure included random intercepts by participant and item, and random slopes of SOA, S1 type and relatedness by participant.

We found no effect of block order or significant block order by S1 type interaction (both t's < 2). There was a significant effect of trial such that participants sped up within each block (estimate = -1.52e-04, SE = 1.31e-05, t = -11.62, CI [-0.0002 -0.0001]). There was no main effect of relatedness (estimate = 2.5e-04,

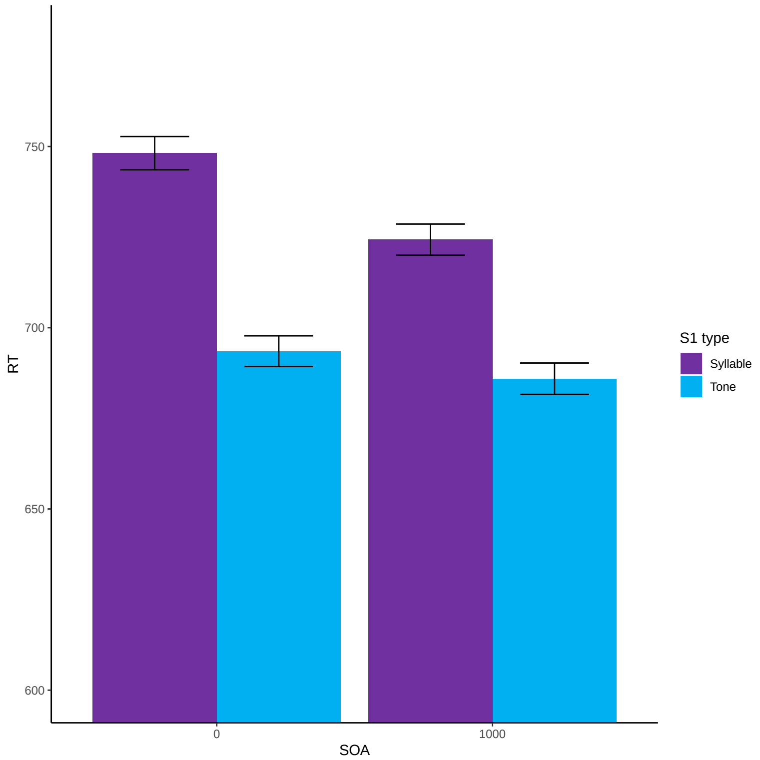

Figure 5.2: Experiment 1: Task 1 RTs (in ms) by SOA and S1 type. Error bars are standard errors and are calculated within-participant. Note that the y-axis does not begin at zero.

SE = 1.03e-03, t = 0.24, CI [-0.002 0.002]). There was a significant effect of S1 type (estimate = 1.49e-02, SE = 4.42e-03, t = 3.36, CI [0.006 0.024]), such that participants responded more slowly to syllables than tones. There was no effect of SOA (estimate = 7.09e-03, SE = 4.81e-03, t = 1.47, CI [-0.002 0.017]), meaning that participants responded equally fast at both 0ms and 1000ms, and no significant S1 type by SOA interaction (estimate = 1.78e-03, SE = 9.87e-04, t = 1.8, CI [-0.0002 0.004]), indicating that the RT difference between syllables and tones was of the same magnitude at both SOAs.

2.6% of the data were task 1 errors. The error counts are presented in Table 5.1. A linear mixed effects model of this error data was conducted with the same model structure as for the RT data except with no S1 type random slope. There was an effect of trial (estimate = 0.08, SE = 0.04, z = 2.22, p = .03), as participants made more errors as the experiment went on. There was an effect of relatedness (estimate = -0.19, SE = 0.05, z = -3.77, $p < .001$), as participants made more errors in the related condition than the unrelated condition. There were also significant interactions between S1 type and block order (estimate = -0.12, SE = 0.06, z = -2.2, p = .03) and S1 type and SOA (estimate = -0.1, SE = 0.04, z = -2.53, p = .01).

The first interaction arose because there was no difference in error proportions in the order syllable-tone (z = 0.8, p = .42), but more errors in syllable identification than tone identification in the order tone-syllable (z = -2.08, p = .04). There was a higher proportion of errors at SOA 0ms in the syllable compared to the tone condition (z = -2.16, p = .03), whereas the proportion of errors by S1 type at SOA 1000ms was similar (z = 0.53, p = .6). These results parallel those found in the RT analysis. There was no evidence for a speed-accuracy trade off as higher numbers of errors were made in the conditions which have the slowest RTs.

Table 5.1: Experiment 1: Error proportions in task 1 RTs by S1 type, SOA and relatedness.

	S1 syllable		S1 tone	
	SOA 0ms	SOA 1000ms	SOA 0ms	SOA 1000ms
Related	0.06	0.05	0.05	0.06
Unrelated	0.04	0.04	0.04	0.05

Experiment 1: Task 2 latencies – picture naming

Figure 5.3 shows the mean latencies for task 2 by S1 type, SOA and relatedness in the dual-task. The descriptive relatedness effect at SOA 0ms with S1 syllables was 36ms and with S1 tones was 29ms. At SOA 1000ms the effect with S1 syllables was 19ms and with S1 tones was 22ms.

A linear mixed effects model included log-transformed task 2 latencies as the dependent variable. Fixed control predictors included trial, block order, log-transformed task 1 RTs and a block order by S1 type interaction. Log-transformed task 1 RTs were included as a control predictor. Because participants were explicitly instructed to respond to task 2 after task 1, and we measure longer RTs to task 1 syllable identification at SOA 0ms than tone identification, we would expect that task 2 naming latencies would also be longer at SOA 0ms with S1 syllables than S1 tones. We thus included task 1 RTs as a control predictor. The fixed experimental predictors were SOA, S1 type, relatedness, and all interactions. The random effects structure included random intercepts by participant and item, random slopes of SOA, S1 type and relatedness by participant, and random slopes of SOA and S1 type by item.

All control predictors were significant: trial (estimate = -6.41e-05, SE = 7.71e-06, t = -8.32, CI [-7.9e-05 -4.9e-05]), log-transformed task 1 RTs (estimate = 0.28, SE = 4.73e-03, t = 58.57, CI [0.27 0.29]), block order (estimate = 1e-02, SE = 4.46e-03, t = 2.25, CI [0.001 0.02]), and block order by S1 type interaction (estimate =

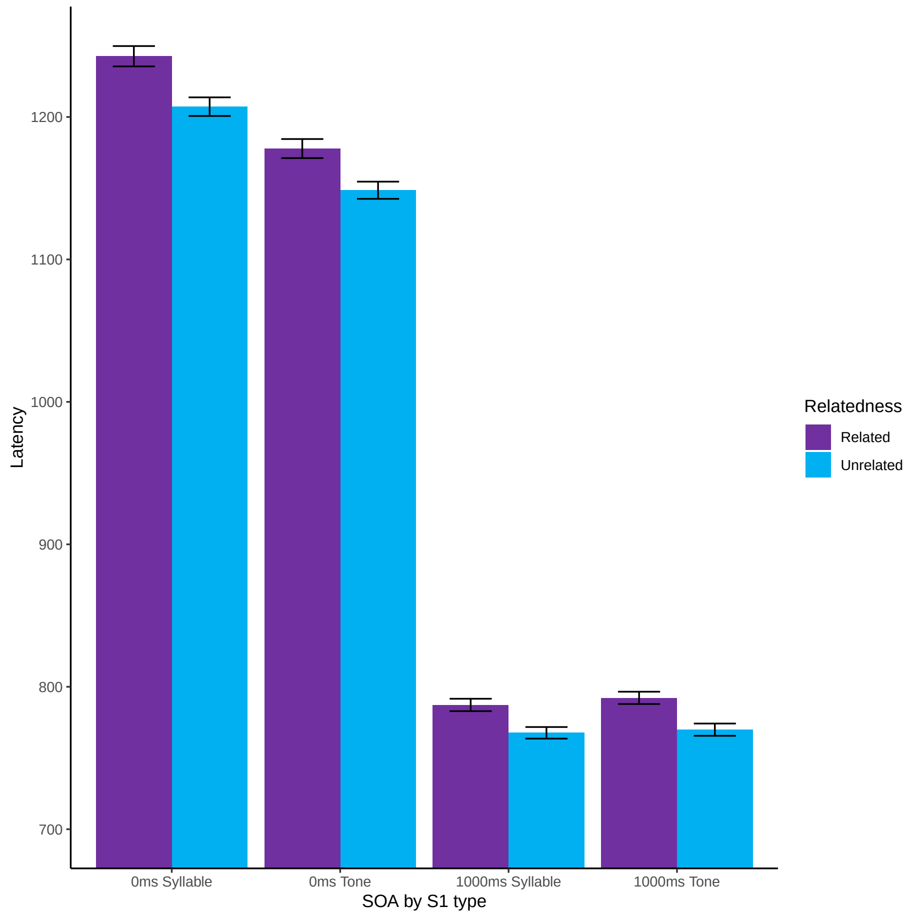

Figure 5.3: Experiment 1: Task 2 mean latencies (in ms) by SOA, S1 type and relatedness. Error bars are standard errors and are calculated within-participant. Note that the y-axis does not begin at zero.

3.32e-03, SE = 1.43e-03, t = 2.33, CI [4.6e-04 6.2e-03]). As with the single task RT analysis, participants who began the experiment with the tone block had similar naming latencies regardless of S1, whereas participants who began with the syllable block were slower to name pictures with S1 as syllables compared to tones. As the order of blocks was a between-participant variable and due to counterbalancing, this difference was likely due to variation in participants. Importantly, including this predictor as a control means that our experimental variables of interest can be interpreted over and above any counterbalancing effects.

The experimental predictor S1 type was not significant ($t < 1$), meaning that naming latencies were similar regardless of S1 type. The predictors SOA (estimate = 8.79e-02, SE = 4.33e-03, t = 20.29, CI [0.079 0.096]) and relatedness (estimate = 5.19e-03, SE = 7.98e-04, t = 6.5, CI [0.004 0.007]) were significant. Participants were significantly slower at SOA 0ms than SOA 1000ms, and participants were significantly slower naming in the related condition compared to the unrelated condition. The interaction between S1 type and SOA was significant (estimate = 5.53e-03, SE = 5.78e-04, t = 9.57, CI [0.004 0.007]). At 0ms latencies with S1 syllables were slower than with S1 tones, with a smaller difference at SOA 1000ms. The SOA by relatedness interaction and relatedness by S1 type interaction did not reach significance, indicating that the relatedness effect was not different between SOAs, and the relatedness effect did not differ by S1 type. The three-way interaction also did not reach significance (all t's < 1).

4.4% of the data were analysable task 2 errors, presented in Table 5.2. This excluded unanalysable errors, such as trials with sneezing, yawning, or coughing. A mixed effects model of the analysable error data was conducted with a similar model structure as for the latency data, but without log-transformed task 1 RTs, and with a random slope of relatedness by item. There was a main effect of trial (estimate = 0.14, SE = 0.04, z = 3.62, $p < .001$) as participants made more errors as the experiment progressed, and a significant interaction between S1 type and block order (estimate = -0.13, SE = 0.06, z = -2.39, $p = .02$). There was also a main effect of relatedness (estimate = -0.24, SE = 0.06, z = -3.88, $p < .001$), as participants made more errors in the related condition than the unrelated condition. The interaction between S1 type and SOA was also significant (estimate = -0.11, SE = 0.04, z = -2.71, $p = .007$). With S1 syllables, there was no difference in errors made at SOA 0ms compared to SOA 1000ms (z = -0.55, $p = .58$). With S1 tones, more errors were made at SOA 1000ms than SOA 0ms (z = 2.27, $p = .02$). Thus, there is some evidence for a speed-accuracy trade-off, but only with S1 tones.

Table 5.2: Experiment 1: Error proportions in task 2 latencies by S1 type, SOA and relatedness.

	S1 syllable		S1 tone	
	SOA 0ms	SOA 1000ms	SOA 0ms	SOA 1000ms
Related	0.06	0.04	0.05	0.06
Unrelated	0.03	0.04	0.03	0.04

Discussion

In this experiment we replicated the key findings of the PRP experiment reported by Piai et al. (2014; see also Schnur & Martin, 2012): We found additive effects of SOA and relatedness on picture naming latencies. In other words, we saw semantic interference effects of equal size at both SOAs. This shows that participants first selected the response to S1 and then selected the name of the picture (see Figure 5.1). Importantly, this held for both types of S1, tones and syllables. Thus, the type of S1 did not affect how the participants coordinated the response selection processes for the two stimuli with each other. This is important as it indicates that a key finding of dual-task experiments using tones as S1 was replicated with syllables as S1.

For task 1 (syllable or tone identification), we observed longer RTs to syllables than to tones. This may be due to the syllables being harder to identify than the tones, or, as suggested in the Introduction, due to interference from concurrent picture processing. While we cannot rule out the latter explanation, the former seems more plausible for two reasons. Firstly, in the single identification task, RTs were longer for syllables than tones. This effect was unexpected as a pre-test and previous use of the stimuli had shown no RT difference between these tones and syllables. Secondly, in the analysis of task 1 RTs, the interaction between S1 type and SOA was not significant. Yet, an interference effect from picture naming onto identification should only be observed if the two tasks were carried out in parallel (at the 0ms SOA), and not if they were performed in sequence (at the 1000ms SOA). Thus, it appears that the participants found the syllables harder to identify than the tones.

For task 2 (picture naming), we found an interaction of S1 type and SOA. Naming latencies were longer in the syllable condition than in the tone condition at the 0ms SOA, but not at the 1000ms SOA. Note that in the analysis of picture naming latencies the effect of identification RT was controlled for. Thus, the interaction was not a direct consequence of the longer identification RTs for syllables discussed above. The interaction indicates that concurrent responding to a syllable interfered more with the naming task than concurrent responding to a tone. Thus, we observed cross-talk between similar tasks, as reported in earlier dual-task studies (e.g., Hommel, 1998; Lien et al., 2007; Logan & Schulkind, 2000; Paucke et al., 2015). We return to the implications of this finding in the General Discussion.

Experiment 2: Task choice paradigm

Experiment 1 showed that participants postponed lexical selection until after response selection for task 1. This is consistent with the view that in dual-task paradigms participants generally prefer to execute response selection processes in sequence rather than in parallel. Theories of dual-task performance commonly assume that the selection stage for a task, where stringent capacity restrictions apply, is preceded by a pre-selection stage, during which multiple cognitive processes can run in parallel (Pashler, 1994; Tombu & Jolicœur, 2003). A number of studies, using the task choice (TC) paradigm described below, have investigated whether lexical selection for a picture name could occur during the pre-selection stage for a non-linguistic stimulus. In Experiments 2 and 3 we used the same paradigm to study whether lexical selection could occur during the early processing of a tone and a syllable.

The TC study that is most relevant to the present research was conducted by Piai, Roelofs, and Schriefers (2015), (see also Janssen, Schirm, Mahon, & Caramazza, 2008; Mädebach, Oppermann, Hantsch, Curda, & Jescheniak, 2011). In this study, the same pictures but tones with different frequencies were used as in Experiment 1 reported above. Tones and pictures were presented with SOAs of 0ms and 1000ms (Experiment 1) or with SOAs of 0ms and 350ms (Experiment 2). No overt response to the tone was required. Instead the tone instructed the participant to name the picture or read aloud the distractor word. As Piai et al. (2011) pointed out, in this task participants could carry out some preparation for both verbal responses, but had to suspend these processes at some point and make the task decision (to read aloud the word or name the picture). They proposed that a good suspension point would be just before the initiation of word form retrieval. This is because word form retrieval for picture naming and for reading aloud requires processing capacity and cannot be easily combined with the task decision (Piai et al., 2011, 2015). Thus, planning processes for the picture name would be suspended after lexical selection. If processing the tone and making the task decision take more time than preparing the picture name up to lexical selection, the semantic interference effect should be absorbed into the cognitive slack created by slower task decision processes. Consequently, the effect should be absent at the 0ms SOA. In contrast, it should be present at the later SOAs, where the task decision precedes lexical selection. This was exactly the pattern Piai et al. (2015) observed: Semantic interference effects were seen at late SOAs but not at the SOA of 0ms. This indicates that lexical selection was combined with the pre-selection processes for the tones .

In Experiment 2 we used the TC paradigm with the same materials as in Experiment 1. The aim was to establish whether we would replicate the pattern seen in the TC experiment by Piai et al. (2015) when S1 were tones, and whether this pattern would also be observed when S1 were syllables. As before, a plausible hypothesis is that syllables and tones should not differ in their effects on picture naming latencies. If this is the case, we should only observe an interaction of SOA and relatedness on the picture naming latencies, with the semantic interference effect being absent at the 0ms SOA and present at the 1000ms SOA. However, Experiment 1 yielded evidence for additional interference of the syllables compared to the tones with the picture naming task, and one might expect this effect to be replicated in the TC paradigm. There should then be a main effect of S1 type, with naming latencies being longer in the syllable than in the tone condition. Finally, because of the additional interference arising in the linguistic condition, participants might choose to schedule the processing of the stimuli differently in the linguistic and non-linguistic condition. While processing may overlap in the tone condition, task choice and lexical selection might occur in sequence in the syllable condition. In the latter case, there would be no slack to absorb the semantic interference effect. Consequently, there should be a three-way interaction, due to additive effects of SOA and relatedness when S1 are syllables, and an interaction of SOA and relatedness when they are tones.

For the word reading task, we did not expect to find a semantic interference effect or an effect of S1 (Glaser & Düngelhoff, 1984; Piai et al., 2015) since in adults word reading is highly automatised and rather immune to distractor effects. Note that in the TC paradigm there are no task 1 RTs to record.

Methodology

Participants

38 participants (M = 22.6 years, SD = 2.4, 31 female) were recruited from the Max Planck Participant Database. All self-reported as right-handed, with no language, sight, or hearing disorders. Participants were paid €12 for participation, and were given sweets to motivate them to stay on task. The experiment was granted ethical approval by the Radboud University Social Sciences ethics committee in accordance with the Declaration of Helsinki.

Materials & Apparatus

The same materials and apparatus were used as in Experiment 1.

Design

The design was the same as Experiment 1, except where indicated. Participants did not carry out any single task identification trials. For the task choice trials, participants were instructed to name the picture if they heard the low tone/'aak' syllable, and read the word if they heard the high tone/'iek' syllable.

One additional constraint was used in generating each participant's input list: there were a minimum of four items between the repetition of any spoken item (either the picture name or the distractor word, depending on what participants should have said on that particular trial).

Procedure

The experimental procedure was identical to Experiment 1 except where indicated. Participants began the experiment with the same picture familiarisation phase as in Experiment 1. Participants then practised the task choice task. Participants first named the picture (with a distractor word present) for four trials, then read the word aloud for two trials, before then being familiarised with the tones/syllables and carrying out 24 practice task choice trials. Written reminders of the tone/syllable mappings were on the table in front of the participants.

Results

Pre-processing and analysis

Data were checked for errors in the same way and with the same error criteria as in Experiment 1. Four participants were removed because they did not follow task instructions (N = 1), or had more than 20% errors (N = 3), resulting in analysable datasets from 34 participants. Speech latencies were pre-processed in the same way as Experiment 1. 6.4% (N = 1119) of the data were removed due to errors. Latencies and errors were analysed in the same way as Experiment 1.

Experiment 2: Naming latencies

Figure 5.4 shows the mean naming latencies by S1 type, SOA, and relatedness. The descriptive relatedness effect at SOA 0ms with S1 syllables was 34ms and with S1 tones was -1ms. At SOA 1000ms the effect with S1 syllables was 27ms and with S1 tones was 5ms.

A linear mixed effects model included log transformed naming latencies as the dependent variable. Fixed control predictors included trial and block order. The fixed experimental predictors were SOA, S1 type, relatedness, and all interactions. The random effects structure included random intercepts by participant and item, and random slopes of SOA, S1 type and relatedness by participant and item.

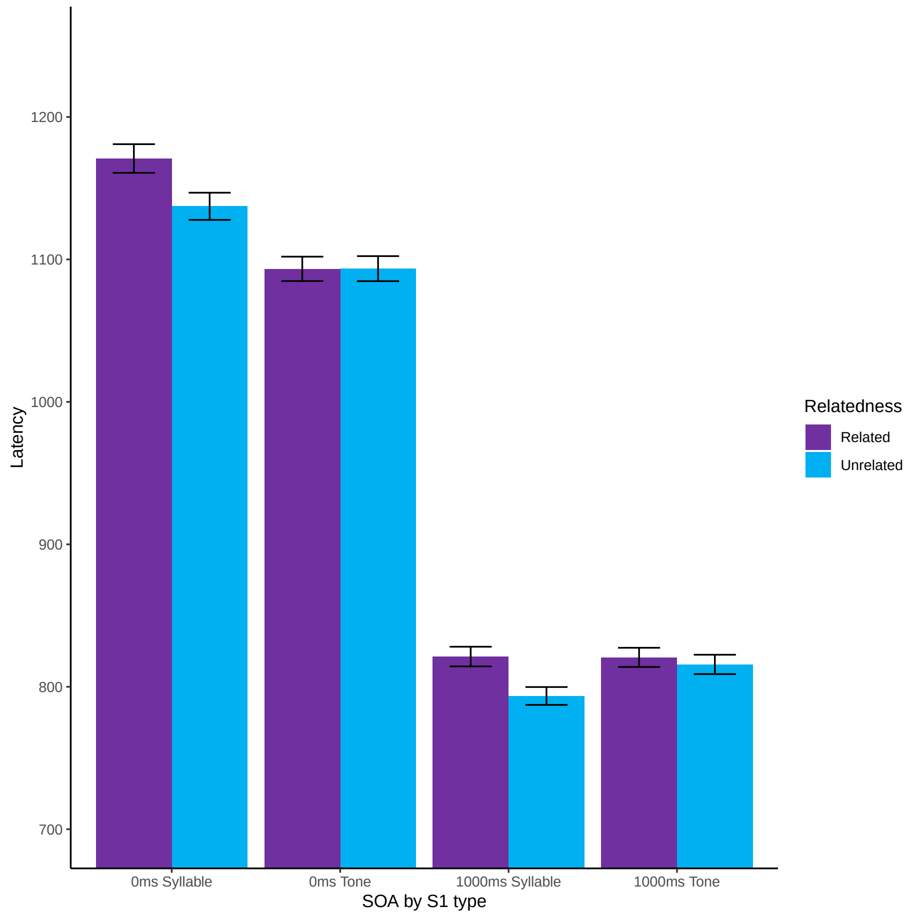

Figure 5.4: Experiment 2: Mean naming latencies (in ms) by S1 type, SOA and relatedness. Error bars are standard errors and are calculated within-participant. Note that the y-axis does not begin at zero.

For the control predictors, the effect of trial just failed to reach significance (estimate = -2.67e-05, SE = 1.36e-05, t = -1.97, CI [-5.3e-05 -8.8e-07]). We found no main effect of block order ($t < 2$).

For the experimental predictors, we found a main effect of SOA (estimate = 6.79e-02, SE = 2.58e-03, t = 26.31, CI [0.063 0.073]) as responses at 0ms were slower than at 1000ms, a main effect of relatedness (estimate = 3.82e-03, SE = 1.32e-03, t = 2.9, CI [0.001 0.006]) as latencies in the related condition were slower than in the unrelated condition, and no main effect of S1 type (estimate = 4.1e-03, SE = 2.47e-03, t = 1.66, CI [-0.0008 0.009]). The interaction between SOA and S1 type was significant (estimate = 6.56e-03, SE = 1.02e-03, t = 6.46, CI [0.005 0.009]), driven by the fact that at 0ms, latencies with S1 syllables were 60ms slower than with S1 tones ($M_{syllable}$ = 1154ms, M_{tone} = 1094ms; $t(49.5) = 4$, $p <$ .001), whereas at 1000ms latencies were more similar (11ms difference; $M_{syllable}$ = 807ms, M_{tone} = 818ms; $t(50.4) = -0.9$, $p = .36$). Relatedness and S1 type also interacted (estimate = 2.85e-03, SE = 1.02e-03, t = 2.8, CI [0.0009 0.005]), demonstrating that the relatedness effect with S1 syllables was significantly larger (t(64.03) = 4, $p < .001$) than with S1 tones (t(62.47) = 0.59, p = .56). SOA and relatedness did not interact ($t < 1$), meaning that the semantic interference effect was the same size at both SOAs. The three-way interaction also did not reach significance ($t < 1$).

8.4% of the naming data were analysable naming task errors, presented in Table 5.3. A mixed effects model of the error data was conducted with a similar model structure as for the naming latency data, but included a random slope of SOA by item. There was a main effect of SOA (estimate = 0.19, SE = 0.05, z = 4.04, $p < .001$), as participants made more errors at SOA 1000ms than at SOA 0ms. There was a main effect of relatedness (estimate = -0.15, SE = 0.06, z = -2.26, p = 0.02), as more errors were made in the related than the unrelated condition. There was main effect of S1 type (estimate = -0.15, SE = 0.05, z = -2.87, p = .004), as more errors were made with S1 syllables than S1 tones. The interaction between SOA and S1 type was significant (estimate = -0.09, SE = 0.04, z = -2.12, p = .03), as there were more errors with S1 syllables at SOA 0ms than with S1 tones (z = -3.28, p = .001), with similar error proportions at SOA 1000ms (z = -1.09, p = .27). The error data is in line with the latency data, and there is no evidence for a speed-accuracy trade-off.

Table 5.3: Experiment 2: Error proportions in naming latencies by S1 type, SOA and relatedness.

	S1 syllable		S1 tone	
	SOA 0ms	SOA 1000ms	SOA 0ms	SOA 1000ms
Related	0.09	0.1	0.06	0.11
Unrelated	0.07	0.09	0.06	0.08

Experiment 2: Reading latencies

Figure 5.5 shows the mean reading latencies by S1 type, SOA, and relatedness. The descriptive relatedness effect at SOA 0ms with S1 syllables was -2ms and with S1 tones was -2ms. At SOA 1000ms the effect with S1 syllables was -3ms and with S1 tones was -8ms.

A linear mixed effects model included log transformed reading latencies as the dependent variable. Fixed control predictors included trial and block order. The fixed experimental predictors were SOA, S1 type, relatedness, and all

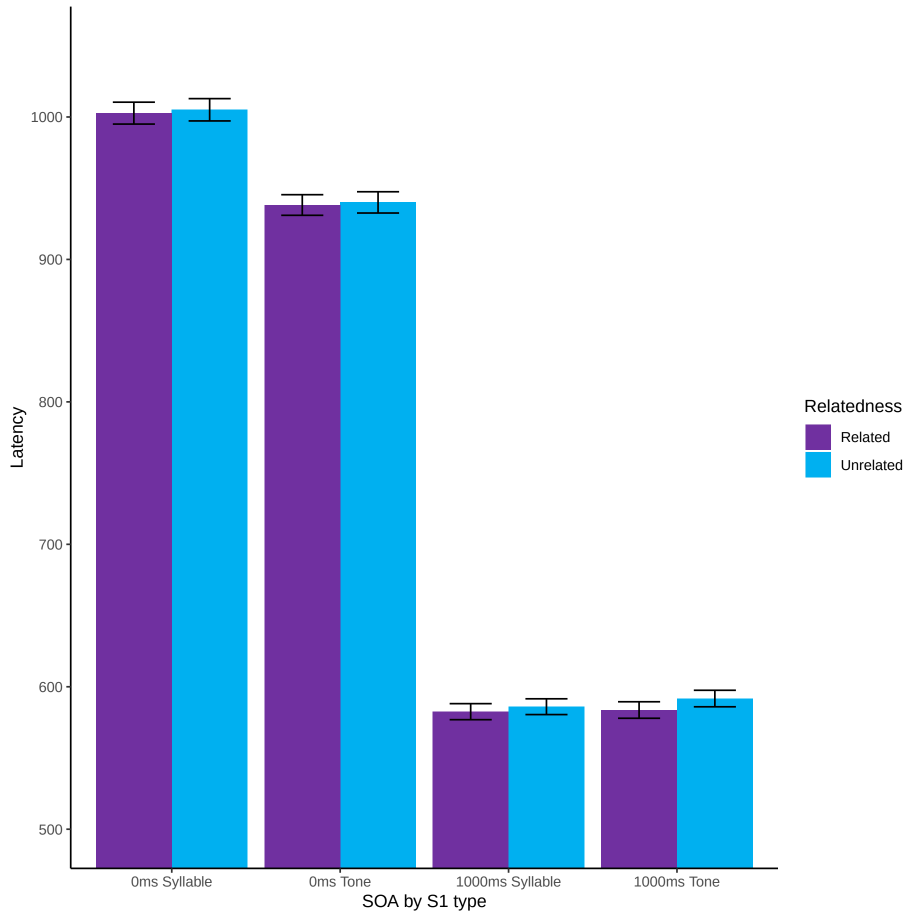

Figure 5.5: Experiment 2: Mean reading latencies (in ms) by S1 type, SOA and relatedness. Error bars are standard errors and are calculated within-participant. Note that the y-axis does not begin at zero.

interactions. The random effects structure included random intercepts by participant and item, random slopes of SOA and S1 type by participant and item, and a random slope of relatedness by participant.

We found no main effects of trial or block order (both t's < 1). There was a significant effect of SOA (estimate = 0.11, SE = 3.06e-03, t = 36.01, CI [0.104 0.116]), as latencies were longer at SOA 0ms than at SOA 1000ms, and a significant effect of S1 type (estimate = 7.12e-03, SE = 2.57e-03, t = 2.77, CI [0.002 0.012]), as latencies were longer with S1 syllables compared to S1 tones. There was no main effect of relatedness (t < 2). The only interaction to reach significance was the SOA by S1 type interaction (estimate = 7.31e-03, SE = 1.03e-04, t = 7.09, CI [0.005 0.093]; all other interactions t < 1). This was because reading latencies at 0ms with S1 syllables were 65ms longer than with S1 tones ($M_{syllable}$ =1004ms, M_{tone} = 939ms; t(44.81) = 5.23, $p < .0001$), whereas at 1000ms the latencies were roughly equal ($M_{syllable}$ = 584ms, M_{tone} = 588ms; t(44.78) = -0.07, p = .95).

2.4% of the reading data were analysable errors, presented in Table 5.4. A mixed effects model of the error data was conducted with a similar model structure as for the reading latency data, but with no random slope of SOA by participant, and no random slopes by item. There were no significant effects.

Table 5.4: Experiment 2: Error proportions in reading latencies by S1 type, SOA and relatedness.

	S1 syllable		S1 tone	
	SOA 0ms	SOA 1000ms	SOA 0ms	SOA 1000ms
Related	0.03	0.02	0.03	0.03
Unrelated	0.02	0.02	0.02	0.02

Discussion

For the picture naming latencies, we obtained a main effect of SOA, indicating that the participants named the pictures faster when the cue to do so preceded rather than coincided with picture onset. There was also a main effect of relatedness, showing that related distractor words interfered more with picture naming than unrelated ones. Importantly, these effects did not interact. Thus, there was no evidence that at the 0ms SOA, the interference effect was absorbed by the cognitive slack created by the response choice task. This may mean that the response choice task was made too fast to create sufficient slack. Alternatively, it may mean that the participants scheduled the tasks differently: Instead of selecting the picture name before making the task choice, as Piai et al. (2015) proposed, participants first selected the task and then proceeded to select the picture name.

There was no triple interaction of SOA, relatedness and S1 type, indicating that the way the response choice task and picture naming were coordinated was not affected by the linguistic or non-linguistic nature of S1. We did, however, find an interaction of S1 type with SOA, showing that at the SOA of 0ms, but not at the SOA of 1000ms, syllables interfered more with picture naming than tones. The same pattern had also been observed in Experiment 1. This interaction was also seen for the reading times, along with a main effect of SOA. Thus, both tasks - picture naming and reading - were hindered more by the presentation of the syllables compared to the tones.

While these results form a coherent pattern, suggesting sequential processing of the two stimuli and little influence of S1 type on the processing strategy, there was one important unexpected finding, namely the interaction of relatedness and S1 type. As Figure 5.4 shows, the semantic interference effect was obtained at both SOAs when S1 was a syllable, but it was not seen at either SOA when S1 was a tone. While the absence of a distractor effect at the 0ms SOA is consistent with the results obtained by Piai et al. (2015), its absence at the 1000ms SOA is surprising and complicates the interpretation of the remaining findings of this experiment. This is because effects at the 0ms SOA are interpreted by comparison with 'baseline' effects at the 1000ms SOA.

Our TC experiment differed from the experiments by Piai et al. (2015) in a number of ways. One potentially important difference is that we varied SOA within blocks (as in our Experiment 1), whereas Piai et al. (2015) used a blocked design. We opted for within-block variation of SOA because our experiments featured an additional variable, S1 type, which was varied between blocks. Both within and between-block manipulations of SOA have been used in earlier TC studies (for within-block manipulations see Besner & Care, 2003; Besner & Risko, 2005; O'Malley & Besner, 2011; Paulitzki, Risko, O'Malley, Stolz, & Besner, 2009; Risko & Besner, 2008; for between-block (but also between-participant) manipulations see Janssen et al., 2008; Mädebach, Oppermann, et al., 2011; Piai et al., 2011). From this literature it is not obvious how this design choice would affect the strength of the semantic interference effect. However, in order to facilitate the comparison of the present results to those obtained by Piai et al. (2015), we conducted an additional experiment varying SOA between rather than within blocks.

