Coordination of spoken language production and comprehension:

How speech production is affected by irrelevant background speech

Funding Body
This research was funded by the Max Planck Society for the Advancement of Science (www.mpg.de/en) as part of the IMPRS for Language Sciences project.

International Max Planck Research School (IMPRS) for Language Sciences
The educational component of the doctoral training was provided by the International Max Planck Research School (IMPRS) for Language Sciences. The graduate school is a joint initiative between the Max Planck Institute for Psycholinguistics and two partner institutes at Radboud University – the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour. The IMPRS curriculum, which is funded by the Max Planck Society for the Advancement of Science, ensures that each member receives interdisciplinary training in the language sciences and develops a well-rounded skill set in preparation for fulfilling careers in academia and beyond. More information can be found at www.mpi.nl/imprs

The MPI series in Psycholinguistics
Initiated in 1997, the MPI series in Psycholinguistics contains doctoral theses produced at the Max Planck Institute for Psycholinguistics. Since 2013, it includes theses produced by members of the IMPRS for Language Sciences. The current listing is available at www.mpi.nl/mpi-series

ISBN: 978-94-92910-49-3
Cover design and lay-out by Jieying He
Printed and bound by Ipskamp Drukkers, Enschede

 The research reported in this thesis was conducted at the Max Planck Institute for Psycholinguistics, in Nijmegen, the Netherlands

Coordination of spoken language production and comprehension:

How speech production is affected by irrelevant background speech

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 voor promoties
in het openbaar te verdedigen op

woensdag 7 juni 2023
om 16.30 uur precies

door

Jieying He
geboren op 20 januari 1993
te Gansu (China)

Promotor:

Prof. dr. A.S. Meyer

Copromotoren:

Dr. L. Brehm (University of California, Verenigde Staten)

Dr. A.K. Creemers (Universiteit van Amsterdam)

Dr. C. Frances (Max Planck Institute for Psycholinguistics)

Manuscriptcommissie:

Prof. dr. A.P.A. Roelofs

Dr. S. Bögels

Prof. dr. I. Ivanova (The University of Texas at El Paso, Verenigde Staten)

Contents

1 | General Introduction 7

1.1 A model relevant for speech production in background speech 8

1.2 Irrelevant speech effects and relevant theories 10

1.3 Modulation of focal task difficulty on the irrelevant speech effect 13

1.4 Current dissertation 15

2 | Concurrent listening affects speech planning and fluency: The roles of representational similarity and capacity limitation 18

2.1 Introduction 20

2.2 Experiment 1 27

2.3 Experiment 2 38

2.4 General Discussion 48

2.5 Conclusion 56

Appendices 57

3 | Conducting language production research online: A web-based study of semantic context and name agreement effects in multi-word production 73

3.1 Introduction 74

3.2 Methods 78

3.3 Results 84

3.4 Discussion 93

Appendices 99

4 | Effects of irrelevant unintelligible and intelligible background speech on spoken language production 103

4.1 Introduction 104

4.2 Experiment 1 110

4.3 Experiment 2 128

4.4 General Discussion 142

4.5 Conclusion 147

Appendices 149

5 | Auditory disruption by irrelevant background sentences on spoken language production 167

5.1 Introduction..168
5.2 Methods...174
5.3 Results...180
5.4 Discussion...192
5.5 Conclusions...198
Appendices..199

6 | General Discussion...207
6.1 Summary of the empirical studies...207
6.2 Lexical selection demand affects speech production performance..........................209
6.3 Irrelevant background speech affects spoken language production........................210
6.4 Speakers can control the processing of background speech 212
6.5 Modulation of the irrelevant speech effect by production difficulty 213
6.6 Recommendations for future research 215
6.7 Conclusions...216

Research Data Statement 219

References...221

Nederlandse Samenvatting..235

English Summary..239

Acknowledgments 243

Curriculum Vitae...247

Publications 249

MPI Series in Psycholinguistics...251

1 | General Introduction

Conversation, characterized by a regular exchange of turns between interlocutors, exists in all cultures and is the most common way of using language (Levinson, 2016). These exchanges involve speakers and listeners working together to reach a common communicative goal (e.g., Clark, 1996; Clark & Wilkes-Gibbs, 1986; Goodwin, 1981). Given the speed at which speakers and listeners switch roles during conversation, there must be an overlap between speech production processes and comprehension ones---speakers begin to plan their speech while listening to their interlocutors (Levinson, 2016). This means conversation requires dual-tasking between speech production and comprehension. Moreover, conversation is carried out in many different environments, such as in a quiet room, in a noisy restaurant, or on a busy train where a lively discussion or a phone conversation may be heard in the background. Background noise has been found to disrupt cognitive performance such as short-term memory (e.g., Banbury & Berry, 1998; Hellbrück & Liebl, 2008) and reading performance (e.g., Cauchard et al., 2012; Hyönä & Ekholm, 2016; Yan et al., 2018). This means, to understand how individuals manage to communicate effectively in everyday situations, it is essential to describe how they plan and comprehend speech in background noise.

So far, some aspects of this question have been explored. First, speech comprehension in noise has been extensively investigated, showing that the influence of noise on speech comprehension (or recognition) varies significantly according to the properties and types of noise (e.g., fluctuating versus steady noise, Moore, 2012; speech-like sounds, Bronkhorst, 2000; and speeded speech, Gordon-Salant & Fitzgibbons, 2004; degraded speech, Sharp et al., 2006), the type and number of speakers (e.g., multi-talker babble from male, female, or child, Cullington & Zeng, 2008; Pittman & Wiley, 2001; Sperry et al., 1997), and the listener's age (e.g., Duquesnoy & Plomp, 1980) and hearing ability (e.g., Badri et al., 2011; Humes & Roberts, 1990; Quist-Hanssen et al., 1978). Meanwhile, previous studies have also investigated the influence of background noise on qualities of speech production and demonstrated that individuals, for example, speak more loudly in noise than in quiet (i.e., *the Lombard effect*, Egan, 1972; Lane et al., 1970). However, none of these studies have investigated how speakers plan speech in the presence of verbal background noise (i.e., background speech). Thus, this dissertation explored how background speech interferes with speech production, which is both an essential and neglected aspect of the cognitive processes involved in conversation carried out in noisy contexts.