Experiment 3: Blocked task choice paradigm

Experiment 3 was the same in design and materials to Experiment 2 except that SOA was blocked rather than variable.

Methodology

Participants

38 participants (M = 21 years, SD = 2.9, 30 female) were recruited from the Max Planck Participant Database. All self-reported as right-handed, with no language, sight, or hearing disorders. Participants were paid €12 for participation, and were given sweets to motivate them to stay on task. The experiment was granted ethical approval by the Radboud University Social Sciences ethics committee in accordance with the Declaration of Helsinki.

Materials, Apparatus & Procedure

The same materials, apparatus, and procedure were used as in Experiment 2.

Design

The design was the same as for Experiment 2, except that SOA was blocked. Block order was counterbalanced such that a participant would either carry out blocks in the order 0ms - 1000ms, or 1000ms - 0ms. The same block order was used for both S1 types (e.g., participant 1 carried out the S1 syllable 0ms block, then the S1 syllable 1000ms block, then the S1 tone 0ms block, then the S1 tone 1000ms block).

Results

Pre-processing and analysis

Data were checked for errors in the same way and with the same error criteria as in Experiments 1 and 2. One participant was removed (more than 20% errors), leaving analysable datasets from 37 participants. Speech latencies were also pre-processed in the same way. 6.5% (N = 1231) data was removed due to errors. Data were analysed and are presented in the same way as Experiment 2.

Experiment 3: Naming latencies

Naming latencies for Experiment 3 are presented in Figure 5.6. At SOA 0ms, there was a descriptive interference effect of 20ms with S1 syllables and of 18ms in with S1 tones. At SOA 1000ms, there was a descriptive interference effect of 24ms with S1 syllables and of 15ms with S1 tones.

A linear mixed effects model included log transformed naming latencies as the dependent variable. Fixed control predictors included trial and block order. The fixed experimental predictors were SOA, S1 type, relatedness, and all interactions. The random effects structure included random intercepts by participant and item, random slopes of S1 type and relatedness by participant and item, and a random slope of SOA by participant.

For the control predictors trial and order, there were no significant main effects ($t < 2$). For the experimental predictors, we found main effects of SOA (estimate = 5.24e-02, SE = 3.1e-03, t = 16.9, CI [0.046 0.058]), as latencies were longer at SOA 0ms than SOA 1000ms, and relatedness (estimate = 4.77e-03, SE = 1.64e-03, t = 2.9, CI [0.002 0.008]), as latencies were longer in the related condition than

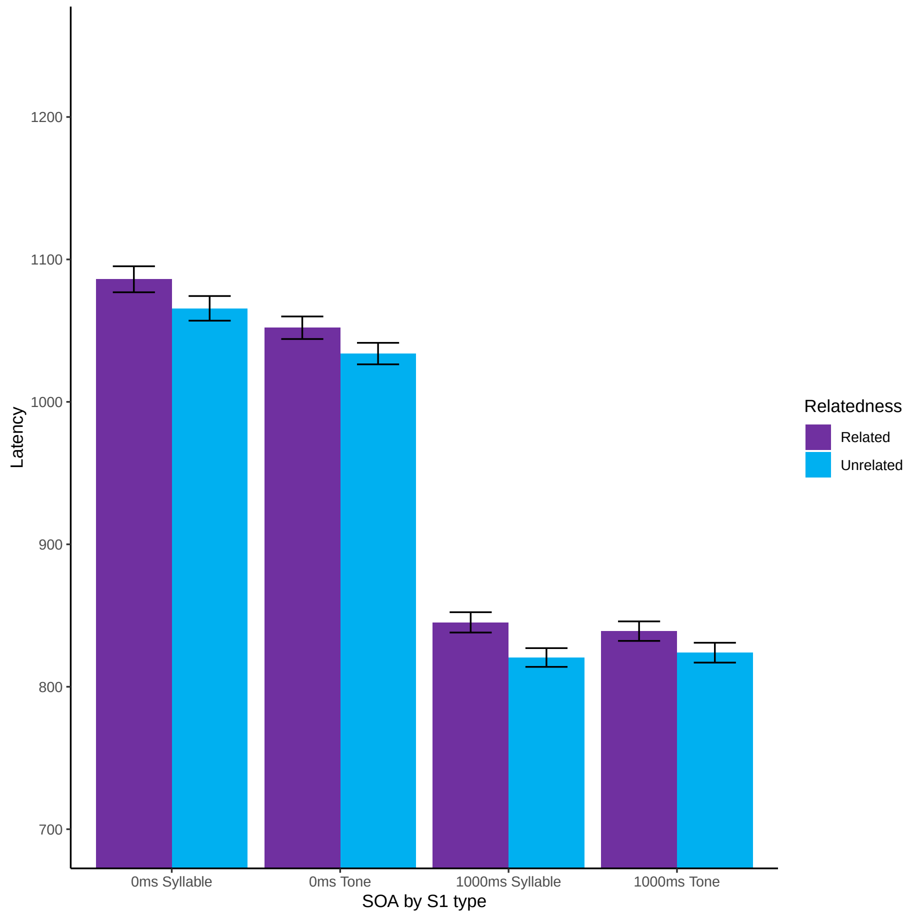

Figure 5.6: Experiment 3: Mean naming latencies (in ms) by S1 type, SOA and relatedness. Error bars are standard error and are calculated within-participant. Note that the y-axis does not begin at zero.

the unrelated condition. There was no effect of S1 type (estimate = 3.39e-03, SE = 2.6e-03, t = 1.3, CI [-0.002 0.009]), meaning that latencies were similar regardless of S1 type. The interaction between S1 type and SOA was significant (estimate = 3.15e-03, SE = 9.7e-04, t = 3.2, CI [0.001 0.005]), driven by the fact that at SOA 0ms, latencies with S1 syllables were 33ms slower than latencies with S1 tones ($M_{syllable}$= 1076ms, M_{tone}= 1043ms; t(48.39) = 2.35, *p* = .02), whereas at SOA 1000ms the difference was smaller at 2ms ($M_{syllable}$= 833ms, M_{tone}= 831ms; t(47.86) = 0.08, *p* = .93). All other interactions were not significant (t $<$ 1). The lack of significant interactions involving relatedness indicates that the relatedness effect did not vary by S1 type or by SOA.

5.6% of the naming data were analysable errors, presented in Table 5.5. A mixed effects model of the error data was conducted with a similar model structure as for the naming latency data, but with no random slopes by item. There was a main effect of SOA (estimate = -0.21, SE = 0.07, z = -3, *p* = .003), as there were more errors at SOA 0ms than SOA 1000ms. There was a main effect of S1 type (estimate = =0.16, SE = 0.08, z = -2.18, p = .03), as there were more errors with S1 syllables. The interaction between SOA and S1 type was significant (estimate = -0.13, SE = 0.05, z = -2.55, p = .01), as there were more errors with S1 syllables than S1 tones at SOA 0ms (z = -3.15, p = .002), whereas the difference at SOA 1000ms was much smaller (z = -0.41, p = .68). There is no evidence for a speed-accuracy trade-off.

Table 5.5: Experiment 3: Error proportions in naming latencies by S1 type, SOA and relatedness.

	S1 syllable		S1 tone	
	SOA 0ms	SOA 1000ms	SOA 0ms	SOA 1000ms
Related	0.1	0.05	0.06	0.05
Unrelated	0.07	0.04	0.04	0.05

Experiment 3: Reading latencies

Reading latencies are presented in Figure 5.7. At SOA 0ms, there was a descriptive interference effect of 5ms with S1 syllables and of 5ms with S1 tones. At SOA 1000ms, there was a descriptive interference effect of 5ms with S1 syllables and of -5ms with S1 tones.

A linear mixed effects model included log transformed reading latencies as the dependent variable. Fixed control predictors included trial and block order. The fixed experimental predictors were SOA, S1 type, relatedness, and all interactions. The random effects structure included random intercepts by participant and item, and random slopes of SOA, S1 type and relatedness by participant and item.

For the control predictors, there was a main effect of trial (estimate = 5.92e-05, SE = 2.63e-05, t = 2.26, CI [7.75e-06 1.11e-04]), as participants slowed down within each block of the experiment. There was no effect of block order ($t < 1$). For the experimental predictors, we found a main effect of SOA (9.34e-02, SE = 3.83e-03, t = 24.38, CI [0.086 0.1]), as participants read aloud faster at the 1000ms SOA than the 0ms SOA. There were no main effects of S1 type or relatedness (t's < 2). None of the interactions reached significance (all t's < 2).

1.9% of the reading data were analysable errors, presented in Table 5.6. A mixed effects model of the error data was conducted with a similar model structure as for the reading latency data, but with no random slopes of SOA or relatedness by

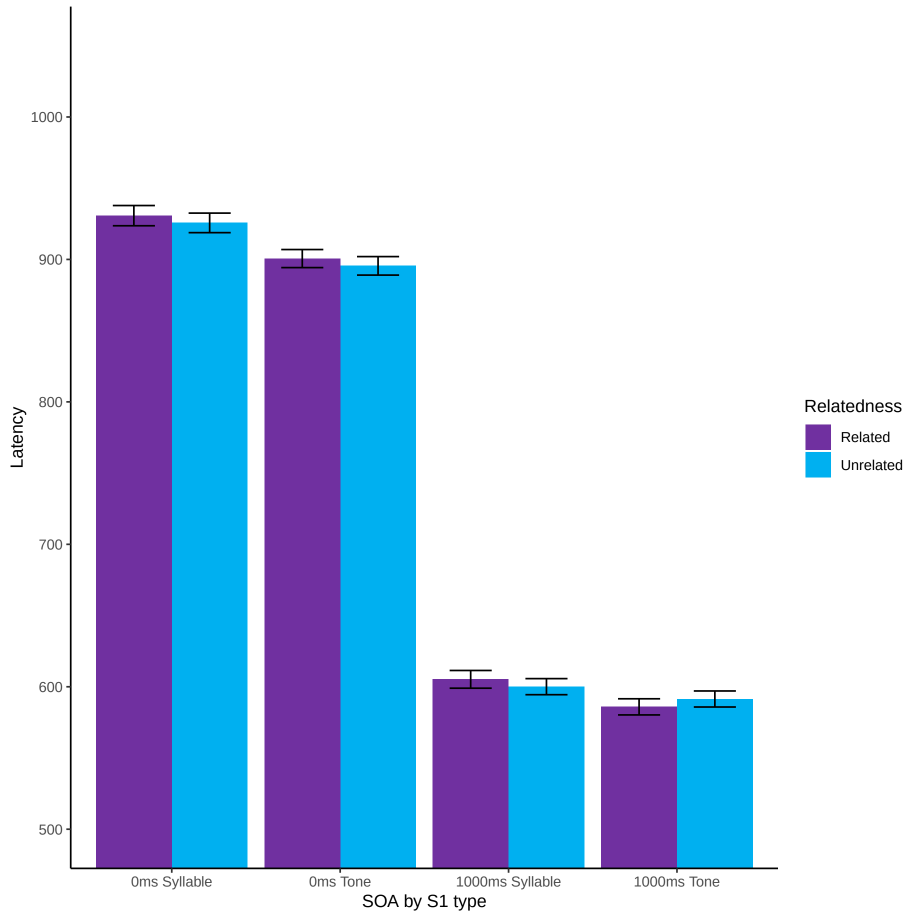

Figure 5.7: Experiment 3: Mean reading latencies (in ms) by S1 type, SOA and relatedness. Error bars are standard error and are calculated within-participant. Note that the y-axis does not begin at zero.

participant or by item. There was a main effect of trial (estimate = 0.24, SE = 0.08, z = 3.19, p = .001) as participants made more errors as the experiment went on. There was a main effect of SOA (estimate = -0.36, SE = 0.08, z = -4.45, $p < .001$), as there were more errors at SOA 0ms than at SOA 1000ms. No other effects were significant. There is no evidence for a speed-accuracy trade-off.

Discussion

The results of Experiment 3 largely correspond to those of Experiment 2. For the picture naming latencies, we obtained the expected effects of SOA and relatedness, with picture naming being faster at the 1000ms than the 0ms SOA, and faster with an unrelated distractor compared to related distractor. These effects did not interact, which, as explained above, may either indicate that at the 0ms SOA the processing of the auditory stimulus did not create enough slack to ab-

Table 5.6: Experiment 3: Error proportions in reading latencies by S1 type, SOA and relatedness.

	S1 syllable		S1 tone	
	SOA 0ms	SOA 1000ms	SOA 0ms	SOA 1000ms
Related	0.03	0.01	0.02	0.01
Unrelated	0.03	0.01	0.02	0.02

sorb the semantic interference effect, or that participants strategically scheduled lexical selection to follow, rather than precede, the choice of task. Importantly, the interaction of relatedness and S1 type, seen in Experiment 2, was not significant. Instead, moderate semantic interference effects were seen in naming for both S1 types and at both SOAs. As before, S1 type interacted with SOA, with syllables interfering more with picture naming than tones, but only at the short SOA.

For the reading latencies, we only observed the expected main effect of SOA, with reading times being shorter at the 0ms SOA than at the 1000ms SOA. Thus, both responses, picture naming and word reading, were initiated faster when the cue preceded the picture compared to when it occurred at the same time, as one would expect. The S1 type effect on reading RTs at the 0ms SOA, seen in Experiment 2, was not replicated.

Bayesian analysis of RTs

Given the inconsistency of the patterns of results obtained in our experiments and the TC experiment reported by Piai et al. (2015), we explored whether the experiments were adequately powered to draw conclusions regarding the presence or absence of the semantic interference effect. If lexical selection occurred in parallel with the processing of S1, most participants should show negligible semantic interference effects. We considered effects (absolute response time differences between the related and unrelated condition) of |10|ms to be negligible. Most published semantic interference experiments obtained effect sizes of at least 20 ms (Ayora et al., 2011; Damian & Martin, 1999; V. S. Ferreira & Pashler, 2002; Glaser & Düngelhoff, 1984; Piai et al., 2011, 2014; Schnur & Martin, 2012; Schriefers et al., 1990), and in the present experiments, the smallest significant interference effect was 15ms. Thus, a 10ms band seemed a good range for negligible effects.

The panels in Figure 5.8 and Figure 5.9 display the semantic interference effect by participant at SOA 0ms in each experiment. The black dotted lines at zero make it easier to see positive and negative effects. The grey band spans 10ms to -10ms to show which effects fall within this negligible band. In the top row of Figure 5.8, showing the results of Experiment 1, most participants have positive semantic interference effects with both S1 types. In the middle row (Experiment 2), with S1 syllables (left middle figure), the interference effects are still largely positive, but there is a wider distribution than in Experiment 1. With S1 tones (middle right figure), the interference effects are almost equally balanced between positive and negative. In the bottom row (Experiment 3), with both S1 types the effect distributions are wide, but most are positive. In Experiments 1 and 2 reported in Piai et al. (2015), displayed in Figure 5.9, almost equal numbers of participants have positive and negative interference effects. Across all experiments, very few individual semantic interference effects are in the negligible band (8-25% of effects by experiment). Thus, from visual inspection of the effect patterns, we conclude that there is no evidence supporting parallel processing.

We carried out a series of Bayesian paired t-tests to test whether Piai et al. (2015) and the present experiments had enough evidence to support the presence or absence of a semantic interference effect. Bayesian two-tailed t-tests carried out in R compared the related and unrelated naming latencies at each SOA with each S1 type using the BEST package (Kruschke & Meredith, 2018), and Bayes Factors were calculated using the BayesFactor package (Morey & Rouder, 2018). For each t-test three chains with approximately 34000 iterations per chain were used for calculation. One chain with 34000 iterations was used for calculation of each Bayes Factor. We set a prior distribution as the log-transformed effect size of 24ms (the average of the 16 semantic interference effects found in the three experiments reported in this paper and the two experiments reported in Piai et al., 2015), with a standard deviation of 8ms. When using the default non-informative prior, the same pattern of results was found. All results are presented in Table 5.7. The column *Difference* reports the difference between conditions on the log (base 10) scale. The column *95% HDI* reports the highest density credible interval. This interval contains the 95% most plausible values of the effect size. The column *% below 0* shows the percentage of the effect sizes which fall below zero. The *% in ROPE* column reports what proportion of the data falls within a region of practical equivalence (ROPE; Kruschke, Aguinis, & Joo, 2012). We set the region of practical equivalence to be approximately 10ms, as displayed in Figures 5.8 and 5.9. Thus, this column tells us what proportion

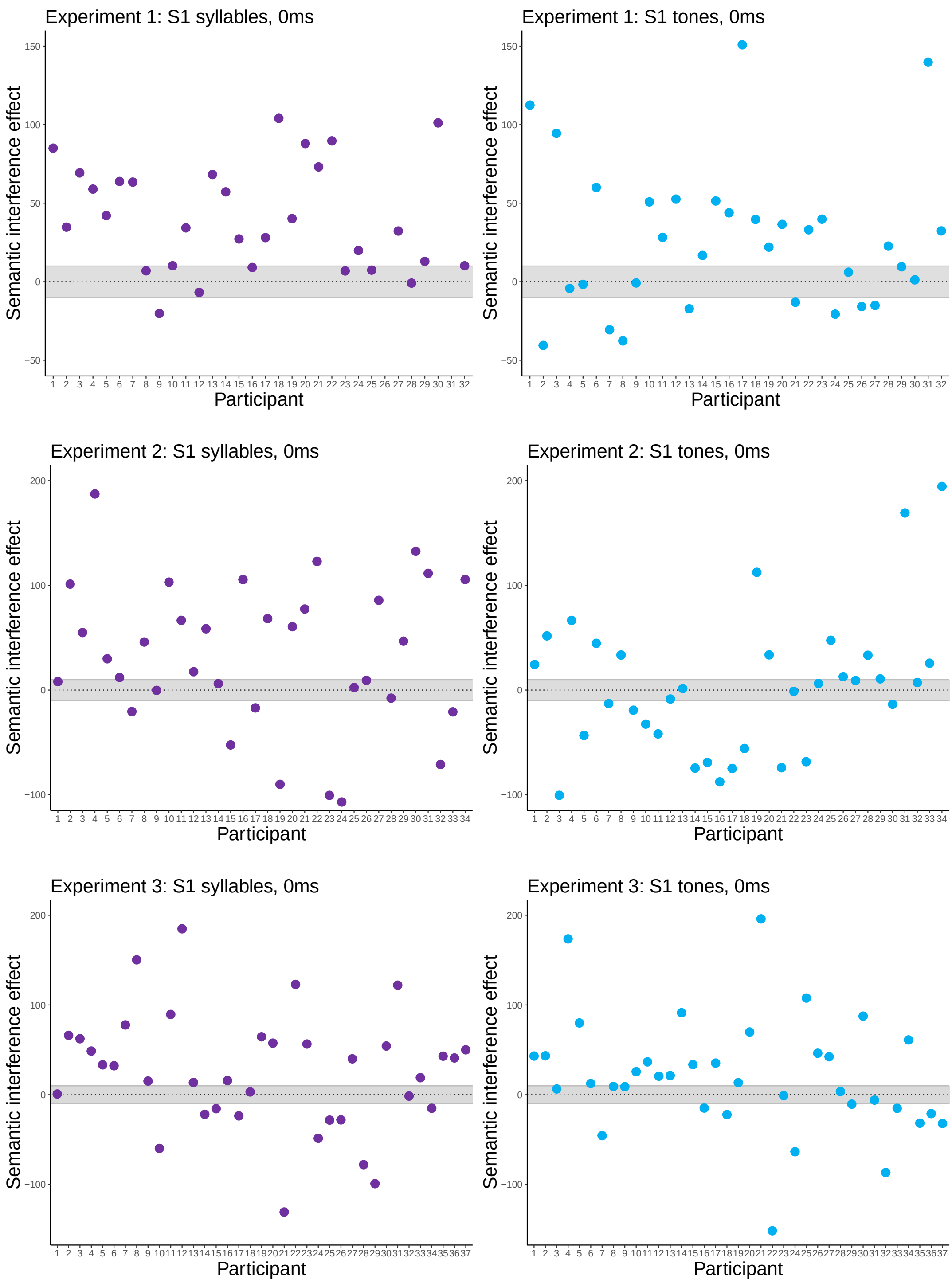

Figure 5.8: Semantic interference effects (in ms) by participant in Experiments 1, 2 and 3 at the 0ms SOA only. The left column of figures (with purple dots) plots S1 syllables in each experiment. The right column (with blue dots) plots S1 tones. Top row = Experiment 1; middle row = Experiment 2; bottom row = Experiment 3. A dotted line is plotted at y = 0ms for ease of interpreting positive and negative effects. The grey band spans y = 10ms to y = -10ms and signifies the band of negligible effects.

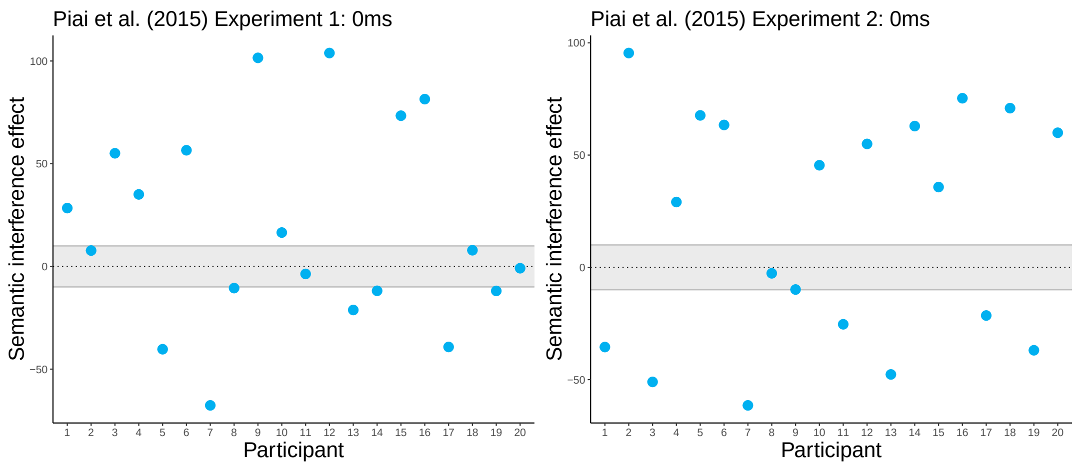

Figure 5.9: Semantic interference effects (in ms) by participant in Experiments 1 (left) and 2 (right) from Piai et al. (2015) at the 0ms SOA only. A dotted line is plotted where y = 0ms for ease of interpreting positive and negative effects. The grey band spans y = 10ms to y = -10ms and signifies the negligible band of effects.

of the difference between related and unrelated conditions is essentially negligible. The final column, *BF*, displays the Bayes Factor for the comparison. Following Kass and Raftery (1995) we take Bayes Factors greater than 3 as indicating support for the alternate hypothesis, as greater than 10 providing strong support for the alternate hypothesis, and less than 1/3 (0.334) as providing support for the null hypothesis.

For Experiment 1 we have fairly strong support at both SOAs with both S1 types for the presence of a semantic interference effect. With S1 tones at 0ms we found slightly weaker support. Almost no data in this experiment fell below zero. Additionally, almost no data fell within the ROPE, meaning that at least 97% of the effects are larger than a negligible effect.

In Experiment 2 we found evidence for the presence of a semantic interference effect with S1 syllables at SOA 1000ms, and inconclusive evidence at SOA 0ms. However, only a small proportion of data fell below zero (less than 1%) and only 2% of the effect fell within the negligible band. This means that the effect is largely meaningful. For S1 tones, we found evidence supporting a *null* effect, meaning that in this condition the absence of the semantic interference effect is meaningful. The Bayes Factors for tests at the two SOAs were both lower than 0.334, supporting the null hypothesis. The HDI for both tests ranged from negative to positive, and almost 50% of the data at 0ms was below zero. Around half of the data from both SOAs fell within the ROPE, meaning that approximately

Table 5.7: Bayesian analysis of the semantic interference effect in the experiments reported here and Piai et al. (2015). Bayes Factors lower than 0.334 provide support for the null hypothesis. Values greater than 3 provide some evidence for the alternate hypothesis. Values greater than 10 provide strong support for the alternate hypothesis. For values between 0.334 and 3 there is not enough evidence for either hypothesis to be supported. Exp = experiment; Difference = mean log difference between related and unrelated conditions; HDI = highest density interval, the 95% most plausible values of the true effect; % below 0 = the proportion of the data which falls below zero; % in ROPE = the proportion of data that falls into a pre-defined region which we deemed a negligible effect, here this is approximately 10ms (0.003 on the log base 10 scale); BF = Bayes Factor.

Experiment	*Difference*	*95% HDI*	*% below 0*	*% in ROPE*	*BF*
Exp 1: S1 syllable, 0ms	0.013	0.007, 0.018	0	0	284.47
Exp 1: S1 tone, 0ms	0.009	0.003, 0.015	0.2	3	7.67
Exp 1: S1 syllable, 1000ms	0.01	0.004, 0.016	0	1	39.3
Exp 1: S1 tone, 1000ms	0.012	0.006, 0.019	0	0	60.36
Exp 2: S1 syllable, 0ms	0.013	0.003, 0.023	0.6	2	2.2
Exp 2: S1 tone, 0ms	0.0007	-0.009, 0.009	49.9	51	0.1
Exp 2: S1 syllable, 1000ms	0.015	0.006, 0.023	0.1	0	30.73
Exp 2: S1 tone, 1000ms	0.002	-0.007, 0.011	29.4	43	0.11
Exp 3: S1 syllable, 0ms	0.01	0.0007, 0.019	1.9	7	0.75
Exp 3: S1 tone, 0ms	0.009	0.0005, 0.017	1.9	8	0.94
Exp 3: S1 syllable, 1000ms	0.013	0.004, 0.023	0.3	1	4.09
Exp 3: S1 tone, 1000ms	0.008	0.0002, 0.016	2.3	9	0.7
Piai et al. (2015): Exp 1, 0ms	0.007	-0.002, 0.017	6.2	16	0.51
Piai et al. (2015): Exp 1, 1000ms	0.023	0.013, 0.034	0	0	204.56
Piai et al. (2015): Exp 2, 0ms	0.007	-0.003, 0.018	7.4	17	0.37
Piai et al. (2015): Exp 2, 350ms	0.019	0.011, 0.027	0	0	87.02

half of the effects were within the negligible band. Therefore, we find support for the null hypothesis.

In Experiment 3 we do not have enough evidence to support either the null or the alternate hypothesis in either condition at the early SOAs, and at the late SOA with S1 tones. The Bayes Factors were inconclusive, and almost 10% of the effects fell within the negligible band. The HDIs were also very wide, estimating a wide range of the size of the plausible effect. Only with S1 syllables at 1000ms did we find evidence, albeit weak, of the presence of the semantic interference effect.

In Experiments 1 and 2 reported in Piai et al. (2015) we found that at the early SOA there was not enough evidence to support either the null or the alternate hypothesis. In comparison with the results from Experiment 2 reported in this paper, only a small amount of the data fell within the ROPE (16% and 17%), and a smaller proportion fell below zero. These results suggest that the data from Piai et al. (2015) are not strong enough to suggest that the absence of semantic interference supports the null hypothesis[4].

In sum, Experiment 1 of the present study, which used the PRP paradigm, provided solid evidence for the presence of semantic interference at both SOAs and regardless of whether the first stimulus was a tone or a syllable. In line with many earlier studies, this indicates that the participants first selected the response to the auditory stimulus and then selected the name of the picture. The remaining experiments, using the TC paradigm mostly provided solid evidence for the presence of semantic interference at the 1000ms SOA. The only exception was the non-linguistic condition of our Experiment 2, where we found solid evidence for the absence of such an effect. As noted above, it is not clear why the effect was absent in this condition. More importantly, the results obtained for the 0ms SOA are inconclusive. None of the TC studies examined here provided convincing evidence for the presence or for the absence of a semantic interference effect at this SOA (except the non-linguistic condition in Experiment 2, but again it is unclear why this was the case). Thus, on the basis of these data, no general claims concerning the way participants schedule the processing of the first and second stimuli can be made.

[4]We also re-analysed Piai et al. (2015)'s data using linear mixed effects models, rather than repeated measures ANOVAs. In this way we were able to concurrently control for participant and item random effects. The mixed model analysis found the same pattern of results for Experiment 1 as reported in Piai et al. (2015), but did not find the same results for Experiment 2. Specifically, we found a main effect of relatedness ($t = 2.42$), and no relatedness by SOA interaction ($t = 1.42$). This indicates that when accounting for participant and item variation, there is evidence for a semantic interference effect at SOA 0ms in Experiment 2.

General discussion

The aim of this study was to examine how participants coordinated lexical selection for a picture name with the processing of a concurrent tone or syllable. In Experiment 1, the psychological refractory period (PRP) paradigm was used to study whether lexical selection could occur in parallel with the selection of the response to the auditory stimulus. Experiments 2 and 3 used the task choice (TC) paradigm to determine whether lexical selection could occur concurrently with the processing of an auditory stimulus which determined the type of response, either picture naming or distractor reading. Below we first focus on the results of Experiment 1, and then turn to those of Experiments 2 and 3 and the Bayesian analysis.

In Experiment 1, semantic interference effects were observed at both SOAs. We interpret this pattern within the framework of dual task theories (D. E. Meyer & Kieras, 1997; Navon & Miller, 1987; Pashler, 1994, 1998; Tombu & Jolicœur, 2003; see Figure 5.1). Pre-selection processes, i.e. visual and early conceptual processes, occurred in parallel for the auditory stimulus and the picture. However, the selection of the response to the tone or syllable and the selection of the picture name occurred in sequence. The semantic interference effect arose because lexical selection took more time in the related than in the unrelated condition. If lexical selection had occurred in parallel with the processing and the response selection for the auditory stimulus, the interference effect would have been absorbed into cognitive slack created by these concurrent processes. In other words, the presence of the semantic interference effect at both SOAs shows that lexical selection followed the selection of the response to the tone or syllable. Our results replicate those of previous PRP studies using similar stimuli and designs (e.g., Ayora et al., 2011; Piai et al., 2011, 2014; Schnur & Martin, 2012). However, in the earlier studies the first stimulus was a tone. Here, we have shown that the same pattern of results was obtained when the first stimulus was a syllable. This is important because PRP research can potentially contribute to our understanding of the way interlocutors coordinate listening and speech planning in conversation. In order to use the paradigm in this context, we first needed to establish whether basic findings obtained with the PRP paradigm can be replicated when the two tasks are linguistic. Our Experiment 1 shows that this is the case.

Additionally we found, not only in Experiment 1 but in all experiments, that at the 0ms SOA the picture naming latencies were longer when the first stimulus was a syllable than when it was a tone. In the statistical analyses of the picture naming latencies, the RTs to the tones or syllables were controlled for. Therefore, it was not the case that lexical selection was simply initiated and completed later because the preceding processes of response selection occurred later for syllables than for tones. At the 1000ms SOA, picture naming latencies after tones and syllables did not differ from one another. This pattern indicates that concurrently presented syllables interfered more with picture naming than tones. In Experiment 2 this was also true for distractor word reading. Most likely additional interference arose in the linguistic condition because the syllables, but not the tones, activated linguistic representations that were relevant to the naming and reading task. The nature of the relevant representations and the precise origin of this cross-talk effect need to be determined in future work. Given that the auditory stimuli were syllables but not words of the participants' native language, it is most likely that the implicated representations were phonological (rather than lexical) and that the effect arose late, during the generation of the phonological forms of the picture names. Fargier and Laganaro (2016) also conducted a dual-task study involving tone or syllable identification as task 1 and picture naming as task 2, and likewise reported longer picture naming latencies in the syllable than the tone condition. The EEG results obtained in this study pointed towards a late, phonological origin of this effect. However, in Fargier and Laganaro's study the syllables corresponded to words of the participants' language; hence additional interference may have occurred at the lexical level. In the present study, the syllables were not words, but they may nevertheless have activated word meanings through spreading activation (Gaskell & Marslen-Wilson, 1997, 2002). The importance of phonological and lexical interference in dual-tasking with two linguistic tasks can now be explored in future work using the same paradigm. Such research would importantly contribute to attaining the goal of understanding the coordination of speaking and listening in everyday conversations.

The PRP experiment (Experiment 1) yielded robust evidence for the presence of semantic interference effects at both SOAs, in the presence of tones as well as syllables. Though there are some exceptions (Dell'Acqua et al., 2007), this pattern of results is in line with the findings reported in earlier studies (e.g., Piai et al., 2014; Schnur & Martin, 2012). The majority of the evidence indicates that response selection for the auditory stimulus precedes the selection of a picture name. The stable pattern of results seen for PRP experiments contrasts sharply with the variable pattern seen in TC experiments compared above (Table 5.7, Experiments 2 and 3, Piai et al. (2015)'s Experiments 1 and 2). The most consistent finding across the TC experiments was the semantic interference effect at the 1000ms SOA, though even this effect was not seen in the non-linguistic condition of our Experiment 2. By contrast the effects seen at the 0ms SOA were highly variable and do not constitute convincing evidence for the presence or for the absence of semantic interference. Consequently, no firm conclusions can be drawn about the research question motivating the use of this paradigm, namely whether lexical selection co-occurred with the early processing of the first stimulus.

Other TC experiments assessing the presence of semantic interference have likewise yielded inconsistent findings. Janssen et al. (2008) presented target pictures with distractor words written in blue or red ink. Depending on the ink colour, participants read the distractor aloud or named the picture. SOAs of 0ms and 1000ms were used, and semantic interference effects were found at both SOAs. In contrast, Mädebach, Oppermann, et al. (2011) used the same method and found no semantic interference effect at the 0ms SOA. Several studies (O'Malley & Besner, 2011; Paulitzki et al., 2009) have used the TC paradigm in conjunction with word and non-word reading tasks in order to study whether lexical and/or prelexical processes can co-occur with the processing of the task choice cue, and have likewise generated somewhat inconsistent findings.

It is always difficult to establish why some paradigms appear to yield more stable results than others. With respect to the PRP and TC paradigms, the comparison made in Table 5.7 is instructive because all studies recruited participants from the same academic community, used very similar materials and equipment, and were closely matched in number of observations. The main difference between the paradigms is obviously the task. We propose that the PRP paradigm is likely to yield more stable results (i.e. conclusive evidence from individual studies and consistency across studies) because the participants are explicitly instructed about the order of processing the auditory and visual stimuli and, perhaps more importantly, their speed of processing the auditory task is monitored.

To elaborate, on each trial of our PRP experiment, participants were presented with three stimuli: an auditory stimulus, a picture, and a distractor word. They were explicitly instructed to respond to the auditory stimulus before naming the picture and to ignore the distractor. These instructions held for all trials and the participants were aware that their performance on both tasks, identification of the auditory stimulus and picture naming, was monitored. Consequently, they should have been highly and consistently motivated to first prioritise processing of the auditory stimulus and then turn to the picture, ignoring the distractor word as much as possible.

In the TC paradigm, the same stimuli were presented, but no overt response to the auditory stimulus was required. However, this stimulus was response-relevant, as it determined how to respond to the visual stimulus, by reading the distractor or naming the picture. Thus, the picture name and distractor word were equally as important on each trial. Piai and colleagues (2011) discussed how task decision and speech planning processes could be scheduled relative to each other. As explained above, they proposed that speakers initially prepared responses to the picture and to the distractor word in parallel, but then suspended these processes to make the task choice. Based on earlier evidence (Reynolds & Besner, 2006; Roelofs, 2008a) they argued that a good suspension point would be prior to word form encoding. This implies that lexical selection for picture names occurred prior to suspension and semantic interference was absorbed into the cognitive slack created by the task choice.

In line with Piai and colleagues' proposal, there are two ways to schedule task processes in the TC paradigm. One way is to make the task choice as early as possible and then proceed with the encoding of either the word or the picture. This strategy should minimise the amount of linguistic encoding carried out in parallel for both stimuli. If such a strategy is adopted, lexical selection for the target picture may not have occurred before task suspension, and consequently a semantic interference effect should be measured. In contrast, if the task choice is made slowly (creating cognitive slack), lexical selection should be completed and no semantic interference effect should be measured. In other words, whether or not a semantic interference effect is observed will depend on how strongly and consistently participants strive to make an early task decision in each trial. The same line of reasoning holds for the PRP paradigm: Task 1 only creates sufficient cognitive slack to absorb differences in the speed of concurrent processes pertaining to task 2 when it takes long enough to complete. In the PRP paradigm, participants are clearly instructed on how to prioritise the tasks and performance on both tasks is monitored. In the TC paradigm participants have to develop a processing strategy themselves and they may vary across trials on how early they make the task decision. We suggest that the higher degree of uniformity of the processing strategies enforced in the PRP paradigm may importantly contribute to the higher consistency of the response strategies within and across participants and, ultimately, of the results across studies.

For psycholinguistic research, the TC paradigm is appealing precisely because it does not force participants to respond overtly to the stimuli they hear. This renders the paradigm more similar to dual-tasking in conversation, where interlocutors listen to others and simultaneously prepare their utterances (Barthel et al., 2016; Bögels et al., 2015; Levinson & Torreira, 2015). However, the dependency of the results on the participants' variable response strategies is problematic. In future research one might aim to develop versions of the TC paradigm addressing this issue. For instance, one could explicitly instruct participants to prioritise processing one of the stimuli, or one could force them to attend early to the pictorial stimulus by presenting it for a very brief period of time. In addition, one might use neurobiological indicators of the allocation of attention to the stimuli (for review see Luck, Woodman, & Vogel, 2000) or the onset of response selection (e.g., Lien et al., 2007).

In conclusion, this study had two goals: (1) to explore the usefulness of dual-task paradigms for research on the coordination of speaking and listening, and (2) to examine how lexical selection for picture naming was combined with the processing of syllables and tones. With respect to the first goal, we found that both paradigms used here were equally as useful for syllable as for tone identification as task 1. Thus, both paradigms can be used in future psycholinguistic research, though the task choice paradigm should be adapted to increase the uniformity of the participants' response strategies or/and to trace these strategies. Concerning the second goal, we did not observe that syllable identification encouraged participants to alter their processing strategy dramatically compared to tone identification. We did however find that syllables interfered more with picture naming than tones. This might not seem too surprising, but is remarkable since only two syllables, not corresponding to words, were used and presented over many trials. One might have thought that under these circumstances the linguistic or non-linguistic nature of the auditory stimuli would not matter much. We found, however, that it did matter, and that the syllables consistently interfered more with picture naming. Future dual-task work can further explore the origin of this cross-talk effect and determine how properties of concurrent speech comprehension affect speech planning.

Acknowledgements

We are grateful to Vitoria Piai for sharing her data and materials, and for being available for many discussions, to Phillip Alday for statistical advice, and to two anonymous reviewers for their valuable comments on the manuscript.

Supplementary Materials

Raw RT and latency models

Experiment 1: Task 1 identification RTs

We carried out a linear mixed effects model with raw RT as the dependent variable. Control predictors included block order, trial, and a block order by S1 type interaction. Experimental predictors included S1 type, SOA, relatedness, and a S1 type by SOA interaction. Random intercepts were fit by participant and item, and there were random slopes of S1 type, SOA and relatedness by participant.

Of the control predictors, we found a significant effect of trial (estimate = -0.27, SE = 0.02, t = -11.43), and block order by S1 type interaction (estimate = 16.05, SE = 7.49, t = 2.14). Of the experimental predictors, we found a significant effect of S1 type (estimate = 20.24, SE = 8.22, t = 2.46), as responses to syllables were longer than to tones. We also found a significant interaction between SOA and S1 type (estimate = 4.15, SE = 1.77, t = 2.34), as the difference between syllable and tone RTs at SOA 0ms was larger than at SOA 1000ms. While this effect is different to that from the log-transformed RT model, the same conclusion can be drawn: the syllable condition was more difficult than the tone condition.

Experiment 1: Task 2 naming latencies

We carried out a linear mixed effects model with raw naming latency as the dependent variable. Control predictors included block order, trial, task 1 RTs, and a block order by S1 type interaction. Experimental predictors included S1 type, SOA, relatedness, and all interactions. Random intercepts were fit by participant and item, with random slopes of S1 type, SOA and relatedness by participant, and random slopes of SOA and S1 type by item.

Of the control predictors, trial (estimate = -0.11, SE = 0.02, t = -5.8), block order (estimate = 18.99, SE = 8.55, t = 2.22), and task 1 RTs (estimate = 0.46, SE = 0.007, t = 70.32) were significant. Of the experimental predictors, there was a significant effect of SOA (estimate = 204.1, SE = 12.8, t = 15.94), as participants were slower at SOA 0ms than at SOA 1000ms. There was a significant effect of relatedness (estimate = 11.93, SE = 1.99, t = 5.99) as participants were slower in the related condition compared to the unrelated condition. Additionally, there was a significant interaction between S1 type and SOA (estimate = 14.09, SE = 1.45, t = 9.73). At SOA 0ms, latencies were slower with S1 syllables compared to with S1 tones. At SOA 1000ms, the latencies were more similar. Importantly, this effect holds with the control predictors in the model. All conclusions are the same as when analysing with log-transformed latencies.

Experiment 2: Naming latencies

We carried out a linear mixed effects model with raw naming latency as the dependent variable. Control predictors included block order and trial. Experimental predictors included S1 type, SOA, relatedness, and all interactions. Random intercepts were fit by participant and item, with random slopes of S1 type, SOA and relatedness by participant and by item.

Neither control predictor was significant. There was a main effect of SOA (estimate = 154.96, SE = 8.02, t - 19.32) as participants were slower to respond at SOA 0ms than SOA 1000ms. There was a main effect of relatedness (estimate = 8.75, SE = 3.11, t = 2.81) as participants took longer to name in the related condition than the unrelated condition. There was also an effect of S1 type (estimate = 12.21, SE = 5.75, t = 2.12), as participants were slower to respond with S1 syllables compared to with S1 tones. The interaction between SOA and S1 type was significant (estimate = 17.36, SE = 2.57, t = 6.77), as participants were slower at SOA 0ms to name with S1 syllables compared to S1 tones, whereas naming latencies were more similar at SOA 1000ms. There was also a significant relatedness by S1 type interaction (estimate = 7.18, SE = 2.57, t = 2.8), as the relatedness effect was present with S1 syllables (t(67.29) = 3.94, p = .0002) and absent with S1 tones (t(65.52) = 0.39, p = .7). The conclusions from these results are the same as when analysing log-transformed latencies.

Experiment 2: Reading latencies

We carried out a linear mixed effects model with raw reading latency as the dependent variable. Control predictors included block order and trial. Experimental predictors included S1 type, SOA, relatedness, and all interactions. Random intercepts were fit by participant and item, with random slopes of S1 type, SOA and relatedness by participant, and SOA and S1 type by item.

Neither control predictor was significant. There was a main effect of SOA (estimate = 193.12, SE = 8.31, t = 23.23), as participants were slower to read the word aloud at SOA 0ms than at SOA 1000ms. There was a main effect of S1 type (estimate = 15.47, SE = 4.85, t = 3.19), as responses were slower with S1 syllables compared to S1 tones. The interaction between SOA and S1 type was significant (estimate = 16.76, SE = 2.14, t = 7.84), as reading latencies were longer at SOA 0ms with S1 syllables compared to S1 tones, whereas at SOA 1000ms latencies were much more similar. No other effects were significant. The results of this model are the same as the log-transformed latency model.

Experiment 3: Naming latencies

We carried out a linear mixed effects model with raw naming latency as the dependent variable. Control predictors included block order and trial. Experimental predictors included S1 type, SOA, relatedness, and all interactions. Random intercepts were fit by participant and item, with random slopes of S1 type, SOA and relatedness by participant, and relatedness and S1 type by item.

Neither control predictor was significant. There was a significant effect of SOA (estimate = 116.61, SE = 7.69, t = 15.17) as latencies at SOA 0ms were longer than at SOA 1000ms. There was a significant effect of relatedness (estimate = 11.04, SE = 3.63, t = 3.04), as latencies were longer in the related condition than in the unrelated condition. The interaction between S1 type and SOA was significant (estimate = 9.2, SE = 2.36, t = 3.89), as latencies at SOA 0ms with S1 syllables were longer than with S1 tones, whereas at SOA 1000ms latencies were more similar. No other effects were significant. This pattern of results is the same as when analysing log-transformed latencies.

Experiment 3: Reading latencies

We carried out a linear mixed effects model with raw reading latency as the dependent variable. Control predictors included block order and trial. Experimental predictors included S1 type, SOA, relatedness, and all interactions. Random intercepts were fit by participant and item, with random slopes of S1 type, SOA and relatedness by participant and by item.

The control predictor trial was significant (estimate = 0.12, SE = 0.05, t = 2.37). Of the experimental predictors, there was a main effect of SOA (estimate = 159.76, SE = 8.83, t = 18.1) as participants were slower to respond at SOA 0ms than at SOA 1000ms. There was a significant interaction between S1 type and SOA (estimate = 4.49, SE = 1.94, t = 2.31), as participants took longer to read with S1 syllables compared to with S1 tones at SOA 0ms, whereas reading latencies were more similar between S1 types at SOA 1000ms. Aside from the interaction, which was not significant in the log-transformed model, the same pattern of results were found when analysing log-transformed latencies. This interaction is found in all other models as well. This interaction is most likely driven by a greater proportion of longer latencies with S1 syllables at SOA 0ms than with S1 tones. When these values are log-transformed, the difference between them is minimised compared to when the raw latencies are used, which is why they are not detected in an analysis with log-transformed latencies as the dependent variable. However, the presence or absence of this interaction here does not change our conclusions from this experiment.

6 Evidence of parallel and reactivated processing in dual-tasking: An MEG study of concurrent syllable identification and picture naming

Abstract

A recent MEG study (Marti et al., 2015) of non-linguistic dual-tasking showed that the time courses of processing in each task did not follow processing chains predicted by existing theories of dual-tasking. Here, we used MEG to determine the time course of processing of two simultaneous linguistic tasks. Participants carried out syllable identification (task 1) and picture naming (task 2), and were instructed to carry out task 1 before task 2. Temporally-generalised multivariate pattern analysis was applied to the MEG recordings in order to resolve the time-course of processing of each task. We found that the time course of processing in syllable identification was similar in a single and a dual-task. However, picture naming was strongly affected by dual-tasking, with processing carried out in parallel with syllable identification and subsequent reactivation. The difference between the parallel and reactivated patterns was equivalent to the reaction-time difference between single and dual-task picture naming. These results do not support the main dual-tasking theories, and suggest a complex pattern of processing in dual-tasking with two linguistic tasks.

Introduction

Recent evidence suggests that people can plan speech while they are listening to an interlocutor (Barthel et al., 2016; Bögels et al., 2015; Bögels et al., 2018; Sjerps & Meyer, 2015). However, the processes underlying the coordination of two linguistic tasks are poorly understood. Evidence from dual-tasking research suggests that when people carry out two tasks at the same time, performance on one or both of the tasks suffers (Pashler, 1994), suggesting that when coordinating speech planning and comprehension at least one task may be hindered. In the current experiment, we took two simple linguistic tasks – picture naming and syllable identification – and carried out an MEG study testing dual-tasking with the two tasks. We traced the time course of brain activation for each task, and determined whether these activation patterns were different for either task in a dual-task compared to when each task was carried out singularly.

Theories of dual-tasking postulate that tasks are composed of three main stages: a perceptual stage, in which a stimulus is perceived; a central stage, in which a stimulus is processed; and an execution stage, in which a response is given (Navon & Miller, 2002; Pashler, 1994; Tombu & Jolicœur, 2003). The perceptual and execution stages can run in parallel with any other stage. The central stage requires capacity, and theories differ on whether this stage runs in parallel or serially with any other stage. The response selection bottleneck theory argues for a bottleneck at the central stage, such that central processing occurs for one task at a time. While the central processes of one task run, central stages of the other task must wait (Pashler, 1994). In contrast, the capacity sharing theory argues that central stages of two tasks can run in parallel, with a finite amount of capacity shared between the tasks (Navon & Miller, 2002; Tombu & Jolicœur, 2003). As capacity is shared, the central processes of both tasks take longer.

These theories make different predictions about how task 1 is affected in a dual-task. Under the response selection bottleneck theory (Pashler, 1994), task 1 is not affected by overlap with task 2, because task 1 processing is shielded. In contrast, the capacity sharing theory (Navon & Miller, 2002; Tombu & Jolicœur, 2003) predicts that task 1 may be affected by overlap with task 2, because task 1 processing is slowed when capacity is shared. Thus, effects of task overlap in task 1 responses allow us to determine if the tasks are carried out in parallel.

A recent MEG study investigating dual-tasking found a pattern of results which does not support either the response selection bottleneck or the capacity sharing theory. Marti et al. (2015) re-analysed an MEG study testing dual-tasking of two simple, non-linguistic tasks (Marti, Sigman, & Dehaene, 2012). Participants carried out tone identification as task 1 and letter identification as task 2. One tone of 1000 or 1100Hz was presented (low and high, respectively) on each trial, and participants responded to the tone via button press. The two letters 'Y' and 'Z' were target letters. These target letters were presented in a letter stream of 12 random letters, with each letter presented for 34ms with a 66ms blank screen interval between letters. The target letters followed tone presentation with different stimulus onset asynchronies (SOAs) in the stream of 1, 2, 4 or 9 letters later. Note that due to the letter + blank screen interval summing to 100ms, letters followed tone presentation by 100, 200, 400 or 900ms. The tone was always presented with the third letter in the stream. Two additional trial types were included: one where no target letters were presented, and one where participants ignored the tone.

Marti et al. (2015) used a novel machine learning technique, temporally- generalised multivariate pattern analysis (MVPA; King & Dehaene, 2014), to analyse the data. In this technique, a classifier is trained to discriminate the MEG topography between two conditions at each sample t. Classifiers trained at sample t are then tested on their ability to decode activity at all other samples t', which gives generalisation across time. By generalising across time, one can determine the neural time course across time of a condition of interest. Classifiers trained to discriminate two conditions can also be tested on a different condition, to determine whether there is decodable information across conditions (cross-decoding). The underlying assumption is that cognitive processes result in stable topographies, and a classifier detects this topography. Thus, if a topography is found at time t in the training dataset and found at time $t + x$ in the testing dataset, this suggests that the cognitive processes in the training dataset occur at a later time (x samples later) in the testing dataset. Marti et al. (2015) trained classifiers for task 1 (discriminating between the 900ms SOA trials and trials where no target letter was presented) and task 2 (discriminating between the 900ms SOA trials and trials where the tone was ignored). These classifiers were then tested on the dual-task trials.

Behaviourally, Marti and colleagues found that responses to task 1 were not affected by the SOA between the two tasks, meaning that tone identification RTs were the same regardless of when the letter was presented. However, task 2 letter identification RTs were affected by SOA, such that with smaller SOAs, longer RTs to letter identification were found. Their classification analysis showed that both classifiers were able to classify over periods of around 200ms, suggesting that the processing of both task 1 and task 2 was made up of successive partially overlapping stages. The performance of the task 1 classifiers was fairly similar regardless of task overlap, though at shorter SOAs the width of accurate decoding was smaller with earlier offsets of successful decoding. In contrast, the performance of the task 2 classifiers was strongly affected by SOA. Between approximately 200-350ms, task 2 classifiers could successfully decode the task in the dual-task trials with weak performance, suggesting that the underlying cognitive processes were weak. Between approximately 350 and 500ms, decoding performance was higher, suggesting stronger cognitive processes. However, the onset of accurate decoding was delayed at shorter SOAs compared to longer SOAs, suggesting that task 2 processing was delayed in the dual-task trials. The width of decoding accuracy was also prolonged at shorter SOAs compared to longer SOAs. After approximately 500ms, the onset of accurate decoding was delayed.

These results suggest initial parallel processing of the two tasks, as up to 350ms classifiers could successfully decode the task. This period was followed by shortened task 1 processes and delayed and lengthened task 2 processes. Neither dual-tasking theory predicts this. The response selection bottleneck theory predicts only a delay in task 2 processes and no effects on task 1 processes. The capacity sharing theory predicts parallel but extended task processing. To explain their data pattern, Marti et al. (2015) suggested that the tasks compete for access to consciousness and attentional resources, and tasks do not passively wait for the bottleneck (Pashler, 1994), nor are fully parallel (Navon & Miller, 2002; Tombu & Jolicœur, 2003).

These findings are important for the current study. They suggest that the coordination of two linguistic tasks may not follow predictions from dual-task theories. However, linguistic tasks may utilise similar representations or processes (cf. Fargier & Laganaro, 2016), resulting in cross-talk between tasks (e.g., Hommel, 1998; Lien et al., 2007; Miller, 2006). To avoid cross-talk, two linguistic tasks may be scheduled sequentially. People appear able to flexibly adjust how they coordinate processing in a dual-task (Lehle & Hübner, 2009), and therefore with two linguistic tasks a serial processing strategy may be favoured.

In the current experiment we tested dual-tasking with two linguistic tasks: syllable identification (task 1) and picture naming (task 2). We presented sounds and pictures with SOAs of 0ms and 1000ms. We applied time-generalised MVPA (King & Dehaene, 2014; Marti et al., 2015) and trained two sets of classifiers to discriminate task 1 syllable identification and task 2 picture naming processing (similarly to Marti et al. 2015). These classifiers were then tested at all time points on the dual-task trials. We were interested in whether 1) we would be able to decode activity at all in either task 1 or task 2; 2) whether the task 1 and task 2 classifiers could decode task processing in the dual-task trials; and 3) whether the time course of decoding was different when carrying out task 1 and task 2 alone compared to in a dual-task.

Methods

Participants

30 participants (13 male, mean age = 24.8 years, SD = 4.93) took part in the experiment and were recruited through the SONA participant recruitment system of Radboud University. All participants were right-handed with normal or corrected-to-normal vision, and reported no hearing, language, or psychiatric disorders. Participants were paid 16 euro for participation. The study was approved by the Ethics Board of Radboud University and was executed in line with the Declaration of Helsinki. For participants who also required an anatomical MRI scan (if there was not one on file), this session took 30 minutes and participants were compensated with an additional 4 euro. Anatomical data are not reported in this chapter.

Materials

There were two tasks: task 1 was a syllable identification task and task 2 was a picture naming task. For the syllable identification task, three sounds were presented. Two sounds were the meaningless syllables 'aak' ([a:k]) and 'iek' ([i:k]), and the other sound was speech-shaped noise. The two syllables 'aak' and 'iek' were recorded by a female native Dutch speaker with a standard Dutch accent. The recordings were time-compressed from 460ms to 305ms using Audacity (Audacity(R), 2014). Five native Dutch speakers listened to the time-compressed recordings and judged them to be natural productions of the syllables. The average pitch of the syllables as measured in Praat (Boersma, 2002) was 222Hz. The speech-shaped noise was generated using an in-house routine. The sounds were concatenated and the spectral content of the concatenated sounds extracted using a fast Fourier transform. This spectrum was then used to modulate broadband white-noise, creating a meaningless sound with a spectral envelope intermediate between the two syllables. This sound was also 305ms long and acted as the speech control sound.

For the picture naming task, ten coloured photographs of everyday objects were selected from Belke (2013; bus *bus*, bal *ball*, jas *coat*, ui *onion*, ring *ring*, hand *hand*, glas *glass*, deur *door*, pet *cap*, stoel *chair*). New photographs for three of the objects were found (*coat, onion, ring*) for higher image quality. A control set of 20 pictures was generated using a diffeomorphing algorithm (Stojanoski & Cusack, 2014) to create two distorted versions of each original picture. The pictures were manipulated using a script from Stojanoski and Cusack (2014; max distortion = 60, nsteps = 8), which deforms images so that they become unrecognisable, while retaining the same low-level visual properties. This technique was specifically designed to maintain similar initial visual processing of a scrambled versus unscrambled image, but to disrupt or even prevent semantic activation. We chose two scrambled images per real image to reduce the possibility of participants learning the correspondence between one scrambled and one real image. The images were sized 300 by 300 pixels.

The experiment was displayed with a grey background (programmed RGB values: 170, 170, 170). Text and the fixation cross were presented in white Arial font, size 20 point. Images were presented in the centre of the screen.

Design

The experiment consisted of three separate tasks. The first task was the picture naming task. Participants carried out 80 picture naming trials – 40 trials contained nameable images and 40 contained unnameable images. Participants named the nameable images and remained silent when seeing an unnameable image. Images were presented in a pseudo-randomised order per participant (using the programme Mix, van Casteren & Davis, 2006; Mix was used for all further randomisations) with the constraints: three other images intervened between presentation of the same image; items with the same initial consonant (bus and bal) had two items between their presentation; and a maximum of five of the same image type (nameable or unnameable) could occur in a row. A trial began with presentation of a fixation cross in the centre of the screen for a random interval of between 600 and 800ms, followed immediately by the image for 100ms. A blank screen was then displayed for a random duration between 900 and 1400ms before the following trial began.

The second task was the syllable identification task. Participants received 75 trials where a sound was presented. ‘aak’ was heard on 25 trials, ‘iek’ on 25 trials, and the speech-shaped noise on 25 trials. Participants pressed a left button for the 'aak' syllable, a right button for the 'iek' syllable, and no button for the noise sound. Sounds were presented in a pseudo-randomised order with the constraint that no more than four trials of the same sound were presented successively. Before the task, participants were familiarised with each of the three sounds. A trial began with a fixation cross presented for a random interval of between 600 and 800ms. The sound was then immediately played. The trial ended either with a response button press or after 1500ms in the case of no response. A blank screen was then displayed for a random duration between 400 and 500ms before the following trial began.

The third task was the dual-task. Participants first practised the dual-task. Participants were instructed to first respond to the sound and then respond to the image, to maximise the length of trials uncontaminated by muscular artifacts from speech. There were six dual-task trial conditions, hereafter labelled conditions A-F, with descriptive names for conditions presented in Table 6.1. These conditions were different combinations of the sound and image types. In condition A (no task), participants heard speech-shaped noise and saw an unnameable image. No responses to either task were required in this condition. In condition B (single task 1), participants heard a syllable and saw an unnameable image. Only a response to task 1 (syllable identification) was required in this condition. In condition C (single task 2), participants heard speech-shaped noise and saw a nameable image. Only a response to task 2 (picture naming) was required in this condition. In condition D (dual-task), participants heard a syllable and saw a nameable image. Responses to both task 1 and task 2 were required. Condition D was our main condition of interest. In conditions A-D, the auditory and visual stimuli were presented with an SOA of 0ms. In conditions E and F, there was a 1000ms SOA between the onset of the auditory stimulus and the onset of the visual stimulus. In condition E (1000ms task 1 and 2), participants heard a syllable and saw a nameable image. Responses to both tasks were required. In condition F (1000ms no task), participants heard speech-shaped noise and saw an unnameable image. No responses to either task were required in this condition.

Marti and colleagues fit their task 1 and task 2 classifiers, trained with only one stimulus at trial onset, to trials at all SOAs. In the current experiment, conditions E and F had only one stimulus input at trial onset, with 1000ms separating task 1 and task 2. Therefore, these conditions indexed isolated task 1 and task 2 processing. We were interested in how the isolated task classifiers would generalise to the dual-task. However, in a dual-task, two stimuli are presented and participants must carry out perceptual and task-specific processing on both stimuli. We also wished to focus on task processing separated from perceptual stimulation, and thus, conditions A, B and C were included to enable training of task 1 and task 2 classifiers with two simultaneous perceptual inputs. Classifiers trained on conditions A, B and C were also tested on the dual-task.

The dual-task consisted of 720 trials, divided into six blocks of 120 trials. Twenty trials from each of the six conditions were presented per block. Trial order within a block was pseudo-randomised per block per participant with the constraints: three intervening items between presentation of the same image; two intervening items between presentation of images beginning with the same phoneme (bus and bal); no more than four trials successively containing the same sound; no more than five successive occurrences of a nameable or unnameable image; no more than eight successive trials with the same SOA. Each block lasted approximately six minutes. Each trial began with a fixation cross presented for a random interval of between 600 and 800ms. The sound was then immediately presented. At the 0ms SOA, the image was displayed simultaneously with the sound for 100ms. A blank screen was then presented for a random interval of between 2150 and 2650ms. At the 1000ms SOA (conditions E and F), a blank screen was displayed immediately after sound onset. 1000ms after sound onset the image was displayed for 100ms. Then a blank screen was displayed for a random interval of between 1150ms and 1650ms.

Participants first practised the dual-task with 36 trials. After 12 of these practice trials, participants were invited to ask the experimenter for clarification of task requirements, and the experimenter reiterated the instructions. Participants then continued with the practice trials, before beginning the experimental dual-task trials proper.

Procedure

Participants were brought to the lab and the experimental procedure was explained. Participants signed an informed consent form, and were given MEG-compatible clothing to change into for the experiment.

Nine Ag/AgCl electrodes were used to measure muscular movements. vEOG and hEOG were measured from above and below the right eye, and at the outer canthi of each eye. Mouth movement-related EMG were measured from above and below the right side of the mouth, approximately on the orbicularis oris. ECG was measured with one electrode placed slightly above the right clavicle

Table 6.1: Information about the dual-task conditions

Condition name	Letter	SOA	Response	Stimuli
No task	A	0	No responses	Noise & unnameable image
Single task 1	B	0	Task 1 response only	Syllable & unnameable image
Single task 2	C	0	Task 2 response only	Noise & nameable image
Dual-task	D	0	Respond to both tasks	Syllable & nameable image
1000ms task 1 and 2	E	1000	Respond to both tasks	Syllable & nameable image
1000ms no task	F	1000	No responses	Noise & unnameable image

and the other on the left of the infracostal line. A ground electrode was placed behind the participant's left ear. Electrode impedences were kept at or below 20 kΩ. Participants were then shown each image that would be presented in the experiment.

Participants were seated upright in the MEG system in a magnetically-shielded booth. Pillows were placed to the left and right of the participants, between them and the chair, under the knees and behind the back, in order to maximise comfort and minimise movement. The procedure was re-explained and participants were asked if they had any questions before being raised into the MEG helmet. All signal recordings were checked and the participant's head was localised before beginning the experiment.

Participants began the experiment with the picture naming task consisting of 80 trials. This was followed by the syllable identification task consisting of 75 trials. These first and second tasks were treated as practice trials and the data are not analysed here. Following this, participants practised the dual-task trials as explained above. Participants were then asked if they had any further questions, before starting the dual-task trials. Participants were able to take a break between each of the six dual-task blocks. In these breaks, the experimenter adjusted the participant's head position if participants had moved from their starting head position.

After the experiment, participants were debriefed and asked about any adverse events. The experiment proper took one hour and the whole experimental session took two hours.

MEG data acquisition and apparatus

MEG data were acquired with a CTF system (VSM/CTF systems, Port Coquitlam, Canada) with 275 axial gradiometers. Five gradiometers (MRF66, MLC11, MLC32, MLF62, MLO33) were disabled for technical reasons. Three head localisation coils were attached to the anatomical landmarks of nasion and left and right preauricular points for each participant. These coils determined the head position relative to the gradiometers. Head position was monitored online with a custom Matlab script (Stolk, Todorovic, Schoffelen, & Oostenveld, 2013). Data were recorded at a sampling rate of 1200Hz and stored for offline analysis.

The experiment was presented using Presentation software (Presentation Neurobs; version 16.4) running on Windows 7. The experiment screen was projected to participants in the MEG room using a PROPixx projector. Sounds were played through air tubes (ER2 insert earphones, Etymotic) attached to plastic ear molds comfortably placed in participants' ears. Verbal responses were recorded using a Sennheiser ME64 microphone present in the magnetically shielded MEG room. Manual responses were collected with fiber optic response pads (Current Design, model HH-2x4-C).

Behavioural preprocessing and analysis

Task 1 syllable identification RTs were measured from the onset of the sound stimulus. Task 2 naming RTs were semi-automatically measured using Praat (Boersma, 2002) and measured from the onset of picture presentation to speech onset. Trials were discarded if RTs were lower than 200ms, responses were incorrect (either incorrect buttons or incorrectly produced names, silence, or anything other than a fluent naming response), or if responses were given in the wrong order (naming before button pressing). The first two trials of each block were removed. Any participant with errors in more than 20% of trials in any of the six conditions was removed from the analysis. This resulted in removal of four participants (two further participants were removed due to bad quality MEG data), leaving data from 24 participants for analysis (excluded trials = 12%; see Appendix Table S1 for trial counts per participant per condition).

Behavioural data were analysed in R (R Core Team, 2017) with linear mixed effects models using the lme4 package (Bates et al., 2015). RTs were log-transformed prior to analysis to reduce skew. Categorical variables were treatment coded. We take |t| greater than 2 to be a significant value. Confidence intervals were calculated using the 'profile' option in the confint.merMod function in lme4. Error data were not analysed.

MEG preprocessing

All data were pre-processed with the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) in Matlab (version 9.0.0, R2016a, The MathWorks, Inc., Natick, MA). Conditions A to D were epoched from 500ms before baseline to 1500ms after. Conditions E and F were epoched from 500ms before baseline to 2500ms after, ensuring enough time after picture onset (1000ms after trial onset) to capture picture naming. Data were high-pass filtered with cut-off at 1Hz (transition width 2Hz, stopband 0-0Hz, passband 2-600Hz) and low-pass filtered with cut-off at 12Hz (transition width 3Hz, passband 0-10.5Hz, stopband 13.5-600Hz) with a Hamming-windowed sinc one-pass, zero-phase finite impulse response filter. Behavioural error trials were removed from the MEG data. Conditions E and F were re-epoched into two trial types: task 1 trials and task 2 trials. For task 1 trials, epochs lasted from -500ms to 1500ms after auditory stimulus (task 1) onset. For task 2 trials, epochs lasted from -500ms to 1500ms after visual stimulus (task 2) onset. Note that this task 2 baseline period included processing relating to task 1. All data were baseline corrected with a baseline window of -500ms to 0ms.