1.1 A model relevant for speech production in background speech

Because speaking over background speech involves both deliberate speech production and involuntary comprehension, it is necessary to understand how production and comprehension coordinate in this context. As outlined in Figure 1.1, during the course of generating an utterance, speakers go through three major stages: conceptualization, formulation, and articulation (Garrett, 1975; Levelt, 1989; Levelt et al., 1999). In the *conceptualization* stage, the speaker decides the message to be conveyed (concept preparation/semantic activation). The output of the conceptualization stage feeds into the *formulation*, where lexical concepts initiate lexical selection, activate lemmas semantically and syntactically, and then the lemmas become syllabified into their phonological forms (lexical selection and phonological encoding; Kempen & Huijbers, 1983). At the *articulation* stage, the phonological form gets phonetically encoded, which results in gestural scores that can be articulated. This means that the speech production process proceeds from lexical concepts to lemmas to lexical output forms and finally to output syllables.

Similar to spoken word production, word comprehension also requires a series of processing stages to translate speech sounds into meaning. This means it recruits several levels of representations including a phonetic component, a phonological component, a lexical component, and a conceptual component, which mediate between incoming speech signals and the computation of meaning (Gaskell & Marslen-Wilson, 1997; McClelland & Elman, 1986). Each component accepts the output from below as input, meaning that comprehending spoken words proceeds in many ways as the reverse of producing spoken words, going from input phonemes to lexical input forms to lemmas and finally to lexical concepts.

Existing literature has shown clear evidence for overlapping representations accessed during spoken word production and comprehension, with shared lexical concepts and lemmas (Glaser & Düngelhoff, 1984; Schriefers et al., 1990) and closely-linked phonological forms (Kittredge & Dell, 2016; Mitterer & Ernestus, 2008), as outlined in Figure 1.1 (adapted from Roelofs, 2014). This indicates individuals draw upon shared or similar representational codes in dual-tasking of speech production and comprehension. This further implies that the processing of background speech may require similar or the same representations that speech production relies on, which might interfere with production performance. Consistent with this proposal, a domain-specific *crosstalk* account (Pashler, 1994; or the *outcome conflict account*: Navon & Miller, 1987) for dual-task processing assumes that if two tasks use shared or similar representational codes at the same time, the representations can come into conflict, leading to impaired performance on one or both tasks. This account thus gives rise to an important question of how shared or similar representations affect speech production in the presence of background speech.

Although speech production is a highly practiced skill, there is clear evidence that it cannot be proceed without attention (for a review, see Roelofs & Piai, 2011). Attention is a limited capacity system consisting of three fundamental components: alerting, orienting, and executive control (Posner & Petersen, 1990; Posner & Rothbart, 2007), all of which play important roles in speech production. Existing work has shown that each process involved in speech production requires some form of attention (e.g., Ferreira & Pashler, 2002; Jongman et al., 2015; Mädebach et al., 2011; Roelofs, 2008). This means that when planning/producing speech, speakers must focus on speech production and shield against irrelevant background information (e.g., irrelevant background speech) that may cause distraction from production.

Prior research has also found that some processes of speech comprehension require attentional resources, including semantic activation (Lien et al., 2008) and at least some aspects of word-form encoding (Cleland et al., 2006; Lien et al., 2008). Moreover, the presence of background speech has been found to impair cognitive tasks (e.g., serial recall, Jones & Morris, 1992; reading, Cauchard et al., 2012) even when it is irrelevant to the focal tasks and should be ignored. This implies that irrelevant background speech may take up limited attentional resources available to speech production, and then disrupt production performance. In line with this proposal, a domain-general *capacity limitation* account for dual-tasking processing (Pashler, 1994; Ruthruff et al., 2003) assumes that when more attentional resources are required by one or both tasks, more interference in task performance should be observed. Hence, this account raises the question of how the amount of available attentional resources influences the cognitive processes involved in speaking in the presence of irrelevant background speech.

In short, speech production and comprehension share or have some similar representations, and each of them requires attentional resources to some degree. The domain-specific crosstalk and domain-general capacity limitation may play important roles in a typical language processing situation. That is, speaking in environments with verbal noise (e.g., in a restaurant or on a train) where speakers plan their speech while ignoring background speech may be affected by these phenomena.

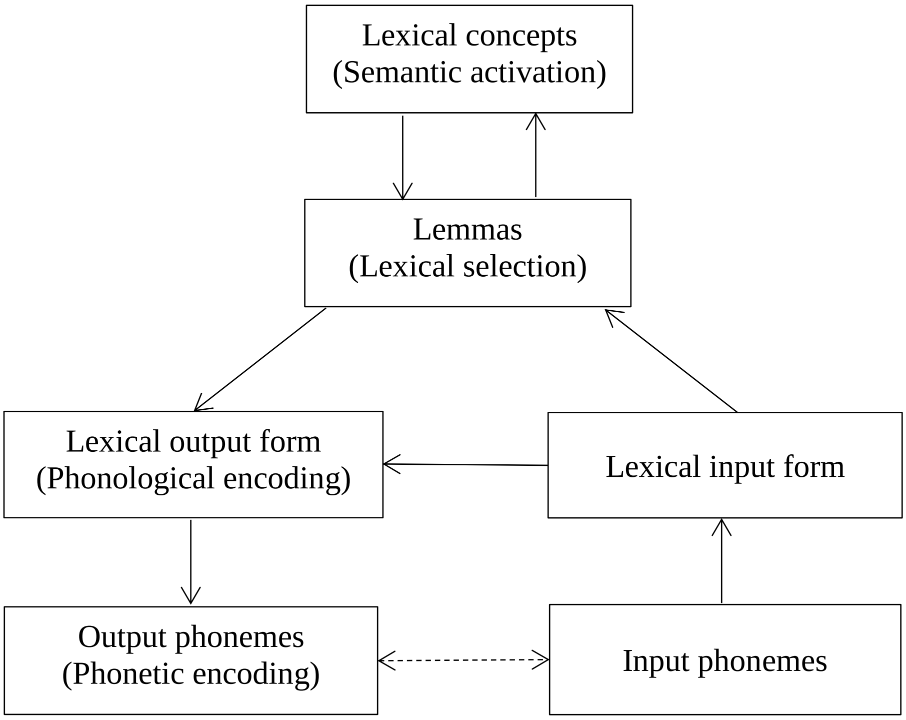

Figure 1.1. A working model of spoken word production and comprehension (adapted from Roelofs, 2014)