MEG analysis: Temporally-generalised MVPA

We analysed the MEG data using temporally-generalised MVPA (King & Dehaene, 2014) implemented in MNE Python (Gramfort et al., 2014) to compare the topographic distributions of MEG activity between two conditions at the sensor level. MVPA was applied using Sci-Kit learn (Pedregosa et al., 2011). On a single subject basis, MEG data were taken from two conditions. The MEG data were scaled by channel for each classifier. A linear support vector machine (SVM; Chang & Lin, 2011) with penalty parameter C equalling 1 was applied to the two conditions to determine the hyperplane which best separated the two conditions at each sample. When training and testing on the same conditions, a five-fold cross-validation procedure was used, where the data were split into five folds and the classifier was trained on four of the folds (80%) and tested on the remaining fifth fold (20%). This procedure was iteratively applied until all folds were used for testing. When classifiers were tested on a different condition to that trained on (cross-decoding), and both the training and testing sets included trials from condition A (no task), a pseudo-cross-validation procedure was used. In this procedure, trial numbers were first equalised across conditions. Then, with equal trial numbers per condition in the training and testing data, the training and testing data were split into five folds (i.e., conditions A, B and D were split into five folds). Classifiers were trained on one of the five folds (e.g., trained to discriminate between conditions A and B) and tested on one fold of the test condition (e.g., tested in one fold of condition D). This procedure was iteratively applied until all five folds were used for training and testing. In this way, the same subset of A was never in both the training and testing sets, and all subsets of B and D were iterated through. Without pseudo-cross-validation in cross-decoding, we found that classification accuracy was almost perfect along the diagonal and in the baseline period. This was due to the classifiers 'memorising' the conditions perfectly, due to condition A being present in the training and testing sets without cross-validation. It is well known that cross-validation is important in classifier analysis but this example highlights that importance.

To generalise across time, the classifiers trained at every sample t were tested at all samples in the test data t'. This provided a training time by testing time matrix. The diagonal of the matrix corresponded to when classifiers were trained and tested on the same time point.

We trained two sets of task classifiers: task 1 classifiers and task 2 classifiers. For task 1 classifiers we trained the classifiers to discriminate between task 1 responses in condition A (no task) vs condition B (single task 1). In condition A noise was presented and in condition B a syllable was presented. For task 2 classifiers we trained the classifiers to discriminate between task 2 responses in condition A vs condition C (single task 2). In condition A an unnameable picture was presented, and in condition C a nameable image was presented. We then tested both the task 1 and task 2 classifiers on their ability to decode activity in condition D (the dual-task condition; where a syllable and nameable image were presented) versus condition A. We were interested in whether, and when, the task 1 and task 2 classifiers were able to decode any information in the dual-task. We also trained isolated task 1 and task 2 classifiers to discriminate between conditions E (1000ms task 1/2) and F (1000ms no task). However, these classifiers were not tested on the dual-task data due to methodological issues (see Results for further discussion).

Classifier accuracy can be evaluated based on the area-under-the-curve (AUC) calculated from receiver operative characteristic (ROC) curves. AUC values describe the ratio of true positives and false positives, with values spanning from 0 to 1. An AUC of 0.5 is at chance level, where the classifier was unable to decode whether the sample belonged to either of the two conditions. An AUC of 1 means the classifier is always correct, and an AUC of 0 indicates reliable misclassification. AUCs less than 0.5 can be observed when classifying across conditions and indicates that the probability of false positives was higher than true positives (King & Dehaene, 2014). Accuracy scores were determined from the average AUCs across the five cross-validation folds.

We determined whether the AUC values were significantly different from the chance value 0.5 using non-parametric one-sample cluster-based permutation tests (Maris & Oostenveld, 2007) with an alpha of 0.05 as implemented in MNE Python. Briefly, cluster-based permutation tests perform a t-test at each time point (sample). T-values for adjacent samples exceeding the significance threshold are grouped into clusters. For each cluster, the sum of the t-values is used as the cluster t statistic. This outcome is compared to a null distribution to control for type I error. The null distribution is derived by randomly scrambling samples 1000 times using the Monte Carlo method and the same procedure is run on these null distributions. The real and null distribution cluster-level statistics are compared and clusters falling into the top or bottom 2.5% are considered significant. AUC scores were scaled to be on the correlation scale (-1 to 1) and then Fisher transformed to put them on an unbounded scale for statistical analysis. Statistical tests were performed across participants. We tested for significant accuracy over the entire trial (-500ms to 1500ms). We discuss any classifier accuracy scores significantly different from chance as determined by the cluster test. Note our analysis technique differs from that of Marti et al. (2015), due to differences in design. As we only tested dual-tasking with one SOA rather than four, we did not compare the onsets and offsets of decoding accuracy across SOA. Marti et al. (2015) also tested the significance of their AUC scores using a threshold derived from the average AUC score in a portion of the baseline window. We chose to test against chance level using cluster-based permutation tests to prevent any accidental bias that could be introduced when manually determining a threshold to test against for significance.

Results

Participants gave a button-press response to the syllable identification task (task 1) in three conditions: single task 1, dual-task, and 1000ms task 1. A spoken response for picture naming (task 2) was given in three conditions: single task 2, dual-task and 1000ms task 2. These conditions are summarised in Table 6.1 and will be referred to by name throughout the results.

Behavioural results

Task 1: Syllable identification

Participants on average took 717.4ms (SD = 228) to respond in single task 1 (when an unnameable image was also presented), 792.2ms (SD = 207) in the dual-task, and 707.5ms (SD = 210) in 1000ms task 1. The data are shown in Figure 6.1.

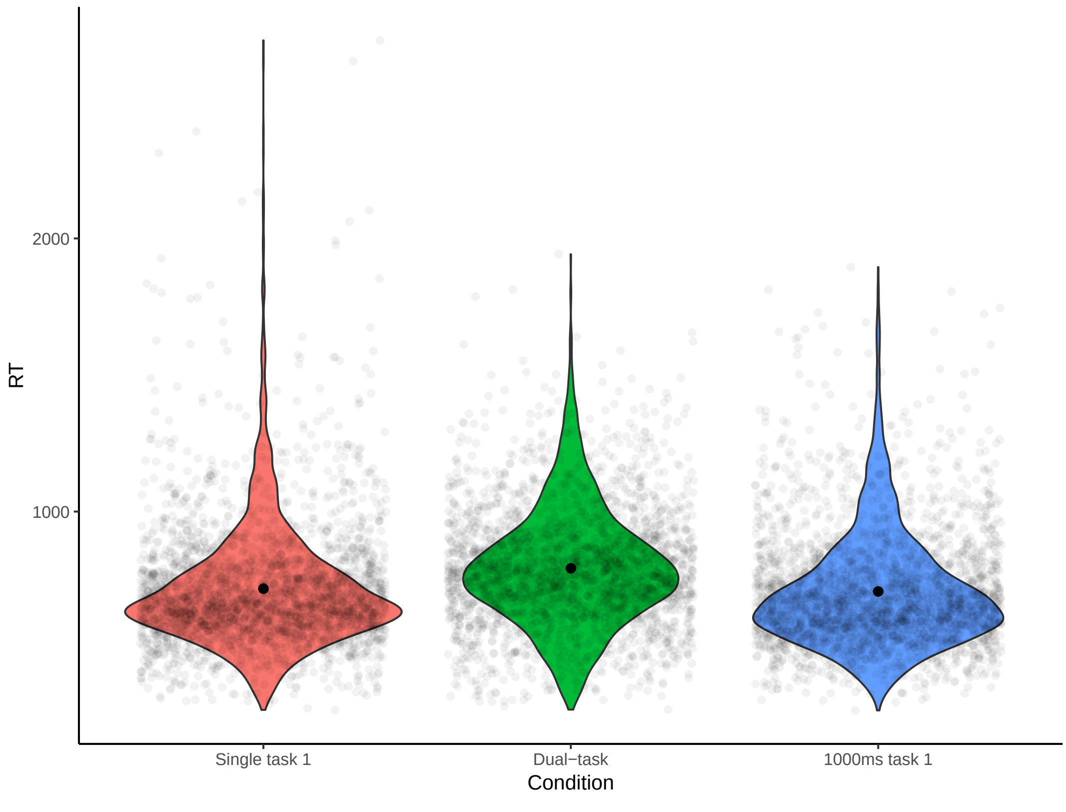

Figure 6.1: Task 1 syllable identification RTs by condition.

A linear mixed effects model was used to analyse the data with log-transformed RT as the dependent variable, a fixed effect of condition, random intercepts by participant, task 1 stimulus and task 2 stimulus, and a random slope of condition by participant. Single task 1 was treated as the baseline condition and was in turn compared to the dual-task and the 1000ms task 1 conditions. RTs in the dual-task condition were significantly longer than in single task 1 (estimate = 0.05, SE = 0.007, t = 6.22, CI [0.031 0.06]). There was no significant difference between single task 1 and 1000ms task 1 (estimate = -0.004, SE = 0.006, t = -0.64, CI [-0.015 0.008]).

These results show that carrying out syllable identification in the context of a dual-task results in slower RTs to the syllable task compared to when carrying out that task when an unnameable image is displayed. Interestingly, the similarity of RTs in the single and 1000ms task 1 conditions suggest that the presence of to-be-ignored visual input does not affect RTs to the auditory task.

Task 2: Picture naming

Mean naming RT was 835.95ms (SD = 182) in single task 2 (when noise was also presented), 1298.61ms (SD = 232) in the dual-task, and 662.15ms (SD = 142) in 1000ms task 2. These results are presented in Figure 6.2.

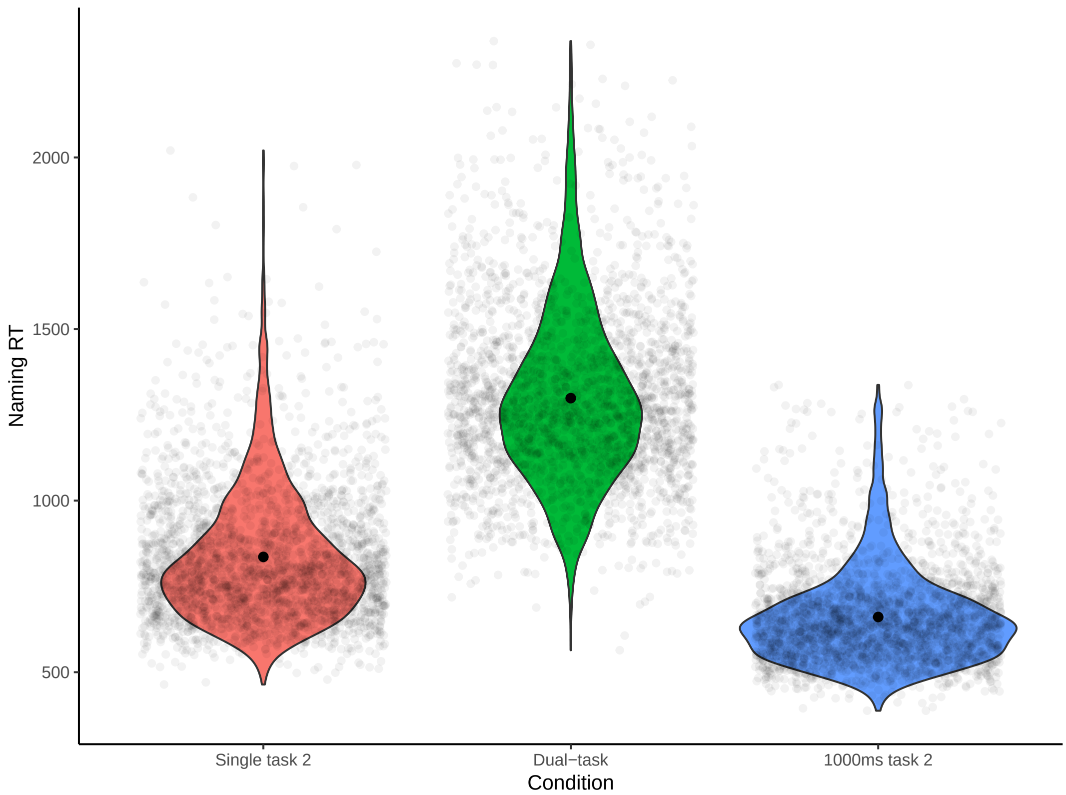

Figure 6.2: Task 2 picture naming RTs by condition.

A linear mixed effects model with log-transformed RT as the dependent variable, a fixed effect of condition, random intercepts by participant, task 1 stimulus and task 2 stimulus, and random slopes of condition by participant and by task

2 stimulus was run. The single task 2 condition was the baseline condition and was in turn compared to the dual-task and 1000ms task 2 conditions. Naming RTs were slower in the dual-task than single task 2 (estimate = 0.19, SE = 0.007, t = 26.55, CI [0.18 0.21]), and faster in 1000ms task 2 than single task 2 (estimate = -0.1, SE = 0.007, t = -14.1, CI [-0.11 -0.09]). Naming RTs were thus fastest when images were presented without auditory input (1000ms task 2), intermediate when an auditory stimulus was heard but not responded to (single task 2), and slowest in the dual-task condition.

The significant difference between the single and 1000ms task 2 conditions is contrary to the absence of such an effect in the task 1 RTs. This suggests that when naming, the presence of an auditory stimulus affects the processes involved in picture naming, even if the auditory stimulus should be ignored. This may be because the general process of naming involves many stages and a distracting stimulus disrupts this process to a greater extent than responding to a syllable when also seeing a distracting stimulus. However, this effect may instead be due to task ordering, as syllable responses should always be given before image responses. Thus, participants may establish a task set which entails responding to task 1 before responding to task 2. In the single task 2 condition, participants must ignore the auditory stimulus, which goes against a possible task set expectation. Violation of the task set may cause interference in the naming process, or may delay the naming process due to participants deciding not to respond to the auditory stimulus. From these data we cannot distinguish between these two possibilities.

Classification results

As described above, a set of task 1 classifiers (syllable identification) were trained to discriminate between the no task and single task 1 conditions. The set of task 2 (picture naming) classifiers were trained to discriminate between the no task and single task 2 conditions. These two sets of classifiers were then tested on their ability to decode information in the dual-task versus no task conditions. We also trained classifiers on 1000ms task 1 and 1000ms task 2 versus 1000ms no task[1].

[1]We trained, and more importantly, tested classifiers on discriminations between conditions for two reasons. Firstly, we were interested in whether, based on the trained discrimination, classifiers could accurately decode dual-task versus no task trials, to indicate that the dual-task trials were more similar to the single task. Secondly, we wished to have ROC-AUC scores as our classification accuracy measure. This is only possible when testing on a discrimination. Note that Marti et al. (2015) also appear to have tested their classifiers on a discrimination, as they obtained ROC-AUC scores, but it is unclear from their methodology exactly what their discrimination conditions were when testing.

As a reminder, classifiers were trained at time t and tested at time t', resulting in a training time by testing time matrix of classification accuracies. Training and testing information is listed in Table 6.2 for convenience[2].

Table 6.2: Classification information

Trained on	Tested on	Name	Why?
EvsF task 1	EvsF task 1	Isolated task 1	Train task 1 with one presented stimulus
AvsB	AvsB	Task 1	Train task 1 syllable identification
AvsB	AvsD	Task 1 on dual-task	Do task 1 classifiers generalise to the dual-task?
EvsF task 2	EvsF task 2	Isolated task 2	Train task 2 with one presented stimulus
AvsC	AvsC	Task 2	Train task 2 picture naming
AvsC	AvsD	Task 2 on dual-task	Do task 2 classifiers generalise to the dual-task?

Classification across time: Task 1 (syllable identification) and Task 2 (picture naming)

Figure 6.3 shows classifier accuracy for the task 1 classifier (left figure) and for the task 2 classifier (right figure). In both figures (and all further figures), grey areas indicate accuracy scores which were not significantly different from chance (AUC = 0.5, with a cluster-level p value of 0.05). Red areas are AUC scores significantly greater than chance, and blue areas are AUC scores significantly lower than chance. The legend shows AUC scores mapped to colours.

For both task 1 and task 2 we found that classifiers decoded in successive, partially overlapping steps (cf. Marti et al., 2015), due to the smooth stream of classification performance along the diagonal. For task 1, classifiers generally decoded over a window of around 200ms (i.e. successful decoding spanned 200ms horizontally across the diagonal line), and for task 2, classifiers decoded over a window of around 300ms. For both tasks, the initial 200-250ms of decoding was more sustained and stable, as seen by above-chance decoding accuracy between 0 and 250ms between training and testing time. During this window, there is a more block-like pattern of accuracy, indicating that during auditory

[2]MEG topographies of conditions A-D are presented in the Appendix for the interested reader.

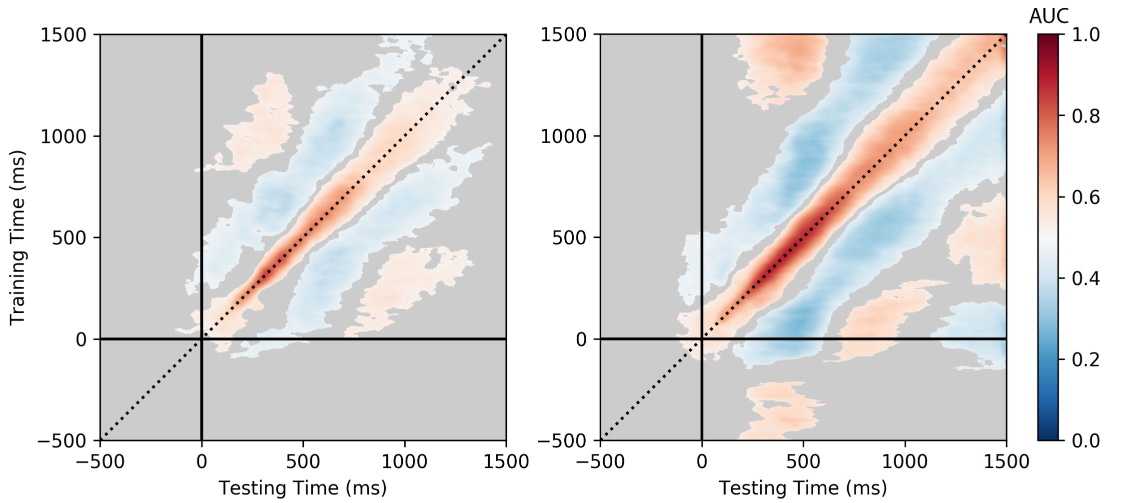

Figure 6.3: Temporally-generalised classifier performance for classifiers trained and tested on the same condition. Left: task 1 syllable identification. Right: task 2 picture naming. The legend refers to AUC scores. Areas in grey are where performance was not significantly different from chance. AUC scores above 0.5 are significantly higher than chance, and below 0.5 significantly lower than chance.

and visual perception of the stimuli there was a sustained and stable information pattern decodable. In task 1 (Figure 6.3 left) there was also a small narrowing of decodable activity at approximately 250ms, following the block-like pattern. This indicates that neural generators underlying processing in this time window were rapidly changing, leading to a shorter width of decodable activity.

For task 1 and task 2, there was significantly lower than chance AUC (in blue) in addition to the high decoding performance around the diagonal. Because AUC also measures bias, this is indicative of the classifier consistently selecting the incorrect label. Conceptually, this is similar to a person consistently calling their left side 'right' and vice versa. Below-chance decoding performance is challenging to interpret (see King & Dehaene, 2014), but may be due to polarity reversals in the signal, or inhibition of activated neural generators (see T. A. Carlson, Hogendoorn, Kanai, Mesik, & Turret, 2011; T. Carlson, Tovar, Alink, & Kriegeskorte, 2013; King, Gramfort, Schurger, Naccache, & Dehaene, 2014, for more discussion about below-chance decoding performance).

For both tasks, classifiers were able to decode information outside of the diagonal. This is shown by the decodable activity in the upper left and bottom right of each figure (after stimulus onset at the zero point), far from the diagonal. This indicates that patterns of activity found early in the trial were also found later.

In task 1, the activity pattern decoded between 0 and 400ms was also found between 700ms and 1200ms. In task 2, the activity pattern decoded between 0 and 300ms was also found between 600 and 1000ms. A possible interpretation for this effect is that perceptual information relating to visual and auditory processing was reactivated during response preparation and execution (task 1 mean RT = 717ms, task 2 mean RT = 836ms). Evidence shows that when cued, perceptual information can be reactivated (see e.g., Sergent et al., 2013), and perceptual simulation research suggests that similar neural generators are used in perception and reactivation of perception of a stimulus (e.g., Nyberg et al., 2001; Wheeler, Petersen, & Buckner, 2000). Thus, if participants reactivated representations associated with perceptual processes, similar topographies would arise, leading to accurate decoding. The timing of this reactivation suggests it is response-related.

We additionally trained and tested the isolated task 1 and task 2 classifiers. These classifiers were trained and tested on conditions E and F, where task 1 and task 2 were separated by 1000ms. Classification accuracy for these classifiers is displayed in the Appendix in Figure S5. In general, the classification pattern followed that described for task 1 and task 2 above. However, in task 2 there was above-chance decoding in the baseline interval. The baseline period here included task 1 processing, making the meaning of any effects after picture onset challenging to interpret. Because of the above-chance baseline accuracy, we did not test the isolated task 1 and task 2 classifiers on the dual-task condition.

Classification across condition: Do task 1 and task 2 classifiers generalise to the dual-task?

We tested the task 1 and task 2 classifiers on their ability to decode information in the dual-task. Figure 6.4 shows classification accuracy for the task 1 classifiers (left figure) and the task 2 classifiers (right figure). For the task 1 classifiers tested on the dual-task data, we see a very similar pattern to Figure 6.3. This similar pattern suggests that in a dual-task, syllable identification is not qualitatively affected compared to when syllable identification is carried out alone.

We see the similarities between the task 1 classifier trained and tested on itself, and tested on the dual-task, more strikingly in Figure 6.5. Only AUC scores significantly above chance are plotted. The training data were always the no task versus single task 1 conditions. Accuracy in green represents classifier accuracy only present in task 1. Accuracy in orange represents classifier accuracy common to task 1 in the single task and the dual-task (i.e. above chance accuracy for this training-testing pair in both the single task and the dual-task). Accuracy in blue

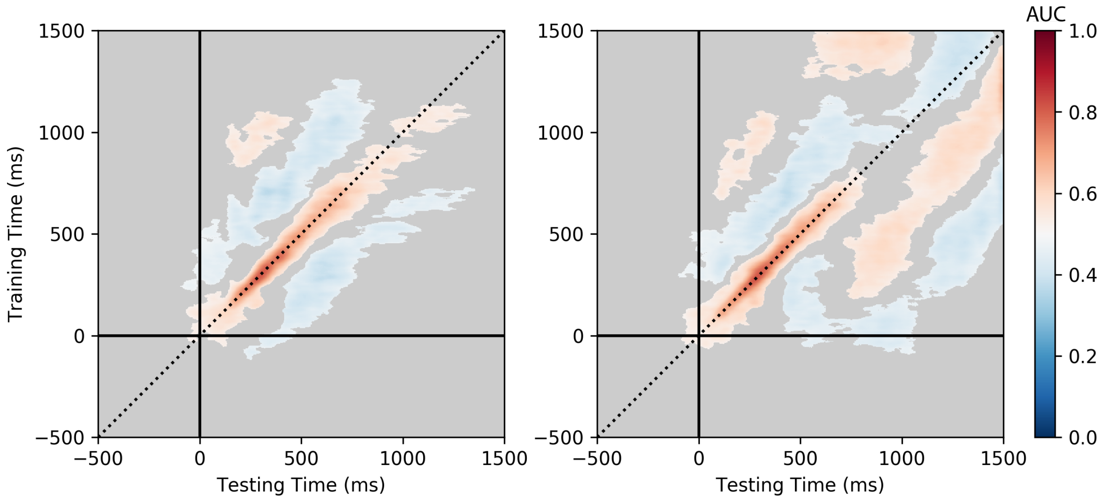

Figure 6.4: Temporally-generalised classifier performance for classifiers trained on the single tasks and tested on the dual-task. Left: task 1 syllable identification classifiers applied to the dual-task. Right: task 2 picture naming classifiers applied to the dual-task. The legend refers to AUC scores. Areas in grey are where performance was not significantly different from chance. AUC scores above 0.5 are significantly higher than chance, and below 0.5 significantly lower than chance.

represents classifier accuracy present only in the dual-task. We found that the majority of the decodable information was equally well classified in the single task and the dual-task, and there was almost no decodable information present only in the dual-task. This shows that during a dual-task, the processing stream of syllable identification is not qualitatively affected by the second task. This may be due to the ease of combining syllable identification with a secondary task, or due to participants strongly prioritising task 1 leading to very similar processing as when carrying out syllable identification alone.

For the task 2 classifiers, we found a different pattern of results (right panel of Figure 6.4). In the dual-task, task 2 classifiers decoded activity along the diagonal from stimulus onset to approximately 700ms, over a window of around 200ms. Because task 1 classifiers were also able to decode task 1 activity during the same time window, this indicates parallel processing of task 1 and task 2. However, classifiers also decoded later activity, such that the topographic activation pattern in task 2 was found in a later reactivated pattern (King & Dehaene, 2014). This reactivation was present for information after 250ms in single task 2. This second decoding pattern was consistent until trial offset. This pattern of reactivation was unexpected as it is not predicted by any theories of dual-tasking (see Discussion).

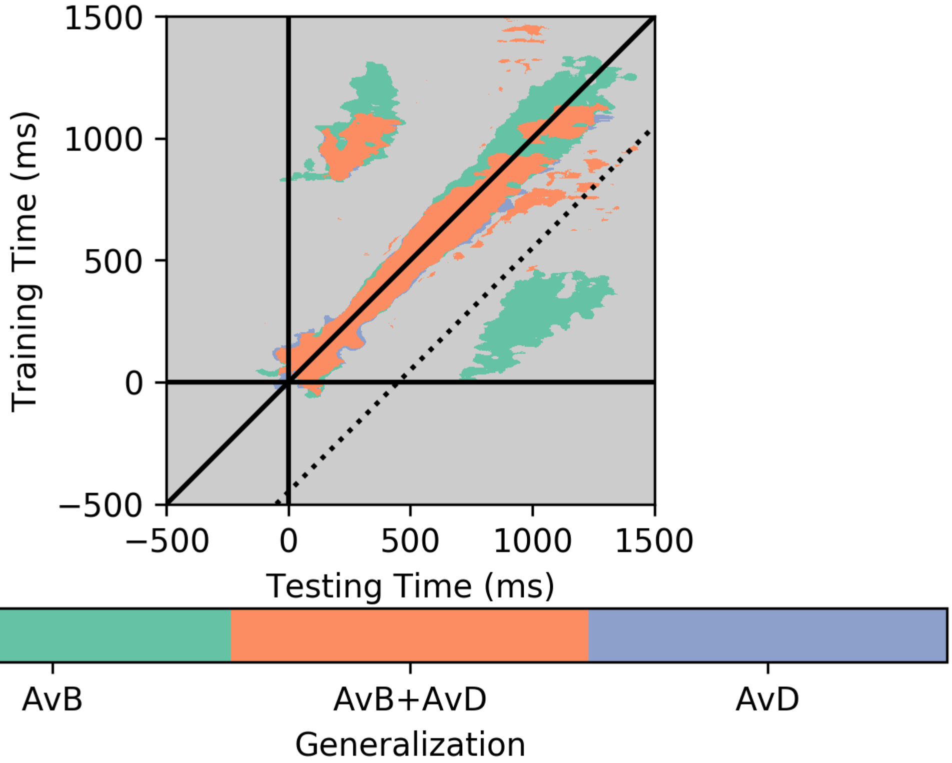

Figure 6.5: Temporally-generalised classifier performance for task 1 syllable identification, tested on the single and the dual-task. Only AUC scores above chance are plotted. Areas in grey are where performance was not significantly different from chance. Activity in green reflects above-chance classifier performance only found in the single task (AvsB). Activity in orange reflects above-chance classifier performance found in both the single and dual-task (AvsB and AvsD). Activity in blue reflects above-chance classifier performance found only in the dual-task (AvsD). The dotted line at 450ms shows no activity along this diagonal.

This pattern is demonstrated more clearly in Figure 6.6. Green areas represent classifier accuracy only present in the single task. Orange areas represent classifier accuracy common to task 2 in the single task and the dual-task. Blue areas represent classifier accuracy present only in the dual-task. The later pattern of activity was present only in the dual-task, and a demonstrative line at 450ms is plotted to indicate that the whole activity pattern is shifted rightwards by approximately 450ms. This indicates that task 2 processes in the dual-task began with a similar task onset (as shown by overlap in orange and green around the diagonal), but were reactivated only in the dual-task. Importantly, this reactivation delay is in line with the difference in naming RTs between single task 2 and the dual-task. Naming RTs in the dual-task were on average 463ms later than in the single task (single task = 836ms, dual-task = 1299ms). Therefore, the re-activation was shifted in time aligned with the naming response in the dual-task.

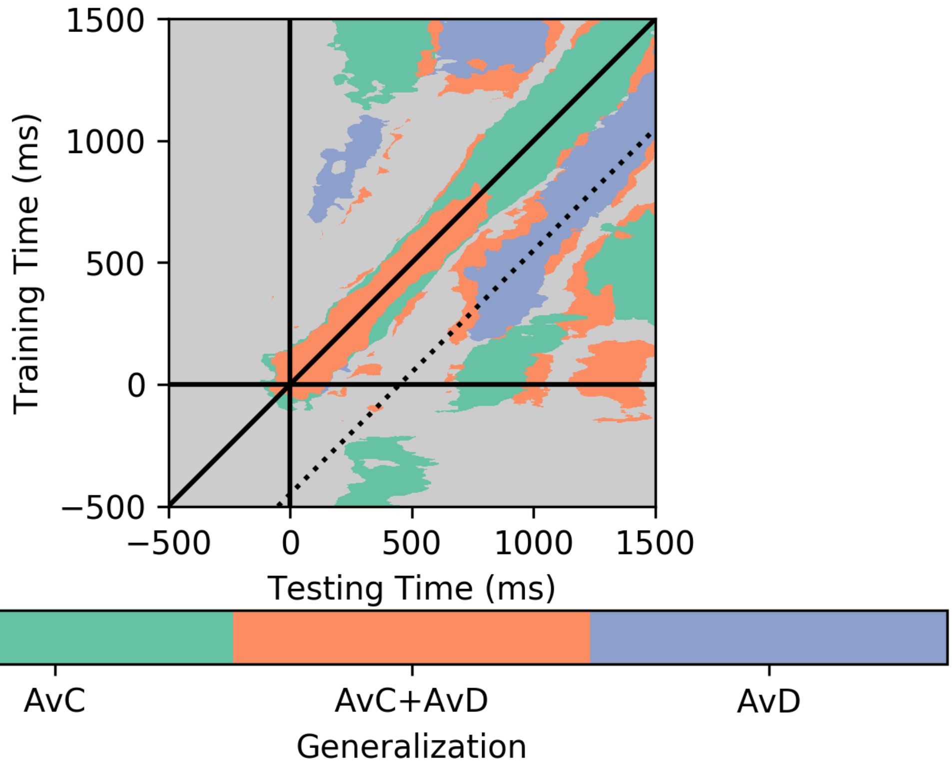

Figure 6.6: Temporally-generalised classifier performance for task 2 picture naming, tested on the single and the dual-task. Only AUC scores above chance are plotted. Areas in grey are where performance was not significantly different from chance. Activity in green reflects above-chance classifier performance only found in the single task (AvsC). Activity in orange reflects above-chance classifier performance found in both the single and dual-task (AvsC and AvsD). Activity in blue reflects above-chance classifier performance found only in the dual-task (AvsD). The dotted line at 450ms shows reactivated activity along this diagonal only in the dual-task.

The onset of reactivation in the dual-task was roughly 700ms. Responses to the syllable were given on average at 792ms. Thus, reactivation was manifest after most of the response planning for task 1 was completed. Only the topographic pattern from single task 2 after 250ms reactivated, indicating that initial processes (most likely involving visual and auditory stimulus perceptual processing) were carried out similarly in a dual-task compared to a single task. This therefore suggests that the linguistic processes involved in picture naming were reactivated, but initial perceptual processes were not[3].

Discussion

In this experiment we aimed to test whether temporally-generalised MVPA could uncover the time course of processing of two tasks in a dual-task, and whether there were any differences in the time course of the tasks when in a dual-task compared to a single task setting. Participants carried out syllable identification as task 1 and picture naming as task 2. We found that syllable identification showed a similar sequence of topographical MEG brain responses when carried out as a single task compared to in a dual-task, whereas the series of topographies elicited by picture naming were strongly affected when dual-tasking.

Regarding our first two questions of whether classifiers would be able to generalise within and across conditions, we found that our classifiers could successfully decode activity in a dual-task when trained on a single task. In a previous dual-tasking experiment where temporally-generalised MVPA was applied, two simple tasks – tone identification and letter identification – were used (Marti et al., 2015). However, the tasks used in the present study are arguably more complicated. Thus, the current study demonstrates that with tasks which are more complex and where both tasks tap language-related resources, temporally- generalised MVPA can reveal differences in the time courses of dual-task versus single task processing.

Regarding task 1, we found that syllable identification was carried out very similarly in dual- and single-task contexts (see Figure 6.5). This suggests that in a dual-task, syllable identification is not qualitatively affected by the dual-task and is carried out broadly with the same task dynamics in a dual-task as when carried out alone. This parallels the results of Marti and colleagues, who also found that their task 1 tone identification task was not strongly affected when carried out in a dual-task compared to a single task. This may be because the processes involved in syllable identification are carried out in parallel with another task with no performance decrements. Alternatively, as participants were instructed to prioritise task 1 over task 2, they may have allocated greater capacity

[3]See Figure S6 where the task 1 classifier was tested on task 2 (i.e. trained on no task versus single task 1 and tested on no task versus single task 2), to determine how much activity was decodable between syllable identification and picture naming. We found that task 1 classifiers did not generalise well to task 2, indicating that the task 1 and task 2 classifiers were decoding different information.

to task 1, leading to similar task performance whether in a dual-task or not. However, despite the broad similarities, there are some decoding differences in task 1 between the single and dual-task (see the green decoded activity in Figure 6.5, which corresponds to activity only found in the single task). We also found that RTs to syllable identification in the dual-task were significantly longer (by 75ms) than in the single task, despite the similarity in topographic brain activation patterns along the diagonal. It may be that processes decoded only in the single task are responsible for this behavioural difference. Further research should investigate how classification and behavioural results can be aligned.

Regarding task 2, we found that picture naming was strongly affected in a dual-task compared to a single task. Classifiers trained on single picture naming were able to decode activity in the dual-task around the diagonal for the first 750ms of task processing (the orange pattern around the diagonal in Figure 6.6). However, there was also reactivation of activity at a later time point in the dual-task (the blue pattern of activity in Figure 6.6). There are three things to note about the reactivated pattern. Firstly, the reactivation pattern is decoded from around 700ms in the dual-task (i.e., it begins at 700ms along the x axis of Figure 6.6). The RT results show that responses to task 1 syllable identification are given on average 792ms into the dual-task. Therefore, the timing of the reactivation suggests that while processes involved in picture naming began with a similar time course in the dual-task to in the single task, indicating parallel processing, some processes were later reactivated *after* the response to task 1 had been given. Secondly, the reactivated pattern corresponds to decodable information from 250ms in the single task (i.e., the pattern begins from 250ms on the y axis of Figure 6.6). Due to this timing, it may be that the reactivated processes are linguistic, post-perceptual, processes. This would suggest that initial perceptual encoding of the visual and auditory stimuli occur at the onset of the trial (as is expected), but are not later reactivated. Thirdly, and most importantly, the timing of the reactivated pattern is in line with the naming RT difference between the single and dual-task. Participants named pictures 463ms later in the dual-task than the single task. The illustrative dotted line in Figure 6.6 is plotted at 450ms, as an approximate demonstration of the diagonal of this process. The fact that the reactivated pattern aligns so closely with the behavioural difference suggests that the reactivated activity contains linguistic processes required during picture naming.

We also find that the reactivation of early picture naming processes begins at the same time that later stages are still being carried out. For example, at 700ms in the dual-task (on the y axis of the right panel of Figure 6.4), we find significant above chance classification of patterns from 250ms (from the x axis; the reactivated pattern) and 600ms (from the x axis; along the diagonal) in the single task. The classification pattern both along the diagonal and in the reactivated stage is smooth and continuous. Both of these findings support the interpretation that picture naming processes are cascading, such that a process does not need to be complete before later processes begin. If this were the case, we would expect a more block-like pattern in the decoded activity (cf. King & Dehaene, 2014). We would also expect no overlap between the parallel and reactivated stages. These results fall in line with language production theories, which assume information cascades from one stage to another without the earlier stage needing to be complete (Dell, 1986; W. J. M. Levelt et al., 1999).

We can speculate on why this reactivation occurs. One possibility is that when syllable and naming processes are carried out simultaneously, naming processes are hindered in some way such that they are not carried out in enough depth to produce satisfactory representations, and hence reactivation of processing is required to complete the task. This implies that reactivation is due to the processes being hindered by concurrent task 1 processing. For example, it may be that lexical processes are carried out, but the output from these processes are not satisfactory for later downstream processes, such as phonological processes. Thus, lexical processes must be carried out again. An alternative but related possibility is that linguistic activation processes involved in picture naming are somewhat automatic, and begin regardless of whether there is a secondary task. However, other linguistic processes, such as selection, require capacity which is not allocated during a dual-task in the same way as in a single task, and thus the linguistic processes must be reactivated later for naming processes to be complete. For example, it may be that general conceptual, lexical and phonological activation of the picture name occurs in parallel with syllable activation processes, but without capacity allocated there is no selection at these levels. Capacity is allocated once task 1 processes are complete, but selection at these levels also requires reactivation at these levels. Thus, the reactivated pattern follows a similar time course to that seen along the diagonal. A third possibility is that during the parallel processing stage a perceptual representation is created for task 2, but further task 2 processing requires an event in task 1 to continue. For example, executing the response in task 1 could be responsible for triggering processing in task 2. This would suggest that the initial task 2 representation is 'retrieved' for later processing and task execution. Note that this third possibility does not assume that parallel processing involves any linguistic stages. However, we argue that this decoded activity must include some linguistic processes because the decodable activity lasts until 750ms into the trial, both in the single task and the dual-task. In the single task, participants name pictures with an average RT of 836ms, strongly suggesting that linguistic processing occurs during the first 700ms. Further research must be carried out to determine which processes are reactivated and why reactivation occurs.

The pattern of results found for task 2 does not fit with either of the main dual-task theories. The response selection bottleneck theory (Pashler, 1994) predicts that post-perceptual processes in task 2 are delayed until task 1 is complete. However, we found parallel processing of syllable identification and picture naming, evidenced by successful decoding along the diagonal for both tasks in the dual-task. The capacity-sharing theory (Navon & Miller, 2002; Tombu & Jolicœur, 2003) predicts parallel processing with extended processing in task 1 and task 2 when the tasks overlap. We do not find evidence for this, as in task 1 and task 2 classifiers decode over a similar time window in the single task and the dual-task (approximately 200ms). If anything, in task 2 the classifiers decode over a qualitatively *smaller* time window in the dual-task compared to the single task between 250 and 500ms, with earlier reductions in classification accuracy, which is opposite to a prediction of extended processing. In order to account for the data pattern under the capacity-sharing theory, we would have to posit an extension to the theory which states that, under some circumstances, there is parallel processing of tasks with no extension in processing time of different stages, but there is also later reactivation of one of the tasks because of the earlier capacity constraints. However, this extension goes against at least one of the fundamental underlying tenets of the capacity-sharing theory, namely that capacity is shared resulting in extended processing. Additionally, neither theory as it stands can account for the later reactivation of task 2 processes. The data also do not fit with the theory proposed by Marti et al. (2015), where they argue that tasks compete for conscious access (or attentional resources), as is evidenced by their data pattern. We find parallel processing within tasks that is decoded over a similar time window. While it may be the case that task 1 and task 2 compete for attentional resources, which results in later reactivation of task 2 processes, there is no evidence of competition when tasks are processed in parallel.

A different theory of dual-tasking which assumes that tasks are scheduled according to condition-action rules (EPIC-SRD; D. E. Meyer & Kieras, 1997) could account for this data pattern by assuming that naming processes are scheduled to occur in parallel, but if some processes are not completed then they are rescheduled for a later stage. Importantly, under the EPIC-SRD framework, there are no capacity limitations underlying scheduling. Thus, task 1 and task 2 can be scheduled to run in parallel, but if one of the tasks (here, task 2) cannot be carried out satisfactorily then this process is later reactivated. Alternatively, it could be that some processes in task 1 and task 2 are scheduled to be carried out in parallel, but due to task instructions other processes are not. This implies that it is not capacity limitations which result in reactivation but strategic scheduling.

More broadly, our results suggest a complex pattern of processing in speech planning in a dual-task. When planning speech while carrying out a concurrent linguistic task, processes in speech planning are carried out both in parallel and more serially, after processes in the other linguistic task are concluded. This suggests that when planning speech in dialogue, it is not the case that all speech planning processes are conducted while listening to an interlocutor, but at least some processes are reactivated after the listening 'task' is finished. Thus, while these results suggest that at least some processes are carried out in parallel with listening (Levinson & Torreira, 2015), this is not the full story. However, one must consider that these results may be driven by naming being task 2 in the dual-task, and thus the same patterns of activity may not be found if naming is prioritised. Future research should determine whether the processing dynamics of naming are affected by different task orders.

One criticism of our analysis technique is that it does not take differences in timings of processes within trials into account. While the classifiers do generalise across time, the underlying assumption is that processes occur with roughly the same time course across trials. However, this may not be the case. During a dual-task, the sequence of processing states may vary across trials for a variety of reasons, such as fatigue, the condition of the previous trial, or the attention allocation to the current trial, and thus while the same task processing stages are carried out, the timing of these stages could vary. An extension of decoding which is temporally unconstrained has recently been proposed (Vidaurre, Myers, Stokes, Nobre, & Woolrich, 2018), which can investigate the sequential order of processing states which may vary in their timing across trials. Further research using this decoding approach may be able to uncover different processing dynamics in linguistic dual-tasking.

In conclusion, our results suggest that in the dual-task of syllable identification and picture naming, syllable identification is not qualitatively affected by the dual-task, and is carried out very similarly as to when carried out alone. In contrast, picture naming is heavily affected by the dual-task, where picture naming processes are carried out in parallel with syllable identification and are reactivated approximately 450ms later. These results do not fit with the response bottleneck or capacity sharing theories of dual-tasking (Navon & Miller, 2002; Pashler, 1994; Tombu & Jolicœur, 2003), and instead suggest that the processing system tries to carry out the tasks in parallel. However, for currently unknown reasons, processes are later reactivated.

Appendix

Table S1: Number of trials per condition by participant

Participant	A	B	C	D	E	F
2	119	117	117	115	117	119
4	119	117	118	109	108	117
5	119	115	118	112	119	117
6	118	115	117	113	118	118
7	119	109	114	100	110	119
8	119	117	119	110	116	114
9	116	108	110	103	108	118
11	105	101	106	108	108	105
13	116	115	118	112	118	119
14	116	118	117	97	112	119
16	118	116	117	102	107	119
17	117	113	109	103	112	116
18	116	118	116	104	119	117
19	119	117	119	113	114	118
20	119	116	117	98	113	117
21	115	117	118	111	120	118
22	117	112	119	103	114	116
23	120	114	113	111	117	118
25	118	113	114	108	115	117
26	116	119	116	113	118	118
27	118	114	112	108	115	117
28	118	110	115	112	118	118
29	118	112	114	103	108	116
30	119	117	119	111	117	118

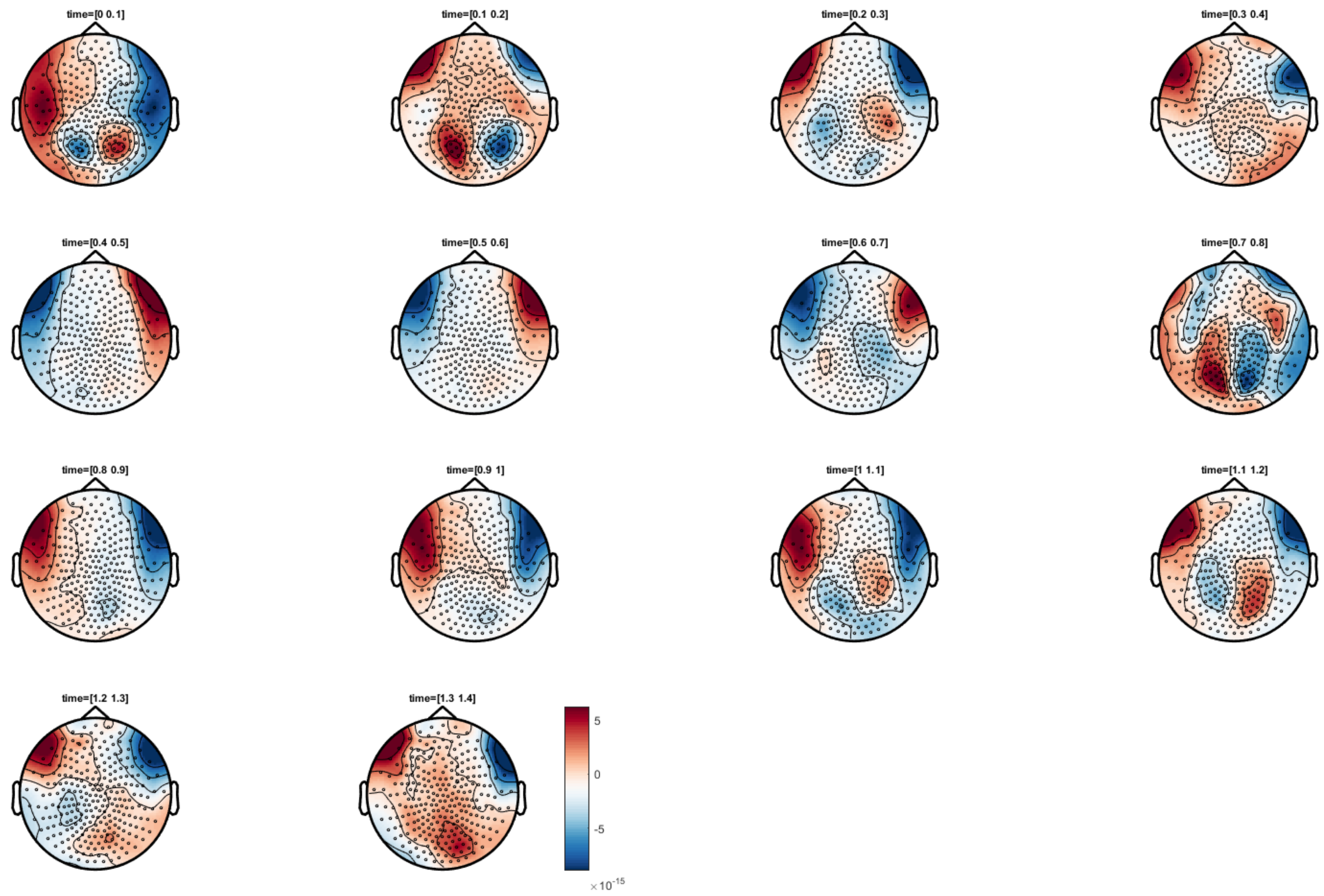

Figure S1: MEG topographic plots of condition A (no task). Plots are every 100ms from 0ms to 1400ms (the title above each image lists the time window in seconds). Axial gradiometer data are plotted.

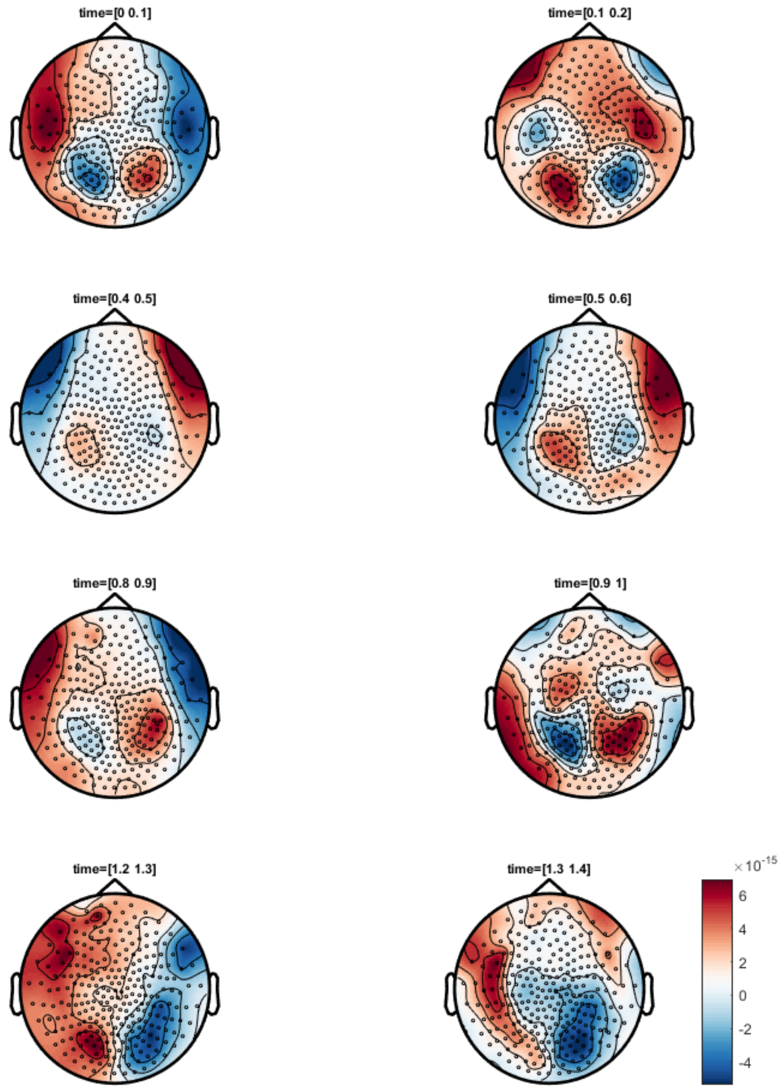

Figure S2: MEG topographic plots of condition B (single task 1). Plots are every 100ms from 0ms to 1400ms (the title above each image lists the time window in seconds). Axial gradiometer data are plotted.

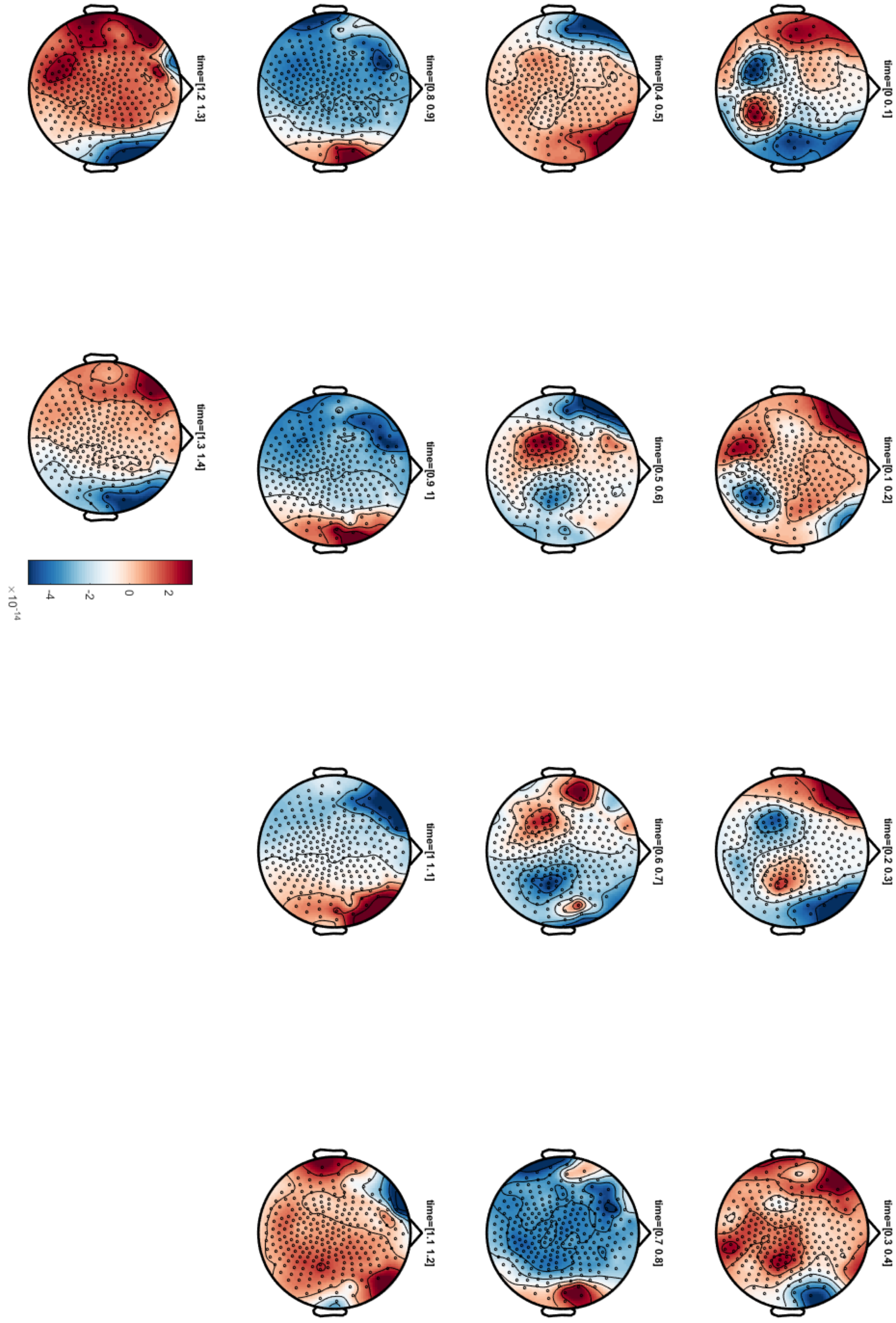

Figure S3: MEG topographic plots of condition C (single task 2). Plots are every 100ms from 0ms to 1400ms (the title above each image lists the time window in seconds). Axial gradiometer data are plotted.

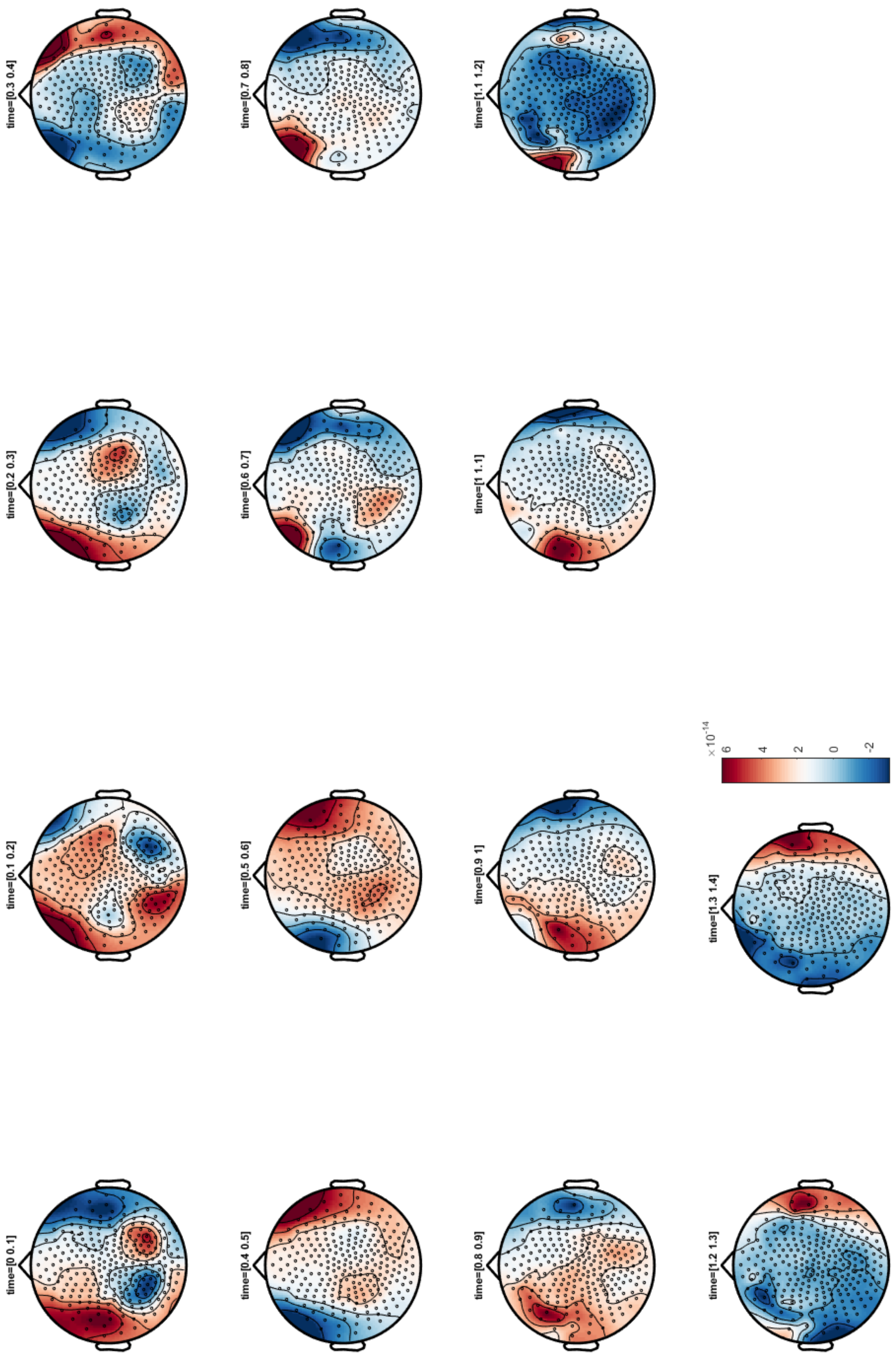

Figure S4: MEG topographic plots of condition D (dual-task). Plots are every 100ms from 0ms to 1400ms (the title above each image lists the time window in seconds). Axial gradiometer data are plotted.

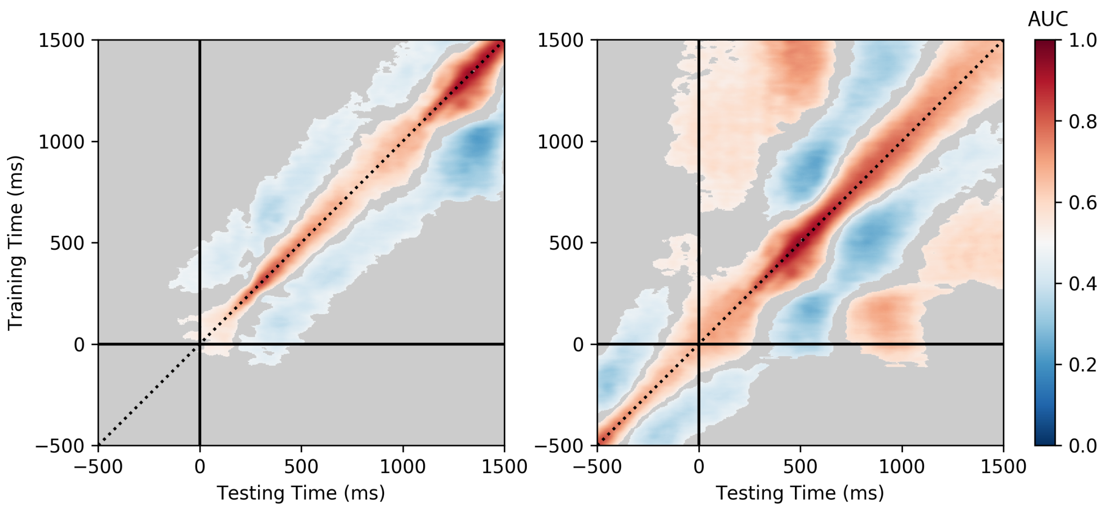

Figure S5: Temporally-generalised classifier performance for classifiers trained and tested on the same condition using data from SOA 1000ms. Left: task 1 syllable identification (when only syllables vs noise were presented). Right: task 2 picture naming (when only nameable vs unnameable images were presented). The legend refers to AUC scores. Areas in grey are where performance was not significantly different from chance. AUC scores above 0.5 are significantly higher than chance, and below 0.5 significantly lower than chance. Note the above-chance baseline decoding performance in the right figure.

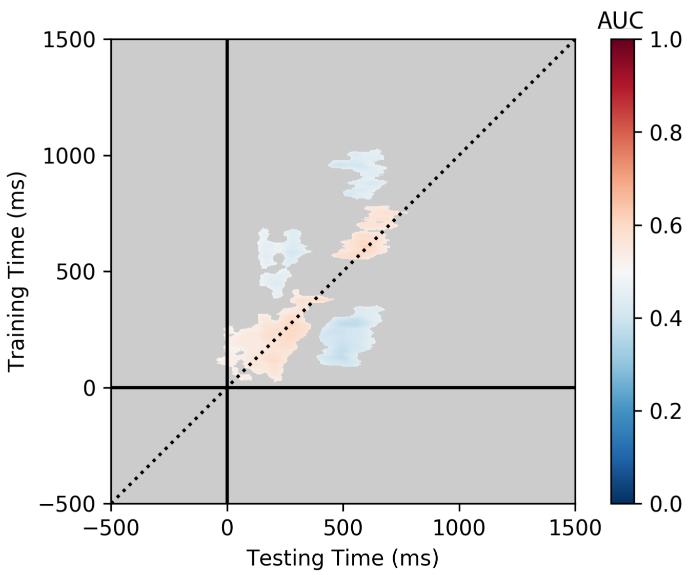

Figure S6: Temporally-generalised classifier performance for classifiers trained on task 1 syllable identification and tested on task 2 picture naming. The legend refers to AUC scores. Areas in grey are where performance was not significantly different from chance. AUC scores above 0.5 are significantly higher than chance, and below 0.5 significantly lower than chance. Note very little decoded activity, indicating that the two classifier sets have decision boundaries at different points.

7 General discussion

In this thesis I sought to investigate how two linguistic tasks are carried out simultaneously, to shed light on the coordination of word planning and comprehension. Despite evidence that in a broad way speaking and listening occur in overlap, there is little evidence testing the second by second coordination of two linguistic tasks when they are forced to occur in parallel. By testing how people coordinate two linguistic tasks in close temporal proximity, we contribute to understanding how speaking and listening happen at the same time.

In Chapter 2 I presented a review of relevant literature and proposed a working model of how word production and comprehension may occur in overlap. The model makes the prediction that lexical concept and lemma selection stages cannot occur in overlap in production and comprehension, but lexical input and output word forms, and input and output phonemes, can be selected in parallel. I proposed that the limitations at these stages are driven by selection processes, not activation processes. This implies that selection processes require capacity. In some cases the amount of capacity required is either too large for another stage to be carried out, or the small amount of capacity allocated to the second task is negligible in the responses measured. A figure displaying this is shown on page 38. Some of the predictions that this model makes were tested in Chapters 3, 4, 5 and 6.

In Chapter 3 I investigated whether dual-tasking with two simple linguistic tasks resulted in mutual interference between the tasks, compared to dual-tasking with one linguistic and one non-linguistic task. In Experiment 1, task 1 was picture naming and task 2 was syllable or tone identification. Three SOAs between tasks were used: 50ms, 300ms and 1800ms. I found that task 1 picture naming RTs were longer when the secondary task was syllable identification compared to tone identification, when the tasks overlapped (SOAs of 50ms and 300ms). Task 2 RTs were also slower for syllable identification than tone identification at all SOAs, despite the tones and syllables being pre-tested to ensure equal difficulty in identifying them both. This pattern of results suggested that concurrent linguistic dual-tasking resulted in mutual interference between the tasks, additive to normal dual-tasking interference. This result also supports one prediction from the model in Chapter 2, namely that two linguistic tasks are processed in parallel. If they were not, we would not expect task 1 naming RTs to be affected by the secondary task.

However, syllables and tones vary in their acoustic complexity, as well as in their linguistic nature. Therefore, in Experiment 2 in Chapter 3 I held the acoustic complexity of the sound constant by presenting the same sounds (sine-wave speech versions of syllables) to all participants. One group of participants were told the sounds were distorted syllables, and another group of participants were told the sounds were computer-generated. I did not find any linguistic interference effects; task 1 naming RTs were the same in both groups, and identification RTs were statistically the same size. This suggested that the linguistic interference found in Experiment 1 may have been driven by the higher acoustic complexity of syllables compared to tones. However, the large amount of variability in the data from Experiment 2 may have masked linguistic effects.

In Chapter 4, I tested again whether linguistic interference between two linguistic tasks would be present if the acoustic complexity of the sounds was held constant. In this experiment, participants carried out picture naming as task 1 and vowel or musical rain identification as task 2. The vowels were synthesised, and the musical rain sounds were created from the synthesised vowels. This resulted in equalised acoustic complexity across the sounds. Participants also carried out a second dual-task condition of size judgement of depicted images as task 1, and vowel or musical rain identification as task 2. Similarly to Experiment 2 in Chapter 3, I found no linguistic interference effects when participants carried out two linguistic tasks (naming and vowel identification) compared to one linguistic and one non-linguistic task (naming and musical rain identification). A Bayesian analysis suggested that this null result was meaningful.

Taken together, the results of Chapters 3 and 4 showed that linguistic interference measured when carrying out two linguistic tasks is not found when the acoustic complexity of the sounds is controlled. This leads to the hypothesis that the acoustic complexity of the sounds drives interference between two linguistic tasks. However, this specific prediction - that acoustic complexity is responsible for this effect - was not tested in this thesis. Future work should investigate this issue.

In Chapter 5 I investigated lexical selection, specifically if this level is affected when carrying out a linguistic task with another linguistic or non-linguistic task. In Experiment 1, participants carried out syllable or tone identification as task

1 (using the same sounds as in Chapter 3), and picture naming as task 2. Pictures were presented with categorically related or unrelated distractor words. I investigated whether semantic interference was found in naming RTs. I found semantic interference of the same size when task 1 was syllable identification as when it was tone identification, and also found the same size semantic interference effect when the two tasks overlapped (SOA 0ms) and when they did not (SOA 1000ms). This suggests that lexical selection is postponed until after the capacity-demanding processes of the identification task are carried out, for both syllable and tone identification as task 1. This provides support for the prediction from the model in Chapter 2, that lexical selection is carried out serially between tasks. If lexical selection were affected by the secondary task, the size of the semantic interference effect would depend on whether the secondary task were linguistic or non-linguistic. This was not the case.