1.2 Irrelevant speech effects and relevant theories

A well-known effect relevant to the present dissertation is the *irrelevant speech effect* (or *irrelevant sound effect*; e.g., Colle & Welsh, 1976; Jones et al., 1992). It refers to the phenomenon that when cognitive tasks are performed in the presence of background speech or non-speech, task performance is impaired. This occurs even when background speech is irrelevant to the task and should be ignored. The irrelevant speech effect was originally observed in short-term memory studies using serial recall (Colle & Welsh, 1976; Salamé & Baddeley, 1982). In a typical serial recall study, a sequence of items (usually six to eight digits or letters) is to be maintained and recalled in the correct order. To test the aforementioned effect, this was done in the presence or absence of irrelevant background speech (e.g., Jones & Morris, 1992). Two qualitatively different auditory distraction effects were obtained: the *changing-state effect* and the *deviation effect.* The first effect refers to the finding that *changing-state* distractor sequences with acoustic changes between successive stimuli (e.g., ABABABAB) disrupt serial recall more than *steady-state* distractor sequences that are composed of a single repeated distractor stimulus (e.g., AAAAAAAA). The deviation effect occurs when expectations about the continuation of auditory distractor sequences are violated, such as a sequence with a single distractor stimulus deviating from a sequence of steady-state distractors (AAAAAAAB) (e.g., Hughes et al., 2007; Lange, 2005; Vachon et al., 2017). Correspondingly, two types of theories have been proposed to explain the two auditory distraction effects: the domain-specific *interference-by-similarity* account (e.g., Jones et al., 1993; Martin et al., 1988; Salamé & Baddeley, 1982, 1989) and the domain-general *attention capture* account (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015).

1.2.1 Interference-by-similarity account

The changing-state effect has been attributed to a conflict between the intentional processing of the to-be-remembered items' order and the obligatory and automatic processing of the irrelevant auditory distractors' order (e.g., Jones et al., 1993; Jones & Macken, 1993; Macken et al., 2009). In other words, the changing-state effect is caused by a conflict between similar, competing representations or processes, which is referred to as the domain-specific *interference-by-similarity* account (also see *interference-by-process* account; e.g., Hughes, 2014; Jones et al.,1993). This account is similar to the crosstalk account for dual-task processing (Pashler, 1994; *outcome conflict*: Navon & Miller, 1987), claiming that shared or similar representations or processes cause interference in task performance.

Interference-by-similarity has been extended to reading research, where reading performance has been shown to be impaired by irrelevant background speech relative to a quiet condition (Cauchard et al., 2012; Hyönä & Ekholm, 2016; Yan et al., 2018). Two specific views attribute this impairment to different sources: the *phonological disruption view* (Salamé & Baddeley, 1982, 1989) and the *semantic disruption view* (Martin et al., 1988). The phonological disruption view hypothesizes that the irrelevant speech effect in reading results from the similarity in the content of phonological codes of reading and background speech (Salamé & Baddeley, 1982, 1989). In contrast, the semantic disruption view assumes that it is a conflict of semantic processing that disrupts reading comprehension (Martin et al., 1988). Importantly, both views have received support from previous studies (e.g., Martin et al., 1988; Vasilev, Kirkby, et al., 2018; Vasilev, Liversedge, et al., 2019).

1.2.2 Attention capture account

In contrast, in the attention capture account (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015), the deviation effect is hypothesized to result from an entirely different process. That is, it is elicited by a violation of an expectation about the continuation of the auditory sequence. This violation of expectation is a necessary condition for attentional capture. When attention is captured by a deviant distractor, the amount of resources available for the primary memorization task is reduced, causing a drop in recall performance (e.g., Hughes et al., 2007; Lange, 2005; Vachon et al., 2017). This is referred to as the domain-general *attention capture* account, which assumes that irrelevant speech disrupts focal task performance because it diverts attention away from the task (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015). This account is compatible with the capacity limitation account for dual-task processing (Pashler, 1994; Ruthruff et al., 2003), assuming that the amount of attentional resources available to focal cognitive tasks decides task performance.

Furthermore, attention capture can be either *stimulus-aspecific* or *stimulus-specific* (Eimer et al., 1996). *Stimulus-aspecific attention capture* occurs when a background sound captures attention because it contrasts with its context (e.g., the sudden onset of speech following a period of silence; Eimer et al., 1996). *Stimulus-specific attention capture* occurs when the particular content of the background sound or speech diverts attention (e.g., when an individual's first name occurs in the stimuli; Eimer et al., 1996; Röer et al., 2013; Wood & Cowan, 1995). There is clear evidence from the irrelevant speech effect in serial recall and reading supporting the attention capture account (e.g., Bell et al., 2012; Cowan, 1995; Hyönä & Ekholm, 2016; Röer et al., 2014, 2015), but these studies do not make a further distinction between stimulus-aspecific and stimulus-specific attentional capture.

In sum, the interference-by-similarity account attributes the irrelevant speech effect to a domain-specific linguistic (e.g., phonological or semantic) similarity (e.g., Jones et al., 1993; Martin et al., 1988; Salamé & Baddeley, 1982, 1989), in contrast to the attention capture account, which attributes the effect to domain-general attentional capture (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015). These different theories, therefore, make different predictions for the influence of irrelevant background speech on speech production, as shown in Table 1.1.

Table 1.1. A summary of theories and the corresponding predictions.

Account	Content	Prediction
Interference-by-similarity account (e.g., Jones et al., 1993; Martin et al., 1988; Salamé & Baddeley, 1982, 1989)		
Phonological disruption view (Salamé & Baddeley, 1982, 1989)	Auditory disruption is caused by shared/similar phonological representations.	Any speech sound, regardless of its intelligibility, should disrupt speech production.
Semantic disruption view (Martin et al., 1988)	Auditory distraction results from shared/similar semantic representations.	Background speech should disrupt speech production only when it is intelligible.
Attention capture account (e.g., Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015)		
Aspecific attention capture view (Eimer et al., 1996)	Attention is captured when stimuli stand out against the context.	Acoustic variation of background speech suffices to disrupt speech production.
Specific attention capture view (Eimer et al., 1996)	Attention is captured by the content of the stimuli itself, such as linguistic information of background speech.	Specific linguistic content of background speech should disrupt speech production.

1.3 Modulation of focal task difficulty on the irrelevant speech effect

When facing auditory disruption caused by irrelevant background speech, speakers have to find a way to reduce the undesired processing of the irrelevant stimuli so that they can perform the cognitive task successfully. There is a body of work indicating that top-down factors, such as increased attention engagement in response to high task difficulty, modulate the processing of irrelevant background information (e.g., Halin et al., 2014; Hughes, 2014; Marsh et al., 2015). These studies have explored how the difficulty of focal cognitive tasks modulates the processing of background sounds, demonstrating that difficult focal tasks reduce or eliminate auditory disruption caused by background sounds via an *attention engagement* mechanism (also referred to as *task engagement*; see Halin et al., 2014; Marsh et al., 2015). That is, difficult focal tasks reduce auditory disruption because they promote a more steadfast locus of attention to the target stimuli. This, in turn, suppresses the attentional orienting response that background sounds may elicit, thus reducing the involuntary processing of background sounds that may interfere with the deliberate cognitive activity. The attention engagement mechanism reflects strategic control, mediated by both external factors, such as task difficulty, and by internal factors, such as motivation and working memory (Hughes et al., 2013; Sörqvist & Rönnberg, 2014). This idea has been extended to the research focusing on simple (e.g., memorizing items with background noise; Hughes et al., 2013; Marsh et al., 2015) and complex (e.g., proofreading with background speech; Halin et al., 2014) linguistic processing.

A similar but more specific account, the *load theory of attention,* has also been proposed to account for the modulation of focal task difficulty in the processing of irrelevant background information (Lavie, 2005; Lavie & Dalton, 2014). Specifically, Lavie and colleagues have explored how the processing of irrelevant information is affected by the load of the focal task in the context of cross-modality (visual and auditory) and within-modality (only visual modality) dual-tasking (Lavie et al., 2004, Lavie, 2005). They found that the processing of background distractors depended critically on the level and type of load involved in the processing of goal-relevant information in both cross-modality and within-modality contexts. That is, a high perceptual load on focal tasks consumes most of the processing resources, which results in reduced processing of irrelevant information relative to a low perceptual load. In contrast, a high load on cognitive control processes reduces the individual's ability to strategically prioritize the focal task and therefore leads to increased distractor processing. Combined, these accounts thus raise the question of whether and how the difficulty of speech production modulates the processing of background speech.

One way to modulate speech production difficulty is by manipulating *name agreement*. Name agreement refers to the extent to which participants agree on the name of a picture. Naming a picture with high name agreement (e.g., a picture of an *apple* is almost always called *apple*) is faster and more accurate than naming a picture with low name agreement (a picture of *a sofa* can also be a *couch, loveseat,* or *lounge*) (Alario et al., 2004; Shao et al., 2014; Vitkovitch & Tyrell, 1995). Following the attention engagement account (Halin et al., 2014; Marsh et al., 2015) and the load theory of attention (Lavie, 2005; Lavie & Dalton, 2014), name agreement should modulate the processing of background speech. The investigation of the modulation of name agreement on irrelevant speech effect would reveal whether speakers use top-down strategies to manage background information when focusing on their speech planning.

1.4 Current dissertation

As mentioned earlier, to understand how individuals manage to have smooth conversations in noisy environments, especially in situations with verbal noise, it is necessary to investigate how they comprehend and produce speech in the presence of background speech. Compared to the influence of background speech on comprehension, little work has investigated how speakers plan/produce utterances in the presence of background speech. This thus leaves the interesting question of whether and how irrelevant background speech interferes with speech production performance. The domain-specific interference-by-similarity account (e.g., Jones et al., 1993; Martin et al., 1988; Salamé & Baddeley, 1982, 1989) predicts that an increase in shared/similar linguistic representations (e.g., semantic and phonology) of background speech with speech production should cause more disruptions, while the domain-general attention capture account (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015) predicts that increased potential of attention capture by background speech should elicit more interference with speech production. In addition, given that the interference by background speech is modulated by an increase in focal task difficulty (e.g., Halin et al., 2014; Hughes, 2014; Marsh et al., 2015), the question arises whether and how speakers shield against interference by background speech when planning speech. The attention engagement account (Halin et al., 2014; Marsh et al., 2015) and the load theory of attention (Lavie, 2005; Lavie & Dalton, 2014) predict that the difficulty of speech production should modulate the interference by background speech. The answers to these questions would advance our understanding of the processes involved in speech production in the presence of background speech, which is relevant for real-life scenarios in which speakers need to ignore or filter out background verbal noise when they speak.

Thus, the present dissertation explored the cognitive mechanisms underlying linguistic dual-tasking, i.e., speaking in the presence of background speech, by focusing on the questions of how background speech influences speech production and what strategies speakers use to overcome this influence. Specifically, the first research question asks how irrelevant background speech affects speech production. To answer this question, this dissertation explored how the representational similarity and attention demand (Chapter 2), stimulus-aspecific (i.e., the presence/absence of pauses) and stimulus-specific (i.e., linguistic richness) variations (Chapter 4), and the interestingness and contextual variation (Chapter 5) of background speech influenced production performance. The second research question asks how speakers shield against disruption by irrelevant background speech. This question was explored by assessing whether the processing of irrelevant background speech was modulated by the difficulty of spoken word production (see Chapters 2, 4, and 5). These explorations provide insights into how domain-specific mechanisms (e.g., interference-by-similarity) and domain-general processing capacity (e.g., attention) are involved in the daily task of speaking in noisy environments.

Chapter 2, from the perspective of linguistic dual-tasking, explored how speech production was affected by concurrent listening (i.e., irrelevant background speech) with varied representational similarity (Experiment 1) and attentional demand (Experiment 2). Two lab-based experiments were conducted: Experiment 1 manipulated representational similarity between auditory stimuli and speech production. Here speakers named sets of pictures in Dutch while ignoring Dutch word lists (high similarity), Chinese word lists (medium similarity), or eight-talker babble (low similarity). Experiment 2 manipulated the attentional demand of auditory stimuli by asking participants to name sets of pictures in Dutch while ignoring Dutch speech (focused-attention condition) or listening to Dutch speech for a later memory task (divided-attention condition). The difficulty of lexical selection in production was manipulated by varying the name agreement (high, low) of pictures to-be-named in Dutch across two experiments. It was predicted that name agreement would modulate the disruption caused by increased representational similarity and attention demand of concurrent listening. The investigation thus assessed the roles of domain-specific representational similarity and domain-general attention limitation in the dual tasking of speaking while listening, and also provided insights into whether dual-tasking interference (i.e., auditory interference) can be diminished by attention engagement in speech production.

Because the COVID-19 pandemic impeded lab-based studies, a web-based study was conducted, which is described in Chapter 3. It highlights how speech production research can be done outside of the laboratory by measuring utterance duration and speech fluency in a multiple-object naming task. Two effects related to lexical selection were examined: name agreement (high, low) and semantic context (homogeneous, heterogeneous). A web-based modified blocked-cyclic naming paradigm was created, in which participants named a total of sixteen simultaneously presented pictures on each trial. This study predicted that the effects of name agreement and semantic context (i.e., it is slower and more error-prone to name the objects in a semantically homogeneous context than in a heterogeneous context) observed in lab-based studies should be replicated. This study thus provides information about the feasibility of conducting language production research in web-based settings.