In Experiments 2 and 3 in Chapter 5 a task choice paradigm was used (Besner & Care, 2003). Participants heard a syllable or a tone (as task 1) and saw a picture-word compound (as task 2). Depending on the syllable or tone heard, participants either named the picture or read the word aloud from the picture-word compound. Previous research had found that making a choice on a non-linguistic task could be carried out in parallel with naming processes (Piai et al., 2015). Specifically, if lexical selection could not be carried out in parallel with choice processes, the process of lexical selection would wait until the choice processes were finished. In this case, semantic interference would be measured in naming RTs. However, if lexical selection were carried out in parallel with choice processes, then any conflicts in lexical selection (which give rise to the semantic interference effect) would be absorbed into the time taken for the choice process, and would not be measured in naming RTs. Piai et al. (2015) found, with an SOA of 0ms, that participants could make a choice on a non-linguistic stimulus in parallel with naming processes, as semantic interference was not measured (Piai et al., 2015). I tested whether the same pattern of results was found with a linguistic choice task and picture naming. However, a different pattern of results from this previous work was found. In Experiment 2 I found semantic interference in naming RTs when the task choice stimuli were syllables at both SOAs (0ms and 1000ms), but no semantic interference at either SOA when the task choice stimuli were tones. In Experiment 3, I found semantic interference effects at both SOAs with both syllables and tones as task choice stimuli. Because the results in Experiments 2 and 3 were not consistent within this study, and did not follow the same pattern as in Piai et al. (2015), I analysed the semantic interference effect in all experiments in Chapter 5 using Bayesian t-tests. I tested whether there was enough evidence for the presence or absence of an effect, and found that there was not enough evidence for either conclusion in the task choice experiments. I suggested that this was due to the variability in response strategies that could be employed by participants in this paradigm. I expand on this point below.

In Chapter 6, I moved from a behavioural to an MEG dual-task study. In this experiment, participants carried out the dual linguistic task of syllable identification and picture naming. I used a novel machine learning analysis technique, temporally-generalised multivariate pattern analysis, to decode the pattern of activity related to syllable identification and the pattern of activity related to picture naming. I then tested whether these patterns were found when participants were dual-tasking with the two tasks. For syllable identification (task 1), the pattern of processing decoded was very similar between the single and the dual-task, indicating that the processes involved in syllable identification were not qualitatively affected by dual-tasking. In contrast, picture naming was strongly affected in the dual-task. Evidence was found for both parallel processing of naming and syllable identification, and later reactivation of naming processes. I speculated on the reasons for reactivation. Processes could be reactivated due to incomplete processing when the tasks were carried out in parallel, due to automatic activation of linguistic processes but capacity-demanding selection processes resulting in later reactivation, due to task 1 processes precluding task 2 processes, or due to strategic scheduling of task components in task 2. These results do not fit with the main theories of dual-tasking, as described further below.

Broader implications

The experiments in this thesis demonstrate that in a dual-task of two linguistic tasks there is a complex pattern of processes which occur in parallel and processes which occur serially. The behavioural evidence suggests that certain linguistic processes can occur in parallel, such as aspects of phonological processing, but other processes, such as lexical selection, cannot. The MEG study reveals an interesting pattern of results with two simple linguistic tasks, such that the secondary task (here, picture naming), has processes which run in parallel and which are later reactivated. This is not predicted by theories of dual-tasking or the working model proposed in Chapter 2. Additionally, the behavioural experiments demonstrate that the type of dual-task given to participants affects how they carry out the dual-task. Each of these points is discussed further below.

Overlap in planning and comprehension

Research investigating the overlap in speech planning and comprehension in conversation has largely shown that planning and comprehension of utterances can co-occur (Barthel et al., 2016; Bögels et al., 2015; Bögels et al., 2018; Boiteau et al., 2014; Sjerps & Meyer, 2015), and that at least in some cases, planning occurs in substantial overlap with comprehension. However, whether specific processes in word planning and comprehension occur in parallel, or more serially (potentially with rapid switching), is poorly understood. The work in this thesis sought to address this issue.

The results from all experiments demonstrate that planning and comprehension can occur in overlap, and thus that comprehension (or word planning) is not entirely postponed until after processes in the other task are complete. Thus, the general finding that planning and comprehension can occur concurrently is supported by tasks testing this at a fine-grained level. The results from Chapter 6 specifically show parallel processing between the tasks. Classifiers were trained and tested at samples corresponding to roughly every 3.3ms. Therefore, unless there was very rapid switching between tasks on the order of a few milliseconds or less, this suggests that processes in two linguistic tasks can be carried out in parallel.

However, some processes, such as lexical selection, are not always carried out in parallel (as in Chapter 5). The finding of reactivation in Chapter 6 also shows tasks are not carried out entirely in parallel. This suggests either some flexibility in the concurrent processing of two tasks, such that some processes are scheduled to be serial, or that there is a structural bottleneck for some processes (i.e., that lexical selection cannot be carried out in parallel with another capacity-demanding task). Thus, there is a complex pattern of serial and parallel processing when two linguistic tasks are carried out concurrently.

The results additionally show that two linguistic tasks interfere with one another, resulting in longer responses to the two linguistic tasks, compared to one linguistic and one non-linguistic task. While the precise reasons for this are unclear (see below), this shows that parallel processing does not arise without a cost. In the experiments reported in this thesis, I enforced concurrent processing as much as possible by using dual-tasking paradigms. However, it may be that if participants were free to choose how to schedule the tasks, they may have scheduled processes more serially.

Some of the results in this thesis seem at first glance to be at odds with one another: there is behavioural evidence of serial processing in some dual-tasks, and neuroimaging evidence of parallel processing in another dual-task (though note the later processing reactivation, meaning these results do not support full parallel processing). I propose that these differences arise because of the flexibility of how people are able to schedule two linguistic tasks, which may depend on the types of task and the frequency of how often parallel processing would need to occur (Israel & Cohen, 2011). Further research should test what factors contribute to this flexibility, because undoubtedly there are constraints on how flexible task scheduling can be. For example, in Chapter 2, specific constraints on certain process are proposed, which should limit scheduling flexibility. The results from this thesis also suggest more broadly that it is not the case that planning occurs entirely in parallel with comprehension (as may be proposed based on Levinson & Torreira, 2015), but there is a complex pattern of serial and parallel processes at play in the overlap of these two tasks.

Acoustic complexity within a dual-task

In Chapters 3 and 4, acoustic complexity was held constant in two experiments, which resulted in no measurable linguistic-specific interference between two linguistic tasks compared to one linguistic and one non-linguistic task. This suggests that acoustic complexity may play a role in driving what appears to be linguistic-specific interference in linguistic dual-tasking, found in Experiment 1 in Chapter 3. However, as stated, acoustic complexity was not manipulated in any experiment. Therefore, while these results predict that acoustic complexity may underlie or contribute to linguistic interference effects, this was not tested and thus remains speculative.

As far as I am aware, there are no picture naming studies which have manipulated the acoustic complexity of a secondary sound, or dual-task studies which have manipulated the acoustic complexity of an auditory task. The lack of research in this area means that I am unable to speculate in an informed way on how acoustic complexity might have an effect on picture naming. However, effects of the acoustic complexity of distracting information in serial recall tasks have been investigated. In general, this literature has found that acoustically complex sounds affect serial recall performance more strongly than acoustically simple sounds (e.g., Jones et al., 1992; Jones, 1995). However, the serial recall paradigm differs from picture naming in that 1) participants memorise lists of words, compared to naming one picture, meaning trial length is longer, and 2) there is less precise control over the exact presentation timing of the distracting auditory information and the stage of word encoding. Therefore, it is hard to say whether accounts of acoustic complexity in the serial recall paradigm would make the same predictions for picture naming. Additionally, without testing whether acoustic complexity has an effect in picture naming, it is premature to speculate on any underlying acoustic complexity effects.

Note that I tried to control acoustic complexity as closely as possible, but an entire thesis could be dedicated to how to define acoustic complexity and how to equate sounds for complexity. The two ways used for control here - presenting the exact same sounds with different instructions, or generating two sounds from the same underlying structure - seem sensible ways to maintain the same level of acoustic complexity while manipulating the linguistic nature of the sounds. However, I acknowledge that there may be other ways to control acoustic complexity.

Integration of results with working model

The working model proposed in Chapter 2 made specific predictions, including that word form and phoneme selection could occur in parallel in planning and comprehension, but conceptual and lexical selection could not. The results of Chapters 3 and 4 suggest that acoustic complexity may play a role in dual-tasking. As suggested in Chapter 4, acoustic complexity may play a role because central attention is required for some early perceptual processes. However, it may be that mapping an acoustically complex sound representation to a response is more challenging than mapping a simpler sound to a representation. In both cases it is unclear how acoustic complexity would result in this effect. This could be disentangled in future studies, and any results would lead to refinement of the working model.

The results of Experiment 1 in Chapter 5 suggest that lexical selection is not carried out in parallel with capacity-demanding aspects of a secondary task. This finding is in line with a prediction from the working model. However, further research should investigate whether this is a structural constraint or a strongly preferred scheduling strategy (i.e., whether it is impossible for lexical selection to overlap or whether it can be done but is strongly preferred to be serially scheduled).

How to integrate the results of Chapter 6 with the working model is challenging. The results from this MEG study suggest that the majority of processes in naming (which occurred in single picture naming after 250ms) are reactivated after initial parallel processing in the dual-task (see Figures 6.5 and 6.6). This pattern of results does not fit with the predictions from the working model. These results instead suggest that any linguistic processes occurring in the first 700ms of planning can be carried out in parallel with linguistic processes in comprehension; there is no structural bottleneck. However, linguistic processes are later reactivated, and are carried out 'serially' as the response to task 1 has been given. The working model makes no predictions about reactivation. Further research should investigate the temporal dynamics of dual-tasking two linguistic tasks, especially with different kinds of linguistic stimuli, to determine whether the working model needs to be fundamentally overhauled.

Integration of results with dual-tasking theories

The results of the current thesis have important implications for theories of dual-tasking. Results from Chapters 3, 4 and 6 show that responses to task 1 are longer at short SOAs, suggesting some parallel processing of the two tasks. This is in line with theories of capacity-sharing (Navon & Miller, 2002; Tombu & Jolicœur, 2003) and goes against a main prediction from the response selection bottleneck model (Pashler, 1994). Thus, our results most strongly suggest that two linguistic tasks interfere with one another due to some parallel processing between the tasks.

Much research has shown that tasks are able to affect one another, resulting in cross-talk (Alards-Tomalin et al., 2017; Eder et al., 2017; Hommel, 1998; Janczyk et al., 2014; Janczyk, 2016; Lien et al., 2007; Logan & Schulkind, 2000; Miller, 2006; Röttger & Haider, 2017; Stelzel & Schubert, 2011; Wickens, 2008). Most of these studies tend to test whether congruency in input (e.g., with two stimuli presented on the same side of the screen) or output (e.g., if two responses require the same hand for responding) results in cross-talk. Few studies test whether overlap in central processes results in cross-talk, partially as representations are assumed to be amodal at a central stage. Therefore, the concept of task similarity does not apply (Hazeltine et al., 2006). However, the results from this thesis suggest that similarity in central stages of processing results in cross-talk, because there was no overlap in input between tasks (auditory sounds vs visual images), or responses (button press responses vs spoken responses). Therefore, this thesis demonstrates that central similarity results in cross-talk, even without overlap in input or output modalities.

Neither of the two main dual-task theories can account for the pattern of data from Chapter 6, where I traced the temporal dynamics of two tasks in a dual-task. The data pattern is neither fully serial nor parallel. The response selection bottleneck theory (Pashler, 1994) predicts that initial perceptual processes overlap between tasks, but central stages of processing should be delayed in task 2 until central processing is finished in task 1. However, these results show parallel processing between task 1 and task 2 from 0ms to 750ms. This is unlikely to only be perceptual processing in task 1 and task 2. Firstly, responses are given to the syllable after 792ms, indicating that central processes in syllable identification must occur during the parallel processing time. Secondly, in single naming, naming occurred after 836ms, and it is thus unlikely that no linguistic processes were carried out between 0 and 750ms (it is this time window that generalises to the dual-task). Therefore, the parallel processing likely includes central processing in both tasks. The results also do not conform to predictions from the capacity-sharing theory (Tombu & Jolicœur, 2003; Kahneman, 1973; Navon & Miller, 2002; Wickens, 2008). This theory predicts that central processes run in parallel but are extended in time. However, I did not find evidence for extended processes when processing in parallel. Additionally, neither theory predicts later reactivation of processes. Under the response selection bottleneck theory, late activation would be predicted if there were no parallel processing (and hence is not reactivation). Under the capacity sharing theory, late activation is not predicted at all. Overall, the pattern of dual-task processing with these two simple linguistic tasks cannot be captured by these dual-task theories.

A different dual-task theory assumes no bottlenecks in processing at any stage. The EPIC model proposed by D. E. Meyer and Kieras (1997) models tasks as composed of condition-action rules, of which an unlimited number can run in parallel. However, due to specific task requirements, such as always responding to task 1 before task 2, an executive scheduler may strategically schedule tasks to run in a more serial fashion, where the response to task 2 is either held in working memory, or is not selected until the task 1 response is given (named the EPIC-SRD model). Due to the fact that this model allows tasks to run entirely in parallel, EPIC-SRD can account for the parallel processing found when dual-tasking in the MEG task. The model can also account for reactivation of later stages, assuming that the condition-action rules were scheduled in this way, or that more conditions were needed to be met for an action, resulting in a later rule process. An alternate possibility is that the set of condition-action rules for picture naming included processes being completed twice, once in parallel with task 1 and once after task 1 was completed, though this seems unlikely from a parsimonious account of how tasks would be coordinated. While EPIC-SRD can account for the pattern of data, the question of why tasks are scheduled in this way remains unclear.

In sum, the results from all studies suggest parallel processing of at least some stages between tasks. However, later reactivation of linguistic processes occurs, which may also occur in the behavioural experiments in this thesis. Further research should determine whether this pattern re-occurs with other combinations of linguistic tasks, and why such a pattern of parallel and reactivated processing would occur. With further research we can extend or develop new dual-tasking theories to understand how people dual-task, and specifically for psycholinguistics, how people dual-task with linguistic tasks.

Methodological challenges & implications

In Chapter 5, I found that one of the paradigms used - the task choice paradigm - did not produce stable results. I suggested that participants may vary the strategies they employ when carrying out this dual-task, but I could not determine these strategies because only one output response was measured. This is further complicated by the fact that participants may switch strategies within the experiment. This is consistent with other evidence that participants are able to flexibly switch strategies in a dual-task (Lehle & Hübner, 2009), and different participants can schedule tasks in different ways (Reissland & Manzey, 2016). For example, some participants may prioritise making the task decision before retrieving any linguistic information about the picture or word, and others may carry out the choice task in parallel with retrieving linguistic information. We cannot determine which strategies were used by participants in these experiments. More research should be carried out with the task choice paradigm to determine how participants schedule the two tasks (as discussed in more depth in Chapter 5). However, more research should also be carried out with other dual-task paradigms, as there could be complex interactions between tasks that might not be measured, or might counteract an effect in the response (see also Cook & Meyer, 2008). The work in this thesis provides important insight into the methodological aspects of how to measure dual-tasking, especially when moving to paradigms such as the task choice paradigm which feels more 'natural' and ecologically valid.

In Chapters 3, 4 and 5, participants were presented with two linguistic sounds and two non-linguistic sounds. I pre-tested all sounds to ensure that they were equal in difficulty for participants. This pre-test contained two parts: participants carried out a simple identification task, where they heard one sound on each trial and responded as quickly as possible, and participants also alternated between identification trials and picture naming trials. By including only the identification trials I sought to determine if there was a base RT difference between the different sets of stimuli. The alternating trials allowed me to see if there were RT differences to the sounds with a more complex task-switching design. I reasoned that pre-testing the sounds in this way would be closer to the way that participants would process the sounds in a dual-task. In the pre-tests, there were no significant differences in RTs between linguistic and non-linguistic sounds in the identification trials, and in the alternating trials. Thus, I concluded that any RT differences between linguistic and non-linguistic sounds in the experiments in Chapters 3, 4 and 5 could not be attributed to difficulty.

However, this way of pre-testing the difficulty of two sounds may not have been optimal. Despite pre-tests showing no differences in difficulty, I did find differences in some experiments. In Experiment 1 in Chapter 5 participants were slower to respond to syllables than tones, whereas in Experiment 1 in Chapter 3 there was no difference in RTs using the same stimuli. This suggests that the combination of tasks may interact with difficulty, such that with an easier combined task (syllable identification and picture naming) no difficulty effect is measured, whereas with a harder combination (syllable identification and picture naming with distractors) a difficulty effect arises. While this may be the case, this does not explain why there was a difference in RTs to syllables and tones in the practice session of Experiment 1 in Chapter 5, before participants had carried out any of the harder dual-task. Additionally, I found no difficulty difference between vowel and musical rain identification in the pre-test of the stimuli in Chapter 4, but did find a difference in the two sessions. Altogether, this suggests that measuring difficulty in this way is not optimal, and more research should be conducted to determine the best way of equating difficulty across stimuli.

In Chapter 6, I carried out an MEG study using a fairly new analysis technique, and there are limitations to this technique. The first is that the specific preprocessing stages can affect how well the classifiers can decode. For example, with the data presented in Chapter 6 I found that with different filter settings there was 'ringing' in the data, resulting in harmonic-type patterns over time, which the classifier could decode (Widmann, Schröger, & Maess, 2015). Also, data which were not baseline corrected resulted in above average baseline decoding in some cases. Classifiers are sensitive to small changes in the signal, meaning that it is paramount to carefully inspect the data to ensure effects are due to cognitive changes and not artifacts in the data. There are also a large number of researcher degrees of freedom related to classifier analysis. The technique is not yet established enough for multiple studies to follow the same protocol. For instance, while many studies utilising classification techniques clean their data before analysis (e.g., Charles, King, & Dehaene, 2014; Heikel, Sassenhagen, & Fiebach, 2018; King et al., 2014; Marti et al., 2015) some experts recommend fitting classifiers to raw data (Grootswagers, Wardle, & Carlson, 2017). These analysis decisions can affect the outcome. I believe more research should be carried out using this technique to ensure a discipline standard, with detailed documentation of analysis pipelines.

Limitations in generalising to conversation & future directions

There are of course limitations to this thesis. One is that while I aimed to test the dual-task of speech planning and comprehension, participants were never tested in a dialogue situation. All participants were tested individually, listening to recordings rather than natural spontaneous speech, and speaking into a microphone rather than directed towards a person. This situation is divorced from dialogue, and thus participants may not have responded as they would in a typical conversation. This could affect the generalisability of the results. However, some research has shown that there is a similar pattern of overlap between planning and comprehension, regardless of whether the speaker is live or prerecorded (Sjerps, Decuyper, & Meyer, under review). Though this study tested overlap more broadly, and not at the fine-grained level tested in this thesis, it is unclear why participants would drastically change their speech planning and comprehension strategies in the lab compared to 'in the wild'. Therefore, while testing participants individually is a limitation, I believe that my results are not reflective of only a 'lab mode' for participants.

A second limitation is that I used the PRP paradigm (Telford, 1931), where two responses are given to two separate tasks, which are carried out together. One could argue that while I show how aspects of production and comprehension can be carried out in parallel or serially using the PRP paradigm, participants may allocate capacity differently if separate overt responses are not given to both tasks, as in dialogue (see Navon & Miller, 2002, for criticism of the PRP paradigm). In order to show that the PRP paradigm would not produce results generalisable to dialogue situations, one should suggest aspects of dialogue which cannot be covered by the PRP paradigm. One possible aspect is that during a conversation people are motivated to listen and respond to their partner. This is not necessarily the case in a PRP experiment. A second possible aspect is that in the paradigm two responses to two separate tasks are given, whereas in a typical conversation, people listen to their interlocutor and plan a response contingent on what their interlocutor says without giving a separate response to their interlocutor's speech. Thus, this difference in stimuli-response pairings may affect processing. Note that I attempted to test contingent responding using the the task choice paradigm, but as explained above, the results were inconclusive. A third possible aspect is that during a conversation, people may prioritise speech planning and comprehension to different extents, and possibly in dynamic ways. For example, if it is more important for a person to fully understand an interlocutor, they will prioritise listening and may avoid planning speech in overlap. However, if it is important for a person to quickly respond, they may prioritise planning speech over listening. These priorities may even change over the course of one conversation. In the PRP paradigm, participants are instructed to always respond to task 1 before task 2. Therefore, results from PRP experiments may only generalise out to specific priority orders of planning and listening.

A third limitation is the choice of tasks. While picture naming is undoubtedly a linguistic task, syllable identification likely does not engage the linguistic system to the same extent as other comprehension tasks. While syllables have been shown to activate phonologically-related neighbouring words (e.g., Gaskell & Marslen-Wilson, 2002; Luce & Pisoni, 1998), which must occur through use of the linguistic system, this use of the linguistic system may be minimal. Therefore, the comprehension task tested in this thesis is not representative of comprehension in everyday life. I specifically chose to use a syllable identification task, where monosyllabic non-words served as the syllables, to reduce interactions between different linguistic levels (such as phonological and semantic levels) when dual-tasking. Because complex interactions between tasks are possible, using simple tasks allows for a first investigation into dual-tasking with two linguistic tasks. However, future research should make use of comprehension tasks which engage more of the comprehension system, ensuring results are more generalisable to everyday conversation.

There are many possible features which could be added to the working model presented in Chapter 2. For example, there is currently no role for executive processes in the model. It has been shown that linguistic processes interface with executive functions (see for example Jongman et al., 2015; Shao et al., 2012, 2013; Vromans & Jongman, 2018), and thus the model must be extended to add executive processes. It is also strongly agreed that prediction plays a large role in language comprehension and production (e.g., Pickering & Garrod, 2009, 2013), yet this is not currently modelled. Speech monitoring is also not addressed in the model, despite evidence that speech monitoring occurs at different linguistic levels (W. J. Levelt, 1993) and requires capacity (Oomen & Postma, 2002). Finally, any model of overlap in planning and comprehension should be computational (e.g., van Paridon, Roelofs, & Meyer, 2019) to model how levels interact and how different subprocesses may run serially or in parallel.

The results from this thesis also pose a number of important questions for linguistic dual-tasking that have not yet been addressed, both regarding the linguistic aspects of dual-tasking and processes relating to task coordination. Regarding linguistic variables, questions which follow from the current results include: how do the different levels of planning and comprehension affect one another? Is interference or facilitation different if the speech being planned and comprehended is related, either phonologically or semantically? Questions regarding general task processing coordination include: do different task priorities change how planning and comprehension are coordinated? Is there a basic default priority in the network to prioritise one processing stream over another? How far can people flexibly change their processing strategies (Lehle & Hübner, 2009; Reissland & Manzey, 2016)? How does task uncertainty affect linguistic dual-task processing (Jongman & Meyer, 2017)? Future research can build on the work in this thesis to give a fuller understand of how planning and comprehension processes are carried out in parallel.

General conclusion

In this thesis I addressed whether word planning and comprehension can be carried out in parallel. Taken together, the results suggest that when two linguistic tasks are carried out concurrently, some processes are carried out in parallel and some processes are delayed. There is interference between linguistic tasks which is additive to typical dual-task interference, yet the source of this interference is currently unknown. While it may be due to the similarity in linguistic processes between tasks, it may also be due to the acoustic complexity of the heard linguistic sound. When investigating at a neural level, the processing stream of picture naming (as the second task) was strongly affected in a dual-task, with both parallel and reactivated processing. These results are not predicted by the main dual-task theories, and future research should investigate the temporal dynamics of linguistic dual-tasking more thoroughly. In sum, this thesis contributes to our understanding of how people coordinate word planning and comprehension with close temporal proximity, and raises further testable predictions for how people are able to communicate so quickly and with such ease in dialogue.

References

Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes*, *24*(5), 713–734. doi: 10.1080/01690960802597250

Adank, P., Van Hout, R., & Smits, R. (2004). An acoustic description of the vowels of northern and southern standard dutch. *The Journal of the Acoustical society of America*, *116*(3), 1729–1738. doi: 10.1121/1.1779271

Alards-Tomalin, D., Walker, A. C., Nepon, H., & Leboe-McGowan, L. C. (2017). Dual-task interference effects on cross-modal numerical order and sound intensity judgments: The more the louder? *The Quarterly Journal of Experimental Psychology*, *70*(9), 1943–1963. doi: 10.1080/17470218.2016.1216139

Allen, P. A., Ruthruff, E., Elicker, J. D., & Lien, M.-C. (2009). Multisession, dual-task psychological refractory period practice benefits older and younger adults equally. *Experimental Aging Research*, *35*(4), 369–399. doi: 10.1080/03610730903175766

Arrington, C. M., Altmann, E. M., & Carr, T. H. (2003). Tasks of a feather flock together: Similarity effects in task switching. *Memory & Cognition*, *31*(5), 781–789. doi: 10.3758/BF03196116

Audacity(R), A. (2014). *Audacity (version 2.0.6).*

Ayora, P., Janssen, N., Dell'Acqua, R., & Alario, F.-X. (2009). Attentional requirements for the selection of words from different grammatical categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1344–1351. doi: 10.1037/a0016373

Ayora, P., Peressotti, F., Alario, F.-X., Mulatti, C., Pluchino, P., Job, R., & Dell'Acqua, R. (2011). What phonological facilitation tells about semantic interference: A dual-task study. *Frontiers in Psychology*, *2*. doi: 10.3389/fpsyg.2011.00057

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi: 10.1016/j.jml.2012.11.001

Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. *Frontiers in Psychology*, *7*, 1858. doi: 10.3389/fpsyg.2016.01858

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Belke, E. (2013). Long-lasting inhibitory semantic context effects on object naming are necessarily conceptually mediated: Implications for models of lexical-semantic encoding. *Journal of Memory and Language*, *69*(3), 228–256. doi: 10.1016/j.jml.2013.05.008

Besner, D., & Care, S. (2003). A paradigm for exploring what the mind does while deciding what it should do. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *57*(4), 311–320. doi: 10.1037/h0087434

Besner, D., & Risko, E. F. (2005). Stimulus-response compatible orienting and the effect of an action not taken: Perception delayed is automaticity denied. *Psychonomic Bulletin & Review*, *12*(2), 271–275. doi: 10.3758/BF03196371

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, *1334*, 1–47. doi: 10.1075/lllt.17.07bes

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot international*, *5*.

Bögels, S., Casillas, M., & Levinson, S. C. (2018). Planning versus comprehension in turn-taking: Fast responders show reduced anticipatory processing of the question. *Neuropsychologia*, *109*, 295–310. doi: 10.1016/j.neuropsychologia.2017.12.028

Boiteau, T. W., Malone, P. S., Peters, S. A., & Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *Journal of Experimental Psychology: General*, *143*(1), 295–311. doi: 10.1037/a0031858

Brisson, B., & Jolicœur, P. (2007). Electrophysiological evidence of central interference in the control of visuospatial attention. *Psychonomic Bulletin & Review*, *14*(1), 126–132. doi: 10.3758/bf03194039

Brown, T. L., Joneleit, K., Robinson, C. S., & Brown, C. R. (2002). Automaticity in reading and the stroop task: Testing the limits of involuntary word processing. *The American Journal of Psychology*, *115*(4). doi: 10.2307/1423526

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, *5*(1). doi: 10.1038/srep12881

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: the first 1000 ms. *Journal of Vision*, *13*(1). doi: 10.1167/13.10.1

Carlson, T. A., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal resolution decoding of object position and category. *Journal of Vision*, *11*(9). doi: 10.1167/11.10.9

Chang, C.-C., & Lin, C.-J. (2011). Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *2*(3), 27.

Charles, L., King, J.-R., & Dehaene, S. (2014). Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli. *Journal of Neuroscience*, *34*(4), 1158–1170. doi: 10.1523/JNEUROSCI.2465-13.2014

Chipunza, C., & Mandeya, A. (2005). Dual-task processing: Effects of task difficulty and stimulus similarity on dual-task performance. *South African Journal of Psychology*, *35*(4), 684–702. doi: 10.1177/008124630503500405

Cleland, A. A., Gaskell, M. G., Quinlan, P. T., & Tamminen, J. (2006). Frequency effects in spoken and visual word recognition: Evidence from dual-task methodologies. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 104–119. doi: 10.1037/0096-1523.32.1.104

Cleland, A. A., Tamminen, J., Quinlan, P. T., & Gaskell, M. G. (2012). Spoken word processing creates a lexical bottleneck. *Language and Cognitive Processes*, *27*(4), 572–593. doi: 10.1080/01690965.2011.564942

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*(4), 886–899. doi: 10.1037/0278-7393.34.4.886

Cutting, J. C. (1997). *The production and comprehension lexicons: What is shared and what is not* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 345–361. doi: 10.1037/0278-7393.25.2.345

Darwin, C. (2005). *Praat scripts.* Retrieved from `http://www.lifesci.sussex.ac.uk/home/ChrisDarwin/Praatscripts/` (Accessed: 31-08-2018)

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage*, *24*(1), 21–33. doi: 10.1016/j.neuroimage.2004.09.039

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. doi: 10.1037/0033-295X.93.3.283

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*(1-3), 287–314. doi: 10.1016/0010-0277(92)90046-K

Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture-word interference effect is not a Stroop effect. *Psychonomic Bulletin & Review*, *14*(4), 717–722. doi: 10.3758/BF03196827

Dent, K., Johnston, R. A., & Humphreys, G. W. (2008). Age of acquisition and word frequency effects in picture naming: A dual-task investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 282–301. doi: 10.1037/0278-7393.34.2.282

Divin, W., Coyle, K., & James, D. (2001). The effects of irrelevant speech and articulatory suppression on the serial recall of silently presented lipread digits. *British Journal of Psychology*, *92*(4), 593–616. doi: 10.1348/000712601162365

Eder, A. B., Pfister, R., Dignath, D., & Hommel, B. (2017). Anticipatory affect during action preparation: Evidence from backward compatibility in dual-task performance. *Cognition and Emotion*, *31*(6), 1211–1224. doi: 10.1080/02699931.2016.1208151

Egorova, N., Shtyrov, Y., & Pulvermuller, F. (2013). Early and parallel processing of pragmatic and semantic information in speech acts: Neurophysiological evidence. *Frontiers in Human Neuroscience*, *7*, 86. doi: 10.3389/fnhum.2013.00086

Elliott, E. M. (2002). The irrelevant-speech effect and children: Theoretical implications of developmental change. *Memory & Cognition*, *30*(3), 478–487. doi: 10.3758/bf03194948

Fairs, A., Bögels, S., & Meyer, A. S. (in preparation). Interference in dual-tasking simple linguistic tasks may be due to linguistic interference or acoustic complexity.