Chapter 4 returned to the investigation of linguistic dual-task processing by focusing on irrelevant speech effects on speech production in online environment. Because the representational similarity effect in Chapter 2 could also be attributed to the stimulus-aspecific variation of auditory stimuli (i.e., background speech), such as acoustic variation (segmented vs. continuous sound stream), this chapter explored how stimulus-aspecific variation (i.e., the presence/absence of pauses) influenced speech production and whether the influence depended on the intelligibility of irrelevant background speech. The name agreement of to-be-named pictures was also manipulated to index lexical selection demand in spoken word production. Two web-based experiments were performed, where native Dutch speakers, who did not speak Chinese, named sets of pictures while ignoring different types of background speech: Experiment 1 explored whether stimulus-aspecific variation (i.e., the presence/absence of pauses) of unintelligible irrelevant speech affected picture naming performance by setting up three background speech conditions: Chinese word list, Chinese sentence, and a quiet control condition. Experiment 2 examined whether the irrelevant speech effects observed in Experiment 1 would be replicated when the background speech was intelligible by replacing Chinese speech with Dutch speech but keeping the quiet control condition. Different predictions were made in each experiment following the domain-specific interference-by-similarity and the domain-general attention capture accounts (see Chapter 4 for details). This exploration thus provided specific evidence for the two types of theories of irrelevant speech effects.

Chapter 4 used relatively boring and uniform background stimuli that may have led to adaptation. To explore whether this had affected the results, Chapter 5 investigated whether the interestingness (funny vs. boring background sentences) and contextual variation (blocks of boring sentences vs. varied blocks mixing boring and funny sentences), would affect speech production performance with a web-based experiment. Again, name agreement (high, low) of to-be-named pictures was also manipulated. Stable effects of interestingness and context were expected, and these effects should be reduced or eliminated when naming low name agreement pictures. The investigation provided insights into how speakers plan their speech and shield against the auditory disruption elicited by background speech in daily life.

Chapter 6 summarizes the key findings from the preceding chapters and discusses their broader theoretical and methodological implications, for example, how these findings can advance our understanding of how people plan their speech in noisy contexts, and how people manage the linguistic dual-tasking processing of speech production and comprehension in daily conversation. This chapter also provides recommendations for future directions within this research field.

2 | Concurrent listening affects speech planning and fluency: The roles of representational similarity and capacity limitation[1]

Abstract

In a novel continuous speaking-listening paradigm, we explored how speech planning was affected by concurrent listening. In Experiment 1, Dutch speakers named pictures with high versus low name agreement while ignoring Dutch speech, Chinese speech, or eight-talker babble. Both name agreement and type of auditory input influenced response timing and chunking, suggesting that representational similarity impacts lexical selection and the scope of advance planning in utterance generation. In Experiment 2, Dutch speakers named pictures with high or low name agreement while either ignoring Dutch words, or attending to them for a later memory test. Both name agreement and attention demand influenced response timing and chunking, suggesting that attention demand impacts lexical selection and the planned utterance units in each response. The study indicates that representational similarity and attention demand play important roles in linguistic dual-task interference, and the interference can be managed by adapting when and how to plan speech.

[1] Adapt from He, J., Meyer, A. S., & Brehm, L. (2021). Concurrent listening affects speech planning and fluency: the roles of representational similarity and capacity limitation. *Language, Cognition and Neuroscience*, *36*(10), 1258-1280. https://doi.org/10.1080/23273798.2021.1925130

2.1 Introduction

Despite conversation being one of the most common ways people communicate in daily life, relatively little experimental work has investigated how people manage to have smooth conversations with interlocutors. A characteristic of natural conversation is turn-taking, with interlocutors alternating between listening and speaking. Evidence from some studies of naturalistic conversation suggests that the gaps between turns are on average around 200 ms (Heldner & Edlund, 2010; Stivers et al., 2009), which shows that speakers do not respond to the partner's end of turn but begin to plan their utterances while listening. This means that conversation requires dual-tasking between speaking and listening (Levinson, 2016). It is known that dual-tasking causes interference in many psychological domains (e.g., Fischer & Plessow, 2015; Pashler, 1994; Strayer & Johnston, 2001), including in simple language tasks (e.g., Fairs et al., 2018; Fargier & Laganaro, 2016, 2019), but the role of dual-tasking in conversation is under-studied.

The present study extends research on linguistic dual-tasking to multi-word production using a novel speaking-listening paradigm in which participants were asked to name sets of six simultaneously shown pictures as quickly as possible while listening to speech. This allowed us to examine how overlapping linguistic representations and attention demand create interference in multi-word production, and to explore how speakers navigate this conflict by changing how they plan speech.

2.1.1 Sources of interference in linguistic dual-tasking

Two major accounts for interference in dual-tasking have been discussed in the literature, falling into the broad classes of domain-specific accounts (e.g., crosstalk) or domain-general accounts (e.g., capacity limitation). We walk through the predictions of both accounts for interference in linguistic dual-tasking below.

Domain-specific accounts of interference (e.g., crosstalk: Pashler, 1994; outcome conflict: Navon & Miller, 1987) suggest that if two tasks (e.g., visual perception and visual imagery) use similar representational codes at the same time, the representations can come into conflict, leading to impaired performance on one or both tasks (Bergen et al., 2007). This account therefore predicts that the degree of interference observed in a dual-task situation depends on the similarity or confusability of the mental representations involved in each task (Navon & Miller, 1987). In this paper, we use the term "representational similarity" to emphasize the role of shared representations between production and comprehension in eliciting interference.

Representational similarity could play a key role in linguistic dual-tasking since production and comprehension draw upon similar representations in the standard multi-stage model of psycholinguistics. In particular, there is clear evidence that representations for lexical concepts and lemmas are shared between production and comprehension. The best evidence for this is the semantic interference that arises in the picture-word interference (PWI) paradigm (Glaser & Düngelhoff, 1984; Schriefers et al., 1990). When naming a picture (e.g., DOG) with a spoken or written related distractor word (e.g., FOX), naming latencies are slowed and error rates increased compared to trials with an unrelated distractor (e.g., RANK; Damian & Martin, 1999; Schriefers et al., 1990). This suggests that there is competition between shared representations for concepts and lemmas across production (the target) and comprehension (the distractor; see Roelofs, 1992, 2003), and highlights the lemma level as an important origin of interference from comprehension on production.

Phonological representations for production and comprehension are also argued to be coupled (Kittredge & Dell, 2016; Mitterer & Ernestus, 2008). Evidence from the PWI paradigm has shown that in naming a picture (e.g., BED) a phonologically related distractor word (e.g., BEND) elicits less interference than an unrelated distractor (e.g., DUKE) (Damian & Martin, 1999; Schriefers et al., 1990). This suggests that comprehending a distractor word pre-activates phonological representations similar to the target, facilitating production when they are related. The implication is that if what is produced instead mismatches what is comprehended, pre-activation of phonological/phonetic representations could also elicit interference.

A representational similarity account of interference in linguistic dual-tasking predicts that a production task should receive more interference from a comprehension task than from a non-linguistic task, and that increased representational similarity between concurrent production and comprehension tasks should lead to increased interference. This prediction is supported by earlier work with the psychological refractory period (PRP) paradigm (e.g., Fairs et al., 2018), in which participants are tested on two discrete tasks (Task 1 and Task 2) and the onset of the Task 2 stimulus follows the onset of the Task 1 stimulus by varying intervals (referred to as stimulus onset asynchrony [SOA]). As the SOA decreases, Task 2 response latencies increase because of increasing task overlap. Performing a picture-naming task alongside syllable-identification results in more interference than performing the same task alongside tone-identification at various SOAs (Fairs et al., 2018). This extra interference occurs because the phonological representations activated by syllables are also used in picture naming. This work therefore demonstrates the importance of representational similarity in linguistic dual-tasking, but leaves open the question of how variation in the similarity of representations between comprehension and production might influence linguistic dual-tasking.

Domain-general accounts of interference suggest that capacity limitation (Pashler, 1994; Ruthruff et al., 2003) can hinder dual-task performance. Two prominent theories of this type have been proposed. The response selection bottleneck model (Pashler, 1994) assumes that performance on each task is staged, and while early and late stages can be processed in parallel, the central response selection stage can only operate on one task at a time, creating a bottleneck. By comparison, the capacity-sharing model (Kahneman, 1973; McLeod, 1977) assumes that even at the central response selection stage, information can be processed in parallel and that interference comes from dividing processing resources unequally, such that when more processing resources are devoted to one task or stimulus, fewer are left for other tasks (Tombu & Jolicœur, 2003). These theories share the general claim that people only have limited capacity or attentional resources to spread across tasks (Kahneman, 1973; Navon & Gopher, 1979). When more capacity is required by one or both tasks, more interference should be observed.

Capacity limitation may play an important role in linguistic dual-tasking because earlier work shows that language production and comprehension both require attention and because attention is required to suppress irrelevant speech input. To elaborate, all levels of language production seem to require attention. Earlier work showed that the amount of available processing resources constrains the cascade of activation from the conceptual to the lexical level in speech planning, suggesting that activating conceptual and lexical representations requires attentional resources (Mädebach et al., 2011). Lexical selection and phonological encoding are also hindered by linguistic dual tasking (Cook & Meyer, 2008; Ferreira & Pashler, 2002; Roelofs, 2008), and sustained attention (the ability to maintain alertness over time) is important for phonetic encoding in production (Jongman et al., 2015).

Some aspects of understanding spoken language also require attention, especially for processes above the word level (Kristensen et al., 2013; Moisala et al., 2015). However, early word recognition processes may occur with little attentional engagement (Dupoux et al., 2003). For instance, dichotic listening studies, where participants are asked to attend to one source of information (e.g., a female voice) while ignoring another source (e.g., a male speaker), have shown that the unattended speech is nonetheless processed to some extent (Aydelott et al., 2015; Dupoux et al., 2003; Rivenez et al., 2006; Rivenez et al., 2008). This means speakers' goals (e.g., attend to or ignore speech input) matter to the comprehension of auditory information. If the speech input is irrelevant, attention (especially executive control; Posner & Rothbart, 2006) is needed to suppress its processing and focus on target task (Dupoux et al., 2003). By contrast, if the speech input is relevant to speaker's goals, attention needs to be divided between processing the speech input and the target task. Therefore, in Experiment 1 we explored how speech planning was influenced by the representational similarity between the irrelevant auditory input and planned speech, and in Experiment 2 we contrasted speech planning when the speech input was relevant versus irrelevant to the speakers' goals.

2.1.2 Flexible planning units in multi-word production

To assess how representational similarity and capacity limitation impact linguistic dual-tasking and to expand on earlier work on interference between single-word production and comprehension (e.g., Fairs et al., 2018; Fargier & Laganaro, 2016, 2019), we designed a novel continuous speaking-listening paradigm. Dutch Participants were asked to name sets of six pictures using lists of nouns (e.g., *snoepje, troon, kasteel, viool, brievenbus, engel*; (*candy, throne, castle, violin, letterbox, angel*), while listening to a stream of linguistic information. This novel paradigm requires participants to retrieve the names of a set of simultaneously presented objects in quick succession and in the correct order, which means they must coordinate planning and articulating a series of words in the presence of the auditory input.

Naming a sequence of objects is different from single object naming because, in order to achieve fluency, speakers need to coordinate the planning and articulation of successive words with each other. Numerous eye tracking studies have shown how speakers usually achieve this: When several objects are to be named, speakers fixate upon them in the order of mention, and their eye gaze runs slightly ahead (by about 400 ms) of their speech (Belke & Meyer, 2007; Griffin & Bock, 2000; Sjerps & Meyer, 2015). In these studies, little processing of the objects can be done without directly fixating upon them, as they are spaced too far apart. This means that the visual-conceptual processing of the second object begins just before the first object name is initiated, and that the further encoding of the second object name happens while the first object name is pronounced. As a result of this tight coordination of word planning and articulation, speakers can name multiple objects fluently without long pauses between their names; this tight temporal coordination of speech planning and articulation requires processing capacity (Jongman et al., 2015). Alternatively, speakers can name sets of objects strictly sequentially, by only initiating the processing of an object after having fully planned and articulated the preceding object's name (Mortensen et al., 2008). This may lead to audible pauses between words. Combined, this means the planning units for multiple-word production can be flexible.

In order to explore whether and how the coordination of the planning and articulation of successive words was affected by the experimental variables, we determined how successive words were grouped into "chunks". We defined a chunk as any sequence of words without pauses of 200 ms or more between them, consistent with previous studies where an interval larger than 200 ms was coded as a silent pause in connected speech (e.g., Belke & Meyer, 2007; Campione & Véronis, 2002; Heldner & Edlund, 2010; Walker & Trimboli, 1982). We assumed that words within a chunk had been planned and coordinated tightly, as described above, with the planning of any following words overlapping with the articulation of the preceding word. By contrast, words separated by pauses had been planned more sequentially.

We quantified response chunking in two ways. The first was the total chunk number per trial, which refers to how many response chunks were produced in total for the six pictures. A perfectly fluent speaker would produce the six object names in one chunk (i.e. without any audible pauses), and a maximally disfluent speaker would produce them in six chunks (i.e. with a pause after each word). The second chunk measure was the first chunk length, which is defined as the number of words in the speaker's first response chunk. This measure is an indicator of the scope of advance planning before utterance onset, with a larger first chunk indicating a larger planned utterance unit. Note again that our view of response chunking does not imply that all words of a chunk are planned at the same time, rather that the planning of adjacent words overlaps enough to ensure that they can be produced without an intervening pause. We predicted that as the task became more demanding, the total chunk number should increase and the first chunk length should decrease. This could either be because that participants were less successful in coordinating the speech planning and articulation of successive words tightly when task demands were high, or because they chose to plan words with less temporal overlap.