Fargier, R., & Laganaro, M. (2016). Neurophysiological modulations of non-verbal and verbal dual-tasks interference during word planning. *PLoS ONE*, *11*(12), e0168358. doi: 10.1371/journal.pone.0168358

Feng, S. F., Schwemmer, M., Gershman, S. J., & Cohen, J. D. (2014). Multitasking versus multiplexing: Toward a normative account of limitations in the simultaneous execution of control-demanding behaviors. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(1), 129–146. doi: 10.3758/s13415-013-0236-9

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current directions in Psychological Science*, *11*(1), 11–15. doi: 10.1111/1467-8721.00158

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*(6), 1187–1199. doi: 10.1037/0278-7393.28.6.1187

Fischer, R., & Plessow, F. (2015). Efficient multitasking: Parallel versus serial processing of multiple tasks. *Frontiers in Psychology*, *6*. doi: 10.3389/fpsyg.2015.01366

Gabriel, D., Gaudrain, E., Lebrun-Guillaud, G., Sheppard, F., Tomescu, I. M., & Schnider, A. (2012). Do irrelevant sounds impair the maintenance of all characteristics of speech in memory? *Journal of Psycholinguistic Research*, *41*(6), 475–486. doi: 10.1007/s10936-012-9204-8

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, *12*(5-6), 613–656. doi: 10.1080/016909697386646

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, *45*(2), 220–266. doi: 10.1016/S0010-0285(02)00003-8

Gaskell, M. G., Quinlan, P. T., Tamminen, J., & Cleland, A. A. (2008). The nature of phoneme representation in spoken word recognition. *Journal of Experimental Psychology: General*, *137*(2), 282–302. doi: 10.1037/0096-3445.137.2.282

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 640–654. doi: 10.1037/0096-1523.10.5.640

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., … Hämäläinen, M. S. (2014). Mne software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. doi: 10.1016/j.neuroimage.2013.10.027

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. doi: 10.1162/jocna01068

Halin, N., Marsh, J. E., & Sörqvist, P. (2015). Central load reduces peripheral processing: evidence from incidental memory of background speech. *Scandinavian Journal of Psychology*, *56*(6), 607–612. doi: 10.1111/sjop.12246

Han, S. W. (2017). The involvement of central attention in visual search is determined by task demands. *Attention, Perception, & Psychophysics*, *79*(3), 726–737. doi: 10.3758/s13414-017-1278-9

Hazeltine, E., Ruthruff, E., & Remington, R. (2006). The role of input and output modality pairings in dual-task performance: Evidence for content-dependent central interference. *Cognitive Psychology*, *52*(4), 291–345. doi: 10.1016/j.cogpsych.2005.11.001

Heikel, E., Sassenhagen, J., & Fiebach, C. J. (2018). Time-generalized multivariate analysis of EEG responses reveals a cascading architecture of semantic mismatch processing. *Brain and Language*, *184*, 43–53. doi: 10.1016/j.bandl.2018.06.007

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555-568. doi: 10.1016/j.wocn.2010.08.002

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1-2), 67–99. doi: 10.1016/j.cognition.2003.10.011

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. doi: 10.1038/nrn2113

Hohlfeld, A., Sangals, J., & Sommer, W. (2004). Effects of additional tasks on language perception: An event-related brain potential investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1012–1025. doi: 10.1037/0278-7393.30.5.1012

Hohlfeld, A., & Sommer, W. (2005). Semantic processing of unattended meaning is modulated by additional task load: Evidence from electrophysiology. *Cognitive Brain Research*, *24*(3), 500–512. doi: 10.1016/j.cogbrainres.2005.03.001

Hommel, B. (1998). Automatic stimulus–response translation in dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(5), 1368–1384. doi: 10.1037/0096-1523.24.5.1368

Hopman, E. W., & MacDonald, M. C. (2018). Production practice during language learning improves comprehension. *Psychological Science*, *29*(6), 961–971. doi: 10.1177/0956797618754486

Humphreys, G. F., Mirković, J., & Gennari, S. P. (2016). Similarity-based competition in relative clause production and comprehension. *Journal of Memory and Language*, *89*, 200–221. doi: 10.1016/j.jml.2015.12.007

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2*. doi: 10.3389/fpsyg.2011.00255

Indefrey, P., & Levelt, W. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101–144. doi: 10.1016/j.cognition.2002.06.001

Israel, M., & Cohen, A. (2011). Involuntary strategy-dependent dual task performance. *Psychological Research*, *75*(6), 513–524. doi: 10.1007/s00426-011-0359-y

Janczyk, M. (2016). Sequential modulation of backward crosstalk and task-shielding in dual-tasking. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(5), 631–647. doi: 10.1037/xhp0000170

Janczyk, M., Pfister, R., Hommel, B., & Kunde, W. (2014). Who is talking in backward crosstalk? Disentangling response- from goal-conflict in dual-task performance. *Cognition*, *132*(1), 30–43. doi: 10.1016/j.cognition.2014.03.001

Janssen, N., Schirm, W., Mahon, B. Z., & Caramazza, A. (2008). Semantic interference in a delayed naming task: Evidence for the response exclusion hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 249–256. doi: 10.1037/0278-7393.34.1.249

Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824–843. doi: 10.3758/s13415-013-0236-9

Jescheniak, J. D., Schriefers, H., Garrett, M. F., & Friederici, A. D. (2002). Exploring the activation of semantic and phonological codes during speech planning with event-related brain potentials. *Journal of Cognitive Neuroscience*, *14*(6), 951–964. doi: 10.1162/089892902760191162

Jones, D. (1995). The fate of the unattended stimulus: Irrelevant speech and cognition. *Applied Cognitive Psychology*, *9*(7), S23–S38. doi: 10.1002/acp.2350090704

Jones, D., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *The Quarterly Journal of Experimental Psychology Section A*, *44*(4), 645–669. doi: 10.1080/14640749208401304

Jongman, S. R., & Meyer, A. S. (2017). To plan or not to plan: Does planning for production remove facilitation from associative priming? *Acta Psychologica*, *181*, 40–50. doi: 10.1016/j.actpsy.2017.10.003

Jongman, S. R., Roelofs, A., & Meyer, A. S. (2015). Sustained attention in language production: An individual differences investigation. *Quarterly Journal of Experimental Psychology*, *68*(4), 710–730. doi: 10.1080/17470218.2014.964736

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.

Kamienkowski, J. E., Pashler, H., Dehaene, S., & Sigman, M. (2011). Effects of practice on task architecture: Combined evidence from interference experiments and random-walk models of decision making. *Cognition*, *119*(1), 81–95. doi: 10.1016/j.cognition.2010.12.010

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *The Quarterly Journal of Experimental Psychology*, *69*(5), 1013–1040. doi: 10.1080/17470218.2015.1053951

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. doi: 10.2307/2291091

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. doi: 10.1016/j.tics.2014.01.002

King, J.-R., Gramfort, A., Schurger, A., Naccache, L., & Dehaene, S. (2014). Two distinct dynamic modes subtend the detection of unexpected sounds. *PloS one*, *9*(1), e85791. doi: 10.1371/journal.pone.0085791

Kittredge, A. K., & Dell, G. S. (2016). Learning to speak by listening: Transfer of phonotactics from perception to production. *Journal of Memory and Language*, *89*, 8–22. doi: 10.1016/j.jml.2015.08.001

Klatte, M., Lachmann, T., Schlittmeier, S., & Hellbrück, J. (2010). The irrelevant sound effect in short-term memory: Is there developmental change? *European Journal of Cognitive Psychology*, *22*(8), 1168–1191. doi: 10.1080/09541440903378250

Klaus, J., Mädebach, A., Oppermann, F., & Jescheniak, J. D. (2017). Planning sentences while doing other things at the same time: Effects of concurrent verbal and visuospatial working memory load. *The Quarterly Journal of Experimental Psychology*, *70*(4), 811–831. doi: 10.1080/17470218.2016.1167926

Kleinman, D. (2013). Resolving semantic interference during word production requires central attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1860–1877. doi: 10.1037/a0033095

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, *15*(4), 722–752. doi: 10.1177/1094428112457829

Kruschke, J. K., & Meredith, M. (2018). Best: Bayesian estimation supersedes the t-test [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=BEST` (R package version 0.5.1)

Lehle, C., & Hübner, R. (2009). Strategic capacity sharing between two tasks: Evidence from tasks with the same and with different task sets. *Psychological Research PRPF*, *73*(5), 707. doi: 10.1007/s00426-008-0162-6

Lenth, R. V. (2016). Least-Squares Means: The *R* Package **lsmeans**. *Journal of Statistical Software*, *69*(1). doi: 10.18637/jss.v069.i01

Levelt, W. J. (1993). *Speaking: From intention to articulation* (Vol. 1). MIT press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01). doi: 10.1017/S0140525X99001776

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in Cognitive Sciences*, *20*(1), 6–14. doi: 10.1016/j.tics.2015.10.010

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*. doi: 10.3389/fpsyg.2015.00731

Liebenthal, E., Binder, J. R., Piorkowski, R. L., & Remez, R. E. (2003). Short-term reorganization of auditory analysis induced by phonetic experience. *Journal of Cognitive Neuroscience*, *15*(4), 549–558. doi: 10.1162/089892903321662930

Lien, M.-C., & Proctor, R. W. (2002). Stimulus-response compatibility and psychological refractory period effects: Implications for response selection. *Psychonomic Bulletin & Review*, *9*(2), 212–238. doi: 10.3758/BF03196277

Lien, M.-C., Ruthruff, E., Cornett, L., Goodin, Z., & Allen, P. A. (2008). On the nonautomaticity of visual word processing: Electrophysiological evidence that word processing requires central attention. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 751–773. doi: 10.1037/0096-1523.34.3.751

Lien, M.-C., Ruthruff, E., Hsieh, S., & Yu, Y.-T. (2007). Parallel central processing between tasks: Evidence from lateralized readiness potentials. *Psycho-*

nomic Bulletin & Review, *14*(1), 133–141. doi: 10.3758/BF03194040

Liepelt, R., Strobach, T., Frensch, P., & Schubert, T. (2011). Improved intertask coordination after extensive dual-task practice. *The Quarterly Journal of Experimental Psychology*, *64*(7), 1251–1272. doi: 10.1080/17470218.2010.543284

Little, J. S., Martin, F. H., & Thomson, R. H. (2010). Speech versus non-speech as irrelevant sound: Controlling acoustic variation. *Biological Psychology*, *85*(1), 62–70. doi: 10.1016/j.biopsycho.2010.05.004

Logan, G. D., Miller, A. E., & Strayer, D. L. (2011). Electrophysiological evidence for parallel response selection in skilled typists. *Psychological Science*, *22*(1), 54–56. doi: 10.1177/0956797610390382

Logan, G. D., & Schulkind, M. D. (2000). Parallel memory retrieval in dual-task situations: I. Semantic memory. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(3), 1072–1090. doi: 10.1037/0096-1523.26.3.1072

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model:. *Ear and Hearing*, *19*(1), 1–36. doi: 10.1097/00003446-199802000-00001

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, *4*(11), 432–440. doi: 10.1016/s1364-6613(00)01545-x

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00226

MacGregor, L. J., Pulvermüller, F., Van Casteren, M., & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communications*, *3*, 711. doi: 10.1038/ncomms1715

Mama, Y., Fostick, L., & Icht, M. (2018). The impact of different background noises on the production effect. *Acta Psychologica*, *185*, 235–242. doi: 10.1016/j.actpsy.2018.03.002

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. doi: 10.1016/j.jneumeth.2007.03.024

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71–102. doi: 10.1016/0010-0277(87)90005-9

Marti, S., King, J.-R., & Dehaene, S. (2015). Time-resolved decoding of two processing chains during dual-task interference. *Neuron*, *88*(6), 1297–1307. doi:

10.1016/j.neuron.2015.10.040
Marti, S., Sigman, M., & Dehaene, S. (2012). A shared cortical bottleneck underlying attentional blink and psychological refractory period. *NeuroImage*, *59*(3), 2883–2898. doi: 10.1016/j.neuroimage.2011.09.063
Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is production: The missing link between language production and comprehension. *Scientific Reports*, *8*(1079). doi: 10.1038/s41598-018-19499-4
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. doi: 10.1016/0010-0285(86)90015-0
McQueen, J. M., Eisner, F., Burgering, M., & Vroomen, J. (in press). Specialized memory systems for learning spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition*.
Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional mri. *Psychological Science*, *22*(9), 1173–1182. doi: 10.1177/0956797611418347
Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(6), 1146–1160. doi: 10.1037/0278-7393.17.6.1146
Meyer, A. S., & Van der Meulen, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin & Review*, *7*(2), 314–319. doi: 10.3758/bf03212987
Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part I. Basic mechanisms. *Psychological Review*, *104*(1), 3–65. doi: 10.1037/0033-295X.104.1.3
Miller, J. (2006). Backward crosstalk effects in psychological refractory period paradigms: Effects of second-task response types on first-task response latencies. *Psychological Research*, *70*(6), 484. doi: 10.1007/s00426-005-0011-9
Miozzo, M., Pulvermüller, F., & Hauk, O. (2015). Early parallel activation of semantics and phonology in picture naming: Evidence from a multiple linear regression meg study. *Cerebral Cortex*, *25*(10), 3343–3355. doi: 10.1093/cercor/bhu137
Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168–173. doi: 10.1016/j.cognition.2008.08.002

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. doi: 10.1006/cogp.1999.0734

Moisala, M., Salmela, V., Salo, E., Carlson, S., Vuontela, V., Salonen, O., & Alho, K. (2015). Brain activity during divided and selective attention to auditory and visual sentence comprehension tasks. *Frontiers in Human Neuroscience*, *19*. doi: 10.3389/fnhum.2015.00086

Morey, R. D., & Rouder, J. N. (2018). Bayesfactor: Computation of bayes factors for common designs [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=BayesFactor` (R package version 0.9.12-4.2)

Mädebach, A., Jescheniak, J. D., Oppermann, F., & Schriefers, H. (2011). Ease of processing constrains the activation flow in the conceptual-lexical system during speech planning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 649–660. doi: 10.1037/a0022330

Mädebach, A., Oppermann, F., Hantsch, A., Curda, C., & Jescheniak, J. D. (2011). Is there semantic interference in delayed naming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 522–538. doi: 10.1037/a0021970

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., & Sams, M. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *NeuroImage*, *30*(2), 563–569. doi: 10.1016/j.neuroimage.2005.10.002

Navon, D., & Miller, J. (1987). Role of outcome conflict in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 435–448. doi: 10.1037/0096-1523.13.3.435

Navon, D., & Miller, J. (2002). Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology*, *44*(3), 193–251. doi: 10.1006/cogp.2001.0767

Nyberg, L., Petersson, K. M., Nilsson, L.-G., Sandblom, J., Åberg, C., & Ingvar, M. (2001). Reactivation of motor brain areas during explicit memory for actions. *NeuroImage*, *14*, 521–528. doi: 10.1006/nimg.2001.0801

Oomen, C. C., & Postma, A. (2002). Limitations in processing resources and speech monitoring. *Language and Cognitive Processes*, *17*(2), 163–184. doi: 10.1080/01690960143000010

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1. doi: 10.1155/2011/156869

O'Malley, S., & Besner, D. (2011). Lexical processing while deciding what task to perform: Reading aloud in the context of the task set paradigm. *Consciousness and Cognition*, *20*(4), 1594–1603. doi: 10.1016/j.concog.2011.08.006

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 358–377. doi: 10.1037/0096-1523.10.3.358

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*(2), 220–244. doi: 10.1037/0033-2909.116.2.220

Pashler, H. (1998). *The psychology of attention*. Cambridge, Mass: MIT Press.

Paucke, M., Oppermann, F., Koch, I., & Jescheniak, J. D. (2015). On the costs of parallel processing in dual-task performance: The case of lexical processing in word production. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1539–1552. doi: 10.1037/a0039583

Paulitzki, J. R., Risko, E. F., O'Malley, S., Stolz, J. A., & Besner, D. (2009). On the role of set when reading aloud: A dissociation between prelexical and lexical processing. *Consciousness and Cognition*, *18*(1), 135–144. doi: 10.1016/j.concog.2008.11.007

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Penolazzi, B., Hauk, O., & Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology*, *74*(3), 374–388. doi: 10.1016/j.biopsycho.2006.09.008

Piai, V., Roelofs, A., & Schriefers, H. (2011). Semantic interference in immediate and delayed naming and reading: Attention and task decisions. *Journal of Memory and Language*, *64*(4), 404–423. doi: 10.1016/j.jml.2011.01.004

Piai, V., Roelofs, A., & Schriefers, H. (2014). Locus of semantic interference in picture naming: Evidence from dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 147–165. doi: 10.1037/a0033745

Piai, V., Roelofs, A., & Schriefers, H. (2015). Task choice and semantic interference in picture naming. *Acta Psychologica*, *157*, 13–22. doi: 10.1016/j.actpsy.2015.02.003

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2). doi: 10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2009). Prediction and embodiment in dialogue. *European Journal of Social Psychology*, *39*(7), 1162–1168. doi: 10.1002/ejsp.663

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. doi: 10.1017/S0140525X12001495

Polich, J., & Donchin, E. (1988). P300 and the word frequency effect. *Clinical Neurophysiology*, *70*(1), 33–45. doi: 10.1016/0013-4694(88)90192-7

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*(2), 253–279. doi: 10.1017/s0140525x9900182x

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*, *17*(6), 884–892. doi: 10.1162/0898929054021111

Pulvermüller, F., Shtyrov, Y., Ilmoniemi, R. J., & Marslen-Wilson, W. D. (2006). Tracking speech comprehension in space and time. *NeuroImage*, *31*(3), 1297–1305. doi: 10.1016/j.neuroimage.2006.01.030

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Reissland, J., & Manzey, D. (2016). Serial or overlapping processing in multitasking as individual preference: Effects of stimulus preview on task switching and concurrent dual-task performance. *Acta Psychologica*, *168*, 27–40. doi: 10.1016/j.actpsy.2016.04.010

Relander, K., Rämä, P., & Kujala, T. (2009). Word semantics is processed even without attentional effort. *Journal of Cognitive Neuroscience*, *21*(8), 1511–1522. doi: 10.1162/jocn.2009.21127

Remez, R. E., Pardo, J. S., Piorkowski, R. L., & Rubin, P. E. (2001). On the bistability of sine wave analogues of speech. *Psychological Science*, *12*(1), 24–29. doi: 10.1111/1467-9280.00305

Remez, R. E., Rubin, P., Pisoni, D., & Carrell, T. (1981). Speech perception without traditional speech cues. *Science*, *212*(4497), 947–949. doi: 10.1126/science.7233191

Reynolds, M., & Besner, D. (2006). Reading aloud is not automatic: Processing capacity is required to generate a phonological code from print. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1303. doi: 10.1037/0096-1523.32.6.1303

Risko, E. F., & Besner, D. (2008). A role for set in the control of automatic spatial response activation. *Experimental Psychology*, *55*(1), 38–46. doi: 10.1027/1618-3169.55.1.38

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1-3), 107–142. doi: 10.1016/0010-0277(92)90041-F

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*(3), 249-284. doi: 10.1016/s0010-0277(97)00027-9

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the stroop task. *Psychological Review*, *110*(1), 88. doi: 10.1037/0033-295x.110.1.88

Roelofs, A. (2008a). Attention, gaze shifting, and dual-task interference from phonological encoding in spoken word planning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1580–1598. doi: 10.1037/a0012476

Roelofs, A. (2008b). Tracing attention and the activation flow in spoken word planning using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 353–368. doi: 10.1037/0278-7393.34.2.353

Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model. *Cortex*, *59*, 33–48. doi: 10.1016/j.cortex.2014.07.001

Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology*, *2*. doi: 10.3389/fpsyg.2011.00307

Roelofs, A., & Piai, V. (2017). Distributional analysis of semantic interference in picture naming. *Quarterly Journal of Experimental Psychology*, *70*(4), 782–792. doi: 10.1080/17470218.2016.1165264

Röttger, E., & Haider, H. (2017). Investigating the characteristics of "not responding": Backward crosstalk in the PRP paradigm with forced vs. free no-go decisions. *Psychological research*, *81*(3), 596–610. doi: 10.1007/s00426-016-0772-3

Sasisekaran, J., & Donohue, C. (2016). Effects of concurrent task demands on language planning in fluent children and adults. *Applied Psycholinguistics*, *37*(06), 1311–1336. doi: 10.1017/S0142716415000582

Schnur, T. T., & Martin, R. (2012). Semantic picture–word interference is a post-perceptual effect. *Psychonomic Bulletin & Review*, *19*(2), 301–308. doi: 10.3758/s13423-011-0190-x

Schriefers, H., Meyer, A., & Levelt, W. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*(1), 86–102. doi: 10.1016/0749-596X(90)90011-N

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(12), 2400–2406. doi: 10.1093/brain/123.12.2400

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action — candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, *10*(4), 295–302. doi: 10.1038/nrn2603

Sergent, C., Wyart, V., Babo-Rebelo, M., Cohen, L., Naccache, L., & Tallon-Baudry, C. (2013). Cueing attention after the stimulus is gone can retrospectively trigger conscious perception. *Current Biology*, *23*(2), 150–155. doi: 10.1016/j.cub.2012.11.047

Shallice, T., McLeod, P., & Lewis, K. (1985). Isolating cognitive modules with the dual-task paradigm: Are speech perception and production separate processes? *The Quarterly Journal of Experimental Psychology Section A*, *37*(4), 507–532. doi: 10.1080/14640748508400917

Shao, Z., Meyer, A. S., & Roelofs, A. (2013). Selective and nonselective inhibition of competitors in picture naming. *Memory & Cognition*, *41*(8), 1200–1211. doi: 10.3758/s13421-013-0332-7

Shao, Z., Roelofs, A., & Meyer, A. S. (2012). Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *The Quarterly Journal of Experimental Psychology*, *65*(10), 1927–1944. doi: 10.1080/17470218.2012.670252

Sigman, M., & Dehaene, S. (2005). Parsing a cognitive task: A characterization of the mind's bottleneck. *PLoS Biology*, *3*(2), e37. doi: 10.1371/journal.pbio.0030037

Sigman, M., & Dehaene, S. (2006). Dynamics of the central bottleneck: Dual-task and task uncertainty. *PLoS Biology*, *4*(7), e220. doi: 10.1371/journal.pbio.0040220

Sigman, M., & Dehaene, S. (2008). Brain mechanisms of serial and parallel processing during dual-task performance. *Journal of Neuroscience*, *28*(30), 7585–7598. doi: 10.1523/JNEUROSCI.0948-08.2008

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, *111*(43), E4687–E4696. doi: 10.1073/pnas.1323812111

Sjerps, M. J., Decuyper, C., & Meyer, A. S. (under review). Initiation of utterance planning in response to pre-recorded and 'live' utterances.

Sjerps, M. J., & Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition*, *136*, 304–324. doi: 10.1016/j.cognition.2014.10.008

Stelzel, C., & Schubert, T. (2011). Interference effects of stimulus–response modality pairings in dual tasks and their robustness. *Psychological Research*, *75*(6), 476–490. doi: 10.1007/s00426-011-0368-x

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592. doi: 10.1073/pnas.0903616106

Stojanoski, B., & Cusack, R. (2014). Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. *Journal of Vision*, *14*(12), 6. doi: 10.1167/14.12.6

Stolk, A., Todorovic, A., Schoffelen, J.-M., & Oostenveld, R. (2013). Online and offline tools for head movement compensation in MEG. *NeuroImage*, *68*, 39–48. doi: 10.1016/j.neuroimage.2012.11.047

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological science*, *12*(6), 462–466. doi: 10.1111/1467-9280.00386

Strijkers, K., Costa, A., & Pulvermüller, F. (2017). The cortical dynamics of speaking: Lexical and phonological knowledge simultaneously recruit the frontal and temporal cortex within 200 ms. *NeuroImage*, *163*, 206–219. doi: 10.1016/j.neuroimage.2017.09.041

Strobach, T., Frensch, P., Müller, H., & Schubert, T. (2012). Age- and practice-related influences on dual-task costs and compensation mechanisms under optimal conditions of dual-task performance. *Aging, Neuropsychology, and Cognition*, *19*(1-2), 222–247. doi: 10.1080/13825585.2011.630973

Strobach, T., Frensch, P. A., & Schubert, T. (2012). Video game practice optimizes executive control skills in dual-task and task switching situations. *Acta psychologica*, *140*(1), 13–24. doi: 10.1016/j.actpsy.2012.02.001

Strobach, T., Liepelt, R., Pashler, H., Frensch, P. A., & Schubert, T. (2013). Effects of extensive dual-task practice on processing stages in simultaneous choice tasks. *Attention, Perception, & Psychophysics*, *75*(5), 900–920. doi: 10.3758/s13414-013-0451-z

Strobach, T., Salminen, T., Karbach, J., & Schubert, T. (2014). Practice-related optimization and transfer of executive functions: A general review and a specific realization of their mechanisms in dual tasks. *Psychological Research*, *78*(6), 836–851. doi: 10.1007/s00426-014-0563-7

Telford, C. W. (1931). The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology*, *14*(1), 1–36. doi: 10.1037/h0073262

Tombu, M., & Jolicœur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 3–18. doi: 10.1037/0096-1523.29.1.3

Tremblay, S., Nicholls, A. P., Alford, D., & Jones, D. M. (2000). The irrelevant sound effect: Does speech play a special role? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1750–1754. doi: 10.1037/0278-7393.26.6.1750

Uppenkamp, S., Johnsrude, I. S., Norris, D., Marslen-Wilson, W., & Patterson, R. D. (2006). Locating the initial stages of speech–sound processing in human temporal cortex. *Neuroimage*, *31*(3), 1284–1296. doi: 10.1016/j.neuroimage.2006.01.004

Vachon, F., & Jolicœur, P. (2012). On the automaticity of semantic processing during task switching. *Journal of Cognitive Neuroscience*, *24*(3), 611–626. doi: 10.1162/jocna00149

Van Assche, E., Duyck, W., & Gollan, T. H. (2016). Linking recognition and production: Cross-modal transfer effects between picture naming and lexical decision during first and second language processing in bilinguals. *Journal of Memory and Language*, *89*, 37–54. doi: 10.1016/j.jml.2016.02.003

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*(4), 584–589. doi: 10.3758/BF03193889

Van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The l2lp model revised. *Frontiers in Psychology*, *6*, 1000. doi: 10.3389/fpsyg.2015.01000

van Paridon, J., Roelofs, A., & Meyer, A. S. (2019). A lexical bottleneck in shadowing and translating of narratives. *Language, Cognition and Neuroscience*. doi: 10.1080/23273798.2019.1591470

Van Selst, M., Ruthruff, E., & Johnston, J. C. (1999). Can practice eliminate the psychological refractory period effect? *Journal of Experimental Psychology: Human Perception and Performance*, *25*(5), 1268–1283. doi: 10.1037/0096-1523.25.5.1268

Vidaurre, D., Myers, N., Stokes, M., Nobre, A. C., & Woolrich, M. W. (2018). Temporally unconstrained decoding reveals consistent but time-varying stages of stimulus processing. *bioRxiv*(260943). doi: 10.1101/260943

Vihman, M. M. (2013). *Phonological development: The first two years*. John Wiley & Sons.

Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, *108*(1), 1–27. doi: 10.1111/bjop.12207

Vromans, R. D., & Jongman, S. R. (2018). The interplay between selective and nonselective inhibition during single word production. *PloS one*, *13*(5), e0197313. doi: 10.1371/journal.pone.0197313

Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 423–440. doi: 10.1037/a0018619

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. doi: 10.1016/s0163-6383(02)00093-0

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences*, *97*(20), 11125–11129. doi: 10.1073/pnas.97.20.11125

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455. doi: 10.1518/001872008X288394

Wickens, C. D., Vidulich, M., & Sandry-Garza, D. (1984). Principles of s-c-r compatibility with spatial and verbal tasks: The role of display-control location and voice-interactive display-control interfacing. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *26*(5), 533–543. doi: 10.1177/001872088402600505

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data–a practical approach. *Journal of Neuroscience Methods*, *250*, 34–46. doi: 10.1016/j.jneumeth.2014.08.002

Nederlandse samenvatting

We besteden veel tijd aan praten en luisteren, maar hoe vaak we dit ook doen en hoe makkelijk het ook lijkt, het gemak waarmee we gesprekken kunnen voeren is behoorlijk verrassend. Mensen zijn heel snel in het opvangen van geluiden die uit iemands mond komen, het herkennen van woorden in deze stroom van geluiden en het snappen van de betekenis van deze woorden. We denken ook na over wat we willen zeggen, zetten de betekenis van onze boodschap om in klanken en produceren vervolgens deze klanken. Al deze ingewikkelde processen gebeuren binnen millisecondes, wat resulteert in schijnbaar moeiteloze conversaties. Als je bedenkt wat we allemaal moeten doen om te praten en te luisteren en hoe snel dit allemaal gaat, is het verrassend dat resultaten van sommige studies suggereren dat we deze twee dingen - denken wat we willen zeggen terwijl we luisteren naar anderen - tegelijk kunnen. Dit zou betekenen dat we kunnen multitasken tijdens conversaties, hoewel het ook bekend is dat multitasken erg moeilijk is. Er is dus een paradox: het is moeilijk om twee dingen tegelijk te doen, maar kennelijk kunnen we plannen wat we willen zeggen terwijl we luisteren naar anderen. In deze thesis onderzoek ik of mensen echt spraak kunnen plannen terwijl ze naar spraak luisteren om te kijken of we inderdaad op deze manier multitasken tijdens conversaties.

Om het antwoord op deze vraag te vinden, heb ik het probleem wat gereduceerd zodat ik experimenten simpel kon houden: proefpersonen benoemden plaatjes (een manier om hen spraak te laten plannen) terwijl ze via een koptelefoon luisterden naar lettergrepen, tonen of andere geluiden (luisteren naar spraak versus geluiden die geen spraak zijn). Het voordeel van experimenten met een simpel ontwerp is dat ze mensen kunnen dwingen om twee simpele taaltaken tegelijk te doen. Daardoor kunnen we precies meten of het plannen van spraak terwijl je naar andere spraak luistert op dezelfde manier verloopt als wanneer je naar andere geluiden luistert, of dat het makkelijker of moeilijker is.

In hoofdstuk 2 van deze thesis analyseerde ik studies die onderzochten hoe we het plannen van of luisteren naar spraak kunnen combineren met andere taken. Eerdere resultaten suggereren dat het lastig is om te plannen wat we willen zeggen terwijl we iets anders aan het doen zijn. Luisteren naar anderen tijdens een andere taak is ook moeilijk, maar misschien niet zo moeilijk als onze eigen spraak voorbereiden. Samengevat toont dit dus aan dat het nogal moeilijk zou moeten zijn om spraak te plannen terwijl we luisteren naar anderen. Toch is er, zoals ik eerder al zei, ook bewijs dat we spraak plannen terwijl we luisteren naar onze gesprekspartner. Daarom stelde ik een model voor dat beschrijft hoe we het plannen en het begrijpen van spraak zouden kunnen combineren. In dit model kunnen sommige processen die deel uitmaken van plannen en begrijpen in parallel uitgevoerd worden, maar sommige ook niet. Enkele hypotheses die voortkwamen uit dit model werden later getest in experimenten die beschreven worden in verschillende hoofdstukken.

In hoofstukken 3 en 4 heb ik eerst getest of mensen plaatjes kunnen benoemen terwijl ze lettergrepen of tonen horen. Daarbij werd aan proefpersonen gevraagd om via een druk op één van twee knoppen aan te geven welke lettergreep of toon ze gehoord hadden. Resultaten toonden aan dat het langer duurt om een plaatje te benoemen terwijl je een lettergreep hoort. Dit suggereert dat het moeilijker is om te plannen wat je wil gaan zeggen terwijl je naar spraak luistert dan wanneer je een toon hoort. Proefpersonen drukten ook later op de knop wanneer ze moesten beslissen welke lettergreep ze gehoord hadden dan wanneer ze tonen moesten onderscheiden, wat zou kunnen verklaard worden doordat het plannen van de naam van het plaatje (taal) het luisterproces verstoorde bij lettergrepen (ook taal), maar niet bij tonen. Het zou echter ook kunnen dat we het om de een of andere reden moeilijker vinden om naar lettergrepen te luisteren en het verschil in reactietijd dus verklaard kan worden doordat lettergrepen 'moeilijke' geluiden zijn en tonen 'makkelijke' geluiden. Om deze alternatieve verklaring uit te kunnen sluiten, heb ik nog twee experimenten uitgevoerd. Proefpersonen werden opnieuw gevraagd om plaatjes te benoemen, maar deze keer hoorden ze geluiden die beter overeen kwamen in hoe 'moeilijk' ze waren (de akoestische eigenschappen van de geluiden werden gematcht). In hoofdstuk 3 hoorden alle proefpersonen dezelfde geluiden die speciaal voor dit experiment gecreëerd werden. Aan de helft van de proefpersonen werd verteld dat dit vervormde lettergrepen waren (taal), de andere helft kreeg de instructie dat ze naar door een computer gegenereerde ruis luisterden (geen taal). Omdat iedereen naar precies dezelfde geluiden luisterde, waren de akoestische kenmerken van geluiden in de 'taal' en 'geen taal' groep identiek. In het experiment beschreven in hoofdstuk 4 hoorden proefpersonen verschillende 'taal' en 'geen taal' geluiden, maar waren deze geluiden wel zo gemanipuleerd dat ze gelijkaardige akoestische kenmerken hadden. Tegen mijn verwachtingen in, vond ik dat als geluiden even 'moeilijk' waren (gelijkaardige akoestische kenmerken), proefpersonen evenveel tijd nodig hadden om het plaatje te benoemen en te beslissen welk geluid ze hoorden. Dit zou dus betekenen dat het benoemen van een plaatje terwijl je naar taal (lettergrepen) luistert niet moeilijk is omdat beide taken taalgerelateerd zijn, maar omdat sommige geluiden nu eenmaal moeilijker te verwerken zijn dan andere. Dit heeft belangrijke gevolgen voor hoe mensen spraak plannen terwijl ze luisteren. Misschien kunnen we meer plannen terwijl we luisteren naar woorden die makkelijker te verwerken zijn.

In hoofdstuk 5 testte ik opnieuw of mensen plaatjes konden benoemen terwijl ze naar lettergrepen of tonen luisterden. In een reeks experimenten zagen proefpersonen plaatjes waar een woord overheen was geschreven. Het woord konden gerelateerd zijn aan het plaatje (een plaatje van een mes met het woord 'zwaard' bijvoorbeeld), of ongerelateerd (een plaatje van een mes met het woord 'kat'). In het eerste experiment wilde ik graag onderzoeken of mensen de correcte naam voor het plaatje konden 'kiezen' terwijl ze naar taal (lettergrepen) of andere geluiden (tonen) luisterden. Het bleek niet uit te maken welke soort geluid er werd afgespeeld. Proefpersonen waren niet in staat de correcte naam te kiezen terwijl ze naar iets anders luisterden. In het tweede en derde experiment moesten proefpersonen telkens beslissen of ze het plaatje gingen benoemen of het woord op het plaatje gingen voorlezen, afhankelijk van welk geluid ze hoorden. Ik heb enkel geanalyseerd hoe lang het duurde om de plaatjes te benoemen. Ik wou weten of mensen de correcte naam voor het plaatje konden selecteren terwijl ze moesten beslissen wat de opdracht was. Als algemene conclusie voor experiment 2 en 3 kon ik stellen dat proefpersonen dit niet konden. Er was echter variatie onder de proefpersonen. Sommigen leken het plaatje wel correct te kunnen benoemen terwijl ze beslisten wat de taak was. Dit is een belangrijke bevinding, aangezien dit iets zegt over de flexibiliteit in de manier waarop mensen plannen wat ze willen ze zeggen terwijl ze naar andere spraak luisteren.