2.1.3 Current study

We performed two experiments with the continuous speaking-listening paradigm, measuring interference in terms of overall picture naming accuracy, response timing (onset latency, speech duration), and response chunking (total chunk number, first chunk length). This provides a multi-faceted picture of what causes interference in linguistic dual-tasking, and what allows speakers to produce fluent speech regardless of interference.

In Experiment 1, we explored the role of representational similarity in linguistic dual-tasking. We manipulated representational similarity with three types of auditory stimuli (Dutch speech, Chinese speech, and eight-talker babble) that participants needed to ignore while naming pictures in Dutch. The irrelevant speech input is likely to cause interference in naming due to code conflict from shared representations since even unattended auditory information disrupts linguistic tasks such as semantic memory, reading, and writing (Marsh et al., 2008, 2009; Oswald et al., 2000; Sörqvist et al., 2012). In addition, increases in code conflict could lead to increases in the capacity required for language production because the unattended auditory words need to be suppressed. These influences are difficult to experimentally disentangle, and both reasons for interference are likely to be important in how representational similarity affects real-world conversations.

Whether because of code conflict or increased capacity demand for suppression, representational similarity is predicted to have a graded impact on interference in production. Auditory Dutch speech (*the high similarity condition*) overlaps with the representations used for production at multiple processing levels and should lead to increased capacity demand for their suppression and therefore to high levels of interference in production. In contrast, auditory Chinese speech (*the moderate similarity condition*) should only activate shared linguistic representations at the phonological/phonetic level, requiring less capacity for suppression and leading to less interference. We contrasted these conditions with a language-like noise condition (eight-talker babble), which was Dutch-like in its acoustic properties only (*the low similarity condition*) and should lead to low capacity demand for suppression. Note that the eight-talker babble was selected as the baseline condition because it is probably the most commonly occurring noise that interferes with speaking but without any linguistic information.

In Experiment 2, we emphasized the impact of capacity limitations on linguistic dual-tasking, which would reveal how much speech planning suffers when speakers attend more or less to their interlocutors. There were two conditions. The focused-attention condition was a replication of the Dutch listening condition in Experiment 1. In the divided-attention condition, participants listened to spoken Dutch words and had to recall whether a specific item was presented in the auditory stream after performing the production task. This is likely to increase the resources allocated to comprehension, and might also cause participants to more strongly activate competing linguistic representations during speech planning. Both of these properties of attention demand would lead to high levels of interference, and again, both reasons for interference are likely to be represented in real-world conversations. The prediction was that regardless of the source of interference, naming performance should be worse in the divided-attention condition than in the focused-attention condition.

In both experiments, we also varied the difficulty of the speech production task by asking participants to name pictures with high or low name agreement. Name agreement is the extent to which participants agree on the name of a picture. Some pictures consistently elicit the same name (e.g., *dog*; high name agreement), but others elicit two or more valid names (e.g., *sofa / couch*, low name agreement). There are other ways of varying the ease of lexical selection in picture naming, for instance, through the use of semantically related or unrelated distractors (e.g., Shao et al., 2013). We opted for varying name agreement because this does not require the use of further distractors in addition to the irrelevant speech and offers a better approximation to object naming in real-life contexts.

The two most common reasons for poor name agreement are that the depicted objects are hard to identify (e.g., a line drawing of a *celery*, commonly misidentified as *rhubarb*) or that the objects have several plausible names (e.g., *sofa* and *couch*; Alario et al., 2004; Vitkovitch & Tyrrell, 1995). Thus, name agreement effects can originate during the visual-conceptual processing of the pictures or the retrieval of their names. We selected pictures that could be easily identified but had multiple names. The long naming latencies associated with these low name agreement items have been attributed to competition among alternative names, which has to be resolved during lexical selection (Alario et al., 2004; Shao et al., 2014). This means that naming low name agreement pictures not only co-activates multiple lemmas, but also requires more processing capacity (e.g., executive control) to inhibit lemma competitors and select the target names.

We predicted that, following earlier work, pictures with low name agreement would be named more slowly than those with high name agreement. More importantly, we also predicted this name agreement effect would interact with representational similarity in Experiment 1: As producing low name agreement pictures involves more competition between lexical candidates and requires more capacity, low name agreement pictures should be more strongly affected by representational similarity than high name agreement pictures. We predicted a similar pattern for the effect of attention demand in Experiment 2: Asking participants to divide their attention between speaking and listening (rather than focusing on speaking alone) should have a stronger impact on pictures with low than high name agreement.

2.2 Experiment 1

To examine the role of shared representations in linguistic dual-tasking, we manipulated representational similarity in a continuous speaking-listening paradigm using three auditory conditions: Dutch speech (high similarity), Chinese speech (moderate similarity), and eight-talker babble (low similarity). We predicted that more interference would be observed as similarity increased. We also manipulated the difficulty of lexical selection in production by varying the name agreement (high, low) of the pictures to be named. We predicted that naming performance would be worse for low name agreement pictures than high name agreement pictures. We predicted an interaction between the two factors, such that a stronger representational similarity effect would be observed for low name agreement pictures than high name agreement pictures. This is because low name agreement pictures elicit more candidate lemmas and therefore require more executive control to select and produce a specific name, which would create more potential conflict with comprehension.

2.2.1 Method

Participants

We recruited 21 native Dutch speakers who had no Chinese experience (16 females) from the Max Planck Institute for Psycholinguistics' database. This sample size was selected because power simulations[2] showed that 20 participants and 126 items would allow 99% power to measure a plausibly small interaction between name agreement and similarity on the onset latency measure. The interaction effect size used in the simulations was a name agreement effect of 50 ms or smaller (SD = 100 ms) in the eight-talker babble and Chinese conditions, but 100 ms or larger (SD = 100 ms) in the Dutch condition. All participants were university students with a mean age of 22 years (range: 19-26) and reported normal or corrected-to-normal

[2] After conducting the experiment, we had a smaller effective sample size than originally anticipated due to many excluded incorrect trials. Further simulations using 21 participants and 84 items (2/3 of the original item number) suggested that 21 participants should still lead to 98% power to observe an interaction where the name agreement effect was 40 ms (100 ms sd) in the eight-talker babble and Chinese listening conditions, and 80 ms (100 ms sd) in the Dutch listening condition.

vision as well as no speech or hearing problems. They provided informed consent and received a payment of 6 € for their participation. The study was approved by the ethics board of the Faculty of Social Sciences of Radboud University.