In hoofdstuk 6 heb ik hersensignalen gemeten van proefpersonen die plaatjes benoemden terwijl ze naar lettergrepen luisterden. De resultaten toonden niet aan dat ze dit tegelijkertijd deden, zoals eerder onderzoek voorspelde, maar ik vond ook niet dat mensen alleen maar luisterden en daarna benoemden, of omgekeerd, zoals te verwachten was op basis van resultaten uit deze thesis die aantoonden dat het lastig is om het plannen van spraak te combineren met luisteren. In plaats daarvan suggereerden de resultaten dat sommige onderdelen van het plannen en luisteren parallel aan elkaar uitgevoerd werden, maar dat processen die betrokken zijn bij het plannen van de naam van het plaatje later opnieuw werden geactiveerd in de hersenen en opnieuw werden uitgevoerd. Dit patroon van resultaten past niet binnen enige theorie over hoe we tegelijk spraak kunnen plannen en kunnen luisteren naar spraak.

Samengevat tonen de resultaten uit deze thesis aan dat het moeilijk is om twee dingen tegelijk te doen. Het is nog moeilijker om twee taalgerelateerde taken - zoals het plannen van en het luisteren naar taal - op hetzelfde moment te doen. Dit doet vermoeden dat we in dagelijkse conversaties niet altijd plannen wat we willen zeggen terwijl we naar iemand anders luisteren, maar dat er veel flexibiliteit is in hoe mensen deze twee dingen cordineren. Het lijkt er dus op dat mensen soms tegelijkertijd kunnen plannen en luisteren, maar soms ook niet.

English summary

We spend a lot of our time talking and listening to other people. Despite how often this happens, and how easy it feels, the ease with which we have conversations is quite surprising. Very rapidly, people take the sounds coming from someone else's mouth, work out where the words are, and what those words mean. People also think about what they want to say, turn the meaning of what they want to say into sounds, and produce those sounds. All of these complex processes happen in milliseconds, leading to effortless conversation. Considering the amount that needs to be done to speak and listen, and how quickly it is done, it is surprising that evidence suggests that we can do these two things - thinking about what we want to say (i.e. planning our speech) while listening to others - at the same time. This means we multitask with planning and listening to speech. Yet, it is well-known that multitasking - doing two things at once - is hard. Thus, we are left with a paradox: it is hard for people to multitask, yet apparently we can plan what we want to say and listen to others at the same time. In the work in this thesis I investigated whether people actually can plan what they want to say while they listen to speech, to see whether this type of multitasking is what we do in conversation.

In order to investigate this question, I scaled down the problem to make the experiments very simple: people named pictures (a way to make them plan speech) while they listened to syllables, tones, or other sounds being played through headphones (to mimic listening to speech or non-speech). A benefit of carrying out experiments with this simple design is that we can force people to carry out two simple language tasks at the same time. We can then precisely measure whether planning speech while listening to speech is the same as when listening to another sound, or whether it is easier, or harder.

In Chapter 2 in this thesis, I reviewed evidence investigating how we combine planning speech or listening to speech with other tasks. The evidence suggests that planning what we want to say while carrying out other tasks is challenging. Listening to other people is also hard, yet maybe not as hard as planning our own speech. Taken together, the findings suggest that it should be quite diffi-

cult to plan what we want to say while we listen to other people. However, as stated above, there is evidence that we do plan our speech while listening to our conversation partner. Therefore, I proposed a model describing how we could combine planning speech and comprehending speech. In this model, some processes involved in speech planning and speech comprehension can be carried out at the same time, and some cannot. Some predictions from this model were tested in later experiments.

In Chapters 3 and 4, I tested whether people can name pictures while they listened to linguistic or non-linguistic sounds. In experiment 1 in Chapter 3, I tested whether people can name pictures while they heard syllables or tones. Participants also had to press one of two buttons depending on which syllables or tones they heard. The results showed that it takes longer for people to name a picture when they hear a syllable, suggesting that planning what they want to say while listening to speech is harder than planning speech when hearing a tone. People also took longer to respond to syllables compared to tones, which may be because planning the picture name interfered with the listening process with syllables, but not with tones. However, these results could be because, for some reason, the syllables are harder sounds to listen to and process than the tones. This implies that the syllables are 'hard' sounds and the tones are 'easy' sounds. To try to rule out this possibility, I carried out two further experiments where again, participants named pictures, but they heard sounds which were more closely matched to one another in their difficulty (the acoustic properties of the sounds were matched). In experiment 2 in Chapter 3, all participants heard the same specially created sounds. One group of participants were told the sounds were distorted syllables (language), and the other group were told they were computer-generated noise (non-language). Because everyone heard the exact same sounds, their acoustics were equal. In the experiment in Chapter 4, I created two computerised vowel sounds (language), and then manipulated these vowel sounds to create two non-language sounds (called 'musical rain'). Because the non-language sounds were generated from the vowels, the sounds were acoustically very similar. In this experiment, all participants heard both language and non-language sounds (like in experiment 1 of Chapter 3). Contrary to what was expected, in experiment 2 in Chapter 3, and in the experiment in Chapter 4, I found that if the sounds were equally difficult to process (i.e. their acoustic properties were very similar) then it took people the same amount of time to name the pictures and to respond to the sounds. This suggests that any difficulties in naming pictures while listening to language sounds (like syllables) are not because both tasks are language-related, but because some sounds are just harder to listen to than others. This has important implications for how people plan speech while listening, as it may be that if some words are easier for us to listen to than others, we can plan more of what we want to say.

In Chapter 5, I again tested whether people could name pictures while listening to syllables or tones. In these experiments I showed participants pictures with words written over the pictures. These words could be related to the pictures (like seeing a picture of a 'knife' and having the word 'sword' written across it), or unrelated (a picture of a knife with the word 'cat' written across it). In the first experiment, I was interested in finding out if people could 'choose' the correct name for the picture while listening to either language sounds (syllables) or different sounds (tones). I found that it didn't matter what kind of sound was played; people were unable to choose the correct picture name at the same time as listening to any sound. In the second and third experiments, people decided whether to name the picture or read the word out loud, based on the sound they heard. For instance, one of the syllables or tones would indicate that participants should name the picture (e.g., 'knife'), and the other syllable or tone would indicate that participants should read the word out loud (e.g., 'sword')'. I only looked at how long it took people to name the pictures. I again wanted to see if people could choose the name of the picture while deciding, based on the syllable or tone they heard, whether they needed to name the picture or read the word out loud. This is because a previous study suggested that people could make these task decisions when listening to tones and naming pictures. I wanted to find out if people could also make these decisions when listening to syllables and naming pictures, as here two language tasks are combined. I found across both the second and third experiments (which varied slightly in their structure) that generally, people could not choose the correct name of the picture at the same time as making the decision about this task based on the sound they heard. However, participants were very variable in how they combined preparing the name of the picture while making these task decisions. Some people seemed able to choose the correct picture name while making the task decision based on the sound, whereas others could not. Importantly, this suggests that there is a lot of flexibility in how people can combine two tasks, and this has implications for how people combine the two tasks of planning what they want to say while they listening to other speech.

In Chapter 6, I recorded the neural responses (brain signals) when people named pictures while listening to syllables. The results did not suggest that people listen to syllables and name pictures at the same time, as predicted by previous evidence. The results also did not suggest that people only listened and then named, or vice versa, as we predicted based on the fact that results from this thesis show it is hard to combine planning speech while listening. Instead, the results suggested that some parts of planning and listening were carried out in parallel, but processes in planning the name of the picture were later reactivated in the brain and carried out again. This pattern of results does not fit with any theories of how we might carry out planning speech and listening to speech at the same time.

In general, the results from this thesis show that doing two things at the same time is hard. It is even harder to do two language things - such as planning and listening to speech - at the same time. This suggests that in everyday conversation, we probably don't always plan what we want to say when we are listening to someone else speak. Instead, there is a lot of flexibility in how people coordinate planning what they want to say and listening to others, and it seems that sometimes people can plan and listen together, and sometimes not.

Résumé en français

Les humains passent une grande partie de leurs temps à écouter les autres. Malgré le fait qu'on le fait fréquemment et sans difficulté apparente, le simple fait de pouvoir discuter est franchement étonnant. On est capable, à une vitesse surprenante, de reconnaitre les sons produit par l'appareillage vocal d'un interlocuteur, de leurs attribuer une identité lexicale, et d'en comprendre le sens. En même temps, on réfléchit à ce qu'on a envie de dire, on convertit nos pensées en paroles, puis en séries de mouvements articulatoires qui produisent enfin les sons désirés. Tous ces processus complexes ont lieu dans l'espace de quelques millisecondes, permettant à une conversation de se dérouler de manière fluide. Si nous nous permettons de considérer toutes ces exigences cognitives, il devient surprenant de constater que nous sommes néanmoins capables de tenir des discussions. Il semblerait que nous puissions simultanément songer à ce que nous allons dire en même temps que d'écouter, et de comprendre les autres. Ceci voudrait dire que nous sommes régulièrement obligés de nous engager à effectuer deux tâches en même temps. La recherche à ce sujet (« multi-tasking » en anglais, une traduction fort insatisfaisante en français serait « multi-tâche ») nous a démontré que l'exécution de multiples processus en parallèle est difficile, et que cela encoure un certain cout pour chacun des processus ainsi effectués. Ils se déroulent de manière plus lente et moins précise. Ainsi sommes-nous confrontés à un paradoxe : nous avons du mal à effectuer des taches en parallèle, mais nous sommes, apparemment, parfaitement capable de planifier ce que nous aurions envie de dire à nos interlocuteurs en même temps que de comprendre ce qu'ils nous disent. La recherche présentée dans cette thèse touche à cette question – est-ce qu'on est réellement capable de planifier ce qu'on a envie de dire pendant qu'on écoute de la parole ? Est-ce ce genre de processus « multi-tâche » qui a lieu, lorsque nous discutons ?

Afin d'investiguer cela, j'ai dû réduire l'échelle du problème. Les expériences que j'ai effectuées sont plutôt simples. Les participants ont dû nommer des images (ce qui les a obligés de planifier de la parole) pendant qu'ils écoutaient des syllabes, des tons, ou d'autres sons, qu'ils devaient distinguer. Ainsi étaient-ils obligés de produire de la parole et traiter un signale auditif simultanément. Ceci m'a permis de mesurer précisément si la planification de la parole en parallèle avec le traitement d'un son se déroule différemment en fonction de si ce son est linguistique ou non. C'est-à-dire, est-ce que le traitement d'un son linguistique exerce une influence relativement plus importante sur notre planification et exécution d'une tache de dénomination qu'un son non-linguistique? Est-ce que les ressources auxquelles nous faisons appelles pour effectuer les deux taches sont les mêmes ? Y-a-t-il une concurrence entre le compréhension et la production de la parole ?

Dans le deuxième chapitre de cette thèse je présente un résumé de la littérature qui a déjà touché à ce sujet. Les données suggèrent que la planification de ce que nous aimerions dire en même temps que l'exécution d'autres taches constituerait effectivement un défi. Par conséquent, je propose un modèle qui décrit comment la planification de la parole et la compréhension pourraient être effectués simultanément, malgré le fait d'éventuellement solliciter les mêmes ressources cognitives. Selon ce modèle, certains processus impliqués dans la planification de la parole et la compréhension pourraient avoir lieu ensemble, et d'autres non. Les prédictions de ce modèle sont testées dans les chapitres qui suivent.

Les chapitres trois et quatre décrivent des expériences qui visent à évaluer si les participants sont capables de nommer des images pendant qu'ils effectuent une tâche de catégorisation syllabique (donc, linguistique) ou tonale (non-linguistique). Les participants ont produit les noms des images plus lentement, lorsque les épreuves contenaient des syllabes plutôt que des tons. Ceci indique que la catégorisation de syllabes empiète plus sur le domaine cognitif linguistique que la catégorisation de tons. Les participants ont aussi répondu plus lentement aux syllabes qu'aux tons, ce qui suggère que la catégorisation de syllabes est plus difficile que la catégorisation des tons, ce qui pourrait indiquer que la dénomination d'images emploie des ressources nécessitées pour la catégorisation syllabique mais pas tonale. Cependant, cette dissociation pourrait simplement résulter du fait que les syllabes sont acoustiquement plus complexes et donc plus difficiles à traiter que les tons. Afin de vérifier si cet effet provenait uniquement de la complexité acoustique j'ai ensuite comparé deux cas de figure. Dans un premier temps j'ai effectué une expérience avec les mêmes stimuli et la même tache, en indiquant simplement à un groupe de participants que tous les sons étaient de sons de parole, et à un autre groupe que les sons étaient des bruits non-linguistiques. Dans un deuxième temps, j'ai effectué une expérience très similaire, cette fois employant des sons syllabiques et tonales qui étaient produits de manière acoustiquement contrôlée, les rendant plus semblables au niveau de leur complexité. Contrairement à mes attentes, j'ai découvert que peu importe la nature linguistique ou non des sons, les participants répondent plus lentement aux images lorsqu'ils sont confrontés à des sons plus complexes. C'est-à-dire que les plus importants ralentissements observés lors que les participants étaient obligés de nommer des images en écoutant des syllabes étaient probablement dû, non pas au fait que les sons étaient linguistiques ou traités de manière linguistique (impliquant un accès simultané aux ressources cognitives dévouées au langage), mais simplement à leur complexité.

Dans le cinquième chapitre j'ai de nouveau demandé à des participants d'effectuer une tâche de dénomination d'image en même temps qu'une catégorisation auditive. Cette fois, les images qui devaient être nommées étaient superposés d'un mot, qui était soit conceptuellement lié à l'image (par exemple une image d'un couteau figurant le mot 'épée') soit pas (par exemple une image d'un couteau figurant le mot 'chat'). Ici les participants ont dû choisir une réponse vocale en fonction des syllabes que leurs étaient présentés. Ils devaient soit nommer l'image ou lire le mot à voix haute. Une précédente étude a démontré que ce genre décision et possible si les sons présentés sont de simples tons. Ainsi, cette expérience a permis de tester s'il le fait d'écouter des syllabes provoque une interférence avec le choix de réponse linguistique, tandis que faire la même chose en écoutant et catégorisant des tons n'en provoque aucune. Il s'est avéré qu'effectivement, les participants n'ont en moyenne pas réussi à effectuer cette tache de sélection en présence de syllabes. Ceci était confirmé lors d'une deuxième et troisième variante de cette expérience. Si, en moyenne les participants ne sont pas capa les d'effectuer la tâche, certains ont néanmoins réussi, ce qui suggère que les capacités de gérer les situations « multi-tâches » sont distribuées de manière inégale à travers la population.

Le sixième chapitre présente une étude dans laquelle j'ai enregistré les signaux neuronaux des participants qui effectuaient la double-tache de dénomination et de discrimination acoustique. J'ai employé une technique qui s'appelle la magnétoencéphalographie qui est sensible aux décharges électriques produites lorsque les assemblages de neurones du cerveau s'activent en réponse à une tâche. Ceci permet d'examiner les patrons d'activité neuronale qui corresponde à, par exemple, la dénomination, la sélection, ou le traitement des stimuli auditifs. Les résultats indiquent qu'au niveau cérébral, les participants ne font pas réellement deux choses en même temps tout au long des épreuves de cette tâche. Cependant, leurs cerveaux n'ont pas non-plus l'air d'alterner entre les deux taches. Il semblerait plutôt que certaines parties de la planification et du traitement acoustique ont lieux simultanément, mais que les processus de planification sont interrompus par les processus de catégorisation des sons, et repris plus tard. Cette séquence de traitement n'est pas prédite par les modelés existants.

Globalement, les résultats présentés dans cette thèse démontrent que l'exécution de deux taches en même temps est difficile. Et que cela est encore plus difficile lorsque les taches, tels la dénomination et l'identification de syllabes, sont de nature linguistique. Ceci suggère que lors d'une discussion, nous ne planifions probablement pas ce que nous voulons dire entièrement en parallèle avec le traitement de la parole de notre interlocuteur. Il existe des contraintes cognitives qui nous l'empêche. Il semblerait qu'il y a beaucoup de souplesse par rapport à la coordination des ressources cognitives lors de la conversation – par moment nous pouvons planifier et comprendre en même temps, mais pas sans cesse.

Acknowledgements

It takes the work of more than one person to write a thesis, even if only one name is on the front. Firstly, this couldn't have been done without amazing supervisors. Antje, you are a phenomenally good scientist. Your comments and criticisms and unfailing support have made me a better researcher, and I especially appreciated your support when none of my results made sense. You also really helped when I wasn't doing well with the stress of the dissertation, for which I cannot express enough thanks. Finally, when I first began my PhD you bought me Yorkshire Tea, and I knew from that moment I could never have a better boss. Sara, you were always available to talk through my ideas (which in the early days made no sense), and I always appreciated your careful reading of my work. Alexis, you were thrown in to supervise a project I had ample enthusiasm for but little experience in, and you were always so supportive, enthusiastic, and positive about it that you made me want to continue, even through the struggles.

Thank you very much to Peter Hagoort, Vitória Piai, and Marina Laganaro, for taking time at quite short notice to read and evaluate my thesis.

I was very privileged to be based at the MPI, with the phenomenal resources there, which allowed all this work to happen. Additionally, no experiments could ever be conducted without the help of the TG, no personal issues could ever be resolved with Operations, and no PhD programme would be carried out without Kevin Lam (and earlier Dirkje and Els). Thanks so much to all!

My two lovely paranymphs, Emma and Caitlin, have been around for the whole thing. Emma, you are a friend from pre-PhD. I remember when we met at the top of the DSB, and I wondered who this exotic British-American person was. We became good friends through the struggles of masters life. To then have you here in Nijmegen has been wonderful (for all the struggles of PhD life!) and you have always been someone I can turn to when happy and sad. To many more years of friendship! Caitlin, we started at roughly the same time, and became friends almost instantly. To multiple meals in multiple houses, movies we've seen, and glasses of wine we've drunk, I am so happy we met. You know a true friend when they help you move house by dismantling your bed, carrying it upstairs, and reassembling it, all while they are really, really sick.

To all my PoL colleagues, past and present, being in our department has been fantastic and I've learnt so much from everyone through the years. Special thanks to Sara, who in one lab meeting asked 'Maybe your effect is because the sounds are different?', which spawned two experiments in this thesis and follow-up studies even now; to Joe and Jeroen, my praat-saviours; to Hans Rutger for creating stimuli; to Suzanne, who put my mind at ease about being an MPI PhD before I began; to Caitlin and Laurel for forest walks, chats, and BOM; additionally to Phillip, who helped me so much when I was going a little crazy towards the end of the writing; to Eirini for Open Science chats which always become regular chats; and to Limor for multiple dinners. And thanks also go to my office mates throughout the years: Markus, Sara, Greta, Federica and Merel. You've all let me bounce ideas off you, given my feedback, and generally been lovely to spend time with. And Markus, I will never forget that mosaic bench!

To the pub quiz team through the years - Erik, Peter, Daniel, Matthias and Ashley - while I sometimes slacked off with my quiz duties, hanging out with you guys on Wednesday nights was a highlight of my PhD. Especially when we introduced pub quiz dinners! And to the members of Bridges Book Club - here I managed to find a wonderful group of people to talk about books and life with, and I'm sad to have to leave.

To Becky, my oldest and bestest of friends. You were supportive every step of the way during my PhD, and made me feel so much better whenever we chatted about anything. Thank you for being you!

And especial thanks to my family. Mum and Graham, when I first started university you always said to me 'If you don't like it, just come home', and I still think about that now. Thank you for being so supportive of my desire to stay in academia and basically be a student forever. To Dad and Kay, thank you for always asking what I'm up to. To Span, special thanks for making me feel better when I was having my breakdown. A Joëlle, Jean-Marie, Emeline et Cédric – vous m'avez si gentiment intégrés au sein de votre famille, je vous en suis infiniment reconnaissante.

Et enfin, je dois les plus grand remerciements à toi, Flo. C'est toi qui m'a encouragé à candidater pour ce poste de doctorante, tu as déménagé à l'étranger avec moi à deux reprises (et tu es sur le point de le refaire une troisième fois), et je n'aurais absolument pas pu finir ce doctorat sans ton soutien pendant tous

ces mois. Cette thèse est autant la tienne que la mienne, et je me réjouis de notre prochaine aventure !

Curriculum Vitae

Amie was born in Portsmouth, on the south coast of the United Kingdom, in 1990. She moved to the University of Edinburgh in 2009 to carry out an undergraduate degree MA Linguistics, completed in 2013. During this degree, in 2011-2012, Amie spent an alternatingly cold and dark, or cold and light, incredibly beautiful Erasmus year at the University of Tromsø - The Arctic University of Norway. After completing her undergraduate degree, Amie went on to obtain an MSc Psychology of Language in 2014, also from the University of Edinburgh. She moved to the Max Planck Institute for Psycholinguistics to start her PhD in 2014. After a rocky start with a project which was abandoned, Amie submitted her PhD thesis towards the end of 2018. Amie currently works as a post-doctoral researcher at Aix-Marseille University in France.

Publications

Fairs, A., Bögels, S. & Meyer, A. S. (2018). 'Dual-tasking with simple linguistic tasks: Evidence for serial processing'. *Acta Psychologica*, 191, 131-148. doi: 10.1016/j.actpsy.2018.09.006.

Hall-Lew, L., Paiva Couceriro, I. & **Fairs, A.** (to appear). 'Credibility without Intelligibility: Implications for Hearing Vernacular Speakers.' In Blake, R & Buchstaller, I. (Eds.) *The Routledge Companion to the Work of John Rickford*. ppTBD. Routledge.

Hall-Lew, L., **Fairs, A.** & Lew, A. (2015). 'Tourists Attitudes towards Linguistic Variation in Scotland'. In E. Togersen, S. Haerstad, B. Maehlum, & U. Roeyneland (Eds.), *Language Variation - European Perspectives V* (pp. 99-110). Amsterdam: Benjamins.

Hall-Lew, L., Lew, A. & **Fairs, A.** (2014). 'Language and Tourism in Sabah, Malaysia and Edinburgh, Scotland'. In B. O'Rourke, N. Bermingham, & S. Brennen (Eds.), *Opening New Lines of Communication in Applied Linguistics: Proceedings of the 46th BAAL Annual Meeting* (pp. 253-259). London: Scitsiugnil Press.

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda I. van Turennout*
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. *Niels O. Schiller*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*
4. The open-/closed class distinction in spoken-word recognition. *Alette Petra Haveman*
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*
6. Gesture and speech production. *Jan-Peter de Ruiter*
7. Comparative intonational phonology: English and German. *Esther Grabe*
8. Finiteness in adult and child German. *Ingeborg Lasser*
9. Language input for word discovery. *Joost van de Weijer*
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*
11. Producing past and plural inflections. *Dirk J.Janssen*
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*
13. From speech to words. *Arie H. van der Lugt*
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*
15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*
16. Language-specific listening: The case of phonetic sequences. *Andrea Christine Weber*

17. Moving eyes and naming objects. *Femke Frederike van der Meulen*
18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*
19. Morphology in speech comprehension. *Kerstin Mauth*
20. Morphological families in the mental lexicon. *Nivja Helena de Jong*
21. Fixed expressions and the production of idioms. *Simone Annegret Sprenger*
22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). *Birgit Hellwig*
23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*
24. Contextual influences on spoken-word processing: An electrophysiological approach. *Danielle van den Brink*
25. Perceptual relevance of prevoicing in Dutch. *Petra Martine van Alphen*
26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*
27. Producing complex spoken numerals for time and space. *Marjolein Henriëtte Wilhelmina Meeuwissen*
28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl Jenny Judith Karin Kemps*
29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*
30. A grammar of Jalonke argument structure. *Friederike Lüpke*
31. Agrammatic comprehension: An electrophysiological approach. *Marijtje Elizabeth Debora Wassenaar*
32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*
33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*
34. Phonetic and lexical processing in a second language. *Mirjam Elisabeth Broersma*
35. Retrieving semantic and syntactic word properties: ERP studies on the time course in language comprehension. *Oliver Müller*
36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*

37. Sensitivity to detailed acoustic information in word recognition. *Keren Batya Shatzman*

38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*

39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*

40. The acquisition of phonological structure: Distinguishing contrastive from non-constrative variation. *Christiane Dietrich*

41. Cognitive cladistics and the relativity of spatial cognition. *Daniel Haun*

42. The acquisition of auditory categories. *Martijn Bastiaan Goudbeek*

43. Affix reduction in spoken Dutch: Probabilistic effects in production and perception. *Mark Plumaekers*

44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Madalla Kooijman*

45. Space and iconicity in German sign language (DGS). *Pamela M. Perniss*

46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*

47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*

48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*

49. Phoneme inventories and patterns of speech sound perception. *Anita Eva Wagner*

50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*

51. A grammar of Savosavo: A Papuan language of the Solomon Islands. *Claudia Ursula Wegener*

52. Prosodic structure in speech production and perception. *Claudia Kuzla*

53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. *Sarah Schimke*

54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*

55. Processing the fine temporal structure of spoken words. *Eva Reinisch*

56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*

57. Processing strongly reduced forms in casual speech. *Susanne Brouwer*
58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. *Miriam Ellert*
59. Lexical interactions in non-native speech comprehension: Evidence from electroencephalography, eye-tracking, and functional magnetic resonance imaging. *Ian FitzPatrick*
60. Processing casual speech in native and non-native language. *Annelie Tuinman*
61. Split intransitivity in Rotokas, a Papuan language of Bougainville. *Stuart Payton Robinson*
62. Evidentiality and intersubjectivity in Yurakaré: An interactional account. *Sonja Gipper*
63. The influence of information structure on language comprehension: A neurocognitive perspective. *Lin Wang*
64. The meaning and use of ideophones in Siwu. *Mark Dingemanse*
65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. *Marco van de Ven*
66. Speech reduction in spontaneous French and Spanish. *Francisco Torreira*
67. The relevance of early word recognition: Insights from the infant brain. *Caroline Mary Magteld Junge*
68. Adjusting to different speakers: Extrinsic normalization in vowel perception. *Matthias Johannes Sjerps*
69. Structuring language: Contributions to the neurocognition of syntax. *Katrien Rachel Segaert*
70. Infants' appreciation of others' mental states in prelinguistic communication: A second person approach to mindreading. *Birgit Knudsen*
71. Gaze behavior in face-to-face interaction. *Federico Rossano*
72. Sign-spatiality in Kata Kolok: How a village sign language of Bali inscribes its signing space. *Connie de Vos*
73. Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning. *Attila Andics*
74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation. *Marijt Witteman*
75. The use of deictic versus representational gestures in infancy. *Daniel Puccini*

76. Territories of knowledge in Japanese conversation. *Kaoru Hayano*

77. Family and neighbourhood relations in the mental lexicon: A cross-language perspective. *Kimberley Mulder*

78. Contributions of executive control to individual differences in word production. *Zeshu Shao*

79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing. *Patrick van der Zande*

80. High pitches and thick voices: The role of language in space-pitch associations. *Sarah Dolscheid*

81. Seeing what's next: Processing and anticipating language referring to objects. *Joost Rommers*

82. Mental representation and processing of reduced words in casual speech. *Iris Hanique*

83. The many ways listeners adapt to reductions in casual speech. *Katja Pöllmann*

84. Contrasting opposite polarity in Germanic and Romance languages: Verum Focus and affirmative particles in native speakers and advanced L2 learners. *Giuseppina Turco*

85. Morphological processing in younger and older people: Evidence for flexible dual-route access. *Jana Reifegerste*

86. Semantic and syntactic constraints on the production of subject-verb agreement. *Alma Veenstra*

87. The acquisition of morphophonological alternations across languages. *Helen Buckler*

88. The evolutionary dynamics of motion event encoding. *Annemarie Verkerk*

89. Rediscovering a forgotten language. *Jiyoun Choi*

90. The road to native listening: Language-general perception, language-specific input. *Sho Tsuji*

91. Infants'understanding of communication as participants and observers. *Gudmundur Bjarki Thorgrímsson*

92. Information structure in Avatime. *Saskia van Putten*

93. Switch reference in Whitesands. *Jeremy Hammond*

94. Machine learning for gesture recognition from videos. *Binyam Gebrekidan Gebre*

95. Acquisition of spatial language by signing and speaking children: A comparison of Turkish sign language (TID) and Turkish. *Beyza Sumer*
96. An ear for pitch: On the effects of experience and aptitude in processing pitch in language and music. *Salomi Savvatia Asaridou*
97. lncrementality and Flexibility in Sentence Production. *Maartje van de Velde*
98. Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture. *Edwin van Leeuwen*
99. The request system in Italian interaction. *Giovanni Rossi*
100. Timing turns in conversation: A temporal preparation account. *Lilla Magyari*
101. Assessing birth language memory in young adoptees. *Wencui Zhou*
102. A social and neurobiological approach to pointing in speech and gesture. *David Peeters*
103. Investigating the genetic basis of reading and language skills. *Alessandro Gialluisi*
104. Conversation electrified: The electrophysiology of spoken speech act recognition. *Rósa Signý Gisladottir*
105. Modelling multimodal language processing. *Alastair Charles Smith*
106. Predicting language in different contexts: The nature and limits of mechanisms in anticipatory language processing. *Florian Hintz*
107. Situational variation in non-native communication. *Huib Kouwenhoven*
108. Sustained attention in language production. *Suzanne Jongman*
109. Acoustic reduction in spoken-word processing: Distributional, syntactic, morphosyntactic, and orthographic effects. *Malte Viebahn*
110. Nativeness, dominance, and the flexibility of listening to spoken language. *Laurence Bruggeman*
111. Semantic specificity of perception verbs in Maniq. *Ewelina Wnuk*
112. On the identification of FOXP2 gene enhancers and their role in brain development. *Martin Becker*
113. Events in language and thought: The case of serial verb constructions in Avatime. *Rebecca Defina*
114. Deciphering common and rare genetic effects on reading ability. *Amaia Carrión Castillo*

115. Music and language comprehension in the brain. *Richard Kunert*

116. Comprehending Comprehension: Insights from neuronal oscillations on the neuronal basis of language. *Nietzsche H.L. Lam*

117. The biology of variation in anatomical brain asymmetries. *Tulio Guadalupe*

118. Language processing in a conversation context. *Lotte Schoot*

119. Achieving mutual understanding in Argentine Sign Language. *Elizabeth Manrique*

120. Talking Sense: the behavioural and neural correlates of sound symbolism. *Gwilym Lockwood*

121. Getting under your skin: The role of perspective and simulation of experience in narrative comprehension. *Franziska Hartung*

122. Sensorimotor experience in speech perception. *Will Schuerman*

123. Explorations of beta-band neural oscillations during language comprehension: Sentence processing and beyond. *Ashley Lewis*

124. Influences on the magnitude of syntactic priming. *Evelien Heyselaar*

125. Lapse organization in interaction. *Elliott Hoey*

126. The processing of reduced word pronunciation variants by natives and foreign language learners: Evidence from French casual speech. *Sophie Brand*

127. The neighbors will tell you what to expect: Effects of aging and predictability on language processing. *Cornelia Moers*

128. The role of voice and word order in incremental sentence processing. Studies on sentence production and comprehension in Tagalog and German. *Sebastian Sauppe*

129. Learning from the (un)expected: Age and individual differences in statistical learning and perceptual learning in speech. *Thordis Neger*

130. Mental representations of Dutch regular morphologically complex neologisms. *Laura de Vaan*

131. Speech production, perception, and input of simultaneous bilingual preschoolers: Evidence from voice onset time. *Antje Stoehr*

132. A holistic approach to understanding pre-history. *Vishnupriya Kolipakam*

133. Characterization of transcription factors in monogenic disorders of speech and language. *Sara Busquets Estruch*

134. Indirect request comprehension in different contexts. *Johanne Tromp*

135. Envisioning language - An exploration of perceptual processes in language comprehension. *Markus Ostarek*

136. Listening for the WHAT and the HOW: Older adults' processing of semantic and affective information in speech. *Juliane Kirsch*

137. Let the agents do the talking: On the influence of vocal tract anatomy on speech during ontogeny and glossogeny. *Rick Janssen*

138. Age and hearing loss effects on speech processing. *Xaver Koch*

139. Vocabulary knowledge and learning: Individual differences in adult native speakers. *Nina Mainz*

140. The face in face-to-face communication: Signals of understanding and non-understanding. *Paul Hömke*

141. Person reference and interaction in Umpila/Kuuku Ya'u narrative. *Clair Hill*

142. Beyond the language given: The neurobiological infrastructure for pragmatic inferencing. *Jana Bašnáková*

143. From Kawapanan to Shawi: Topics in language variation and change. *Luis Miguel Rojas-Berscia*

144. On the oscillatory dynamics underlying speech-gesture integration in clear and adverse listening conditions. *Linda Drijvers*

145. Linguistic dual-tasking: Understanding temporal overlap between production and comprehension. *Amie Fairs*