Apparatus

The experiment was controlled by a desktop computer with Presentation software (Neurobehavioral systems). Auditory stimuli were presented using Sennheiser HD 280-13 headphones. Participants' speech was recorded by using a Sennheiser ME 64 microphone and a digital voice recorder. WebMAUS Basic was used to calculate phonetic segmentation and labels for participants' speech responses (https://clarin.phonetik.unimuenchen.de/BASWebSe-rvices/interface/WebMAUSBasic). Praat software (Boersma & Weenink, 2009) was then used to extract the onsets and offsets of all segmented responses.

Materials

Visual stimuli. 252 pictures (see Appendix A, Table A1) were selected from the MultiPic database of 750 single-object drawings (Duñabeitia et al., 2018), which provides language norms in standard Dutch. Of these, 126 were high name agreement pictures, all with a name agreement percentage of 100%, and 126 were low name agreement pictures, with a name agreement percentage between 50% and 87% (M = 73%, SD = 11%). Independent *t*-tests revealed that the two sets of items differed significantly in name agreement, but not in any of the following 10 psycholinguistic attributes: visual complexity, Age-of-Acquisition (AoA), word frequency (WF), number of phonemes, number of syllables, word prevalence, phonological neighborhood frequency (PNF), phonological neighborhood size (PNS), orthographic neighborhood frequency (ONF), and orthographic neighborhood size (ONS).

The 126 high name agreement and 126 low name agreement pictures were each divided into three subsets and paired with the three auditory conditions (Dutch speech, Chinese speech, eight-talker Babble), meaning that each auditory condition was paired with 42 high name agreement and 42 low name agreement pictures. The high name agreement and low name agreement sets of pictures assigned to each auditory condition were also matched on the above-mentioned 10 attributes.

On each trial of the experiment, six pictures, all with high name agreement or all with low name agreement, were presented simultaneously in a 2 × 3 grid (size: 20cm × 30 cm). The pictures per grid were neither semantically related (i.e. they were from different semantic categories) nor phonologically related (i.e. avoiding the overlap of their 1st phonemes), as judged by a native speaker of Dutch. There were 14 grids for each set of pictures resulting in 42 grids in total. In addition, 36 additional pictures (6 grids) were selected from the same database as practice stimuli.

Auditory Stimuli. For the Dutch speech condition, 252 additional nouns (see Appendix A, Table A2) were selected from the MultiPic database. To pair with the set of 14 picture grids, these 252 Dutch nouns were divided into 14 word lists of 18 nouns. All 14 lists were matched on AoA, WF, number of phonemes, number of syllables and word prevalence. The above-mentioned five lexical variables were also matched between the Dutch nouns in the word lists and the sets of pictures to be named. We estimated that participants would name one picture within the time-span of three auditory words, which was approximately two seconds. This is because naming latencies for pictures can be around one second (e.g., Vitkovitch & Tyrrell, 1995; Shao et al., 2014), the spoken duration (the difference from speech onset and offset of a word) of a one- or two-syllable word may be up to 500 ms (e.g., Damian, 2003), and both utterance onset and articulation may be slowed in dual-tasks contexts. Therefore, to equate the amount of semantic and phonological overlap across trials between planning and listening, we designed the item lists so that any three consecutive Dutch nouns in the auditory condition were neither semantically nor phonologically related to each other, nor to the to-be-named pictures in the same ordinal position, as judged by a native speaker of Dutch. To create practice stimuli, 36 additional Dutch nouns were also selected from the same database to make two word lists. All of the word lists were recorded by a female native Dutch speaker in neutral prosody using Audacity software (http://audacity.sourceforge.net/) at a sample rate of 44100 Hz. Each list was then further processed using Adobe Audition (https://www.adobe.com/products/audition.-html) and Praat to make an audio file lasting 12 seconds by deleting initial and final silences as well as stretching by up to 2.19% or compressing by up to 1.46%.

The Chinese speech lists (see Appendix A, Table A3) were translated from 16 Dutch word lists; items were selected such that the total number of syllables in the Chinese words was matched across lists. The order of nouns in each word list was set again so that no three consecutive Chinese nouns were phonologically related to each other, nor to any Dutch pictures in the same ordinal position. A female native Mandarin Chinese speaker recorded these word lists which were further edited in the same fashion as the Dutch speech to last 12 seconds each.

The eight-talker babble condition was created from a set of 20 semantically anomalous Dutch sentences (see Appendix A, Table A4) based on Smiljanić and Bradlow (2011). Each sentence had an average of eight words (range: six to ten). Babble was made from recordings of eight female native speakers of Dutch between 22 and 30 years old who spoke each sentence in clear, conversational speech. As in Van Engen and Bradlow (2007), four different sentences from each talker were concatenated to create a sound file lasting 12 seconds. A multiple of 100 ms of silence was added to each talker's file (0-700 ms) in order to stagger the talkers once they were mixed together. All eight talkers were then mixed, and the initial 700 ms of the mixed file was removed to eliminate the part of the file that did not contain all eight talkers. The first 100 ms of the completed noise file was faded in. A set of sixteen eight-talker babbles was made; fourteen were used as experimental stimuli and two were used as practical stimuli. All auditory files were matched on intensity (80dB) in Praat.

Design

Representational similarity (Similarity: Dutch speech, Chinese speech, eight-talker babble) and the difficulty of lexical selection in planning (Name agreement: high, low) were both treated as within participant variables; both factors were randomized within experimental blocks and counterbalanced across participants. Items were repeated three times resulting in three blocks each containing 42 trials with one repetition of each auditory condition and each picture grid. Across blocks, the same set of six pictures was paired with all three auditory conditions, and the pictures were presented in a different arrangement within each repetition. Across all participants, the order of trials was randomized with Mix program (van Casteren & Davis, 2006).

Procedure

Participants were tested individually in a soundproof room. A practice session of six trials was followed by the three blocks of experimental trials. Participants took a short break after each block. The whole experiment lasted 30 minutes.

Trials began with a fixation cross presented for 500 ms, followed by a blank screen for 300 ms. Then, a 2 × 3 grid appeared on the screen in which six pictures were presented while a sound file played for up to 12 seconds. Participants named the six pictures one by one in order (first row, followed by second row) as quickly and accurately as possible while ignoring the auditory information. Once finished, they pressed a button to end the trial, at which point a blank screen was presented for 1500 ms.

Analysis

Five dependent measures were coded to index interference in naming. Production *accuracy* indexed the proportion of trials where all six items were named with the correct responses. Picture names were coded as correct if they matched the first or second most common names given to the picture in the MultiPic database (Duñabeitia et al., 2018)[3], were synonymous to one of the two most common names (e.g., *laboratorium / lab*), or contained a diminutive version of one of the two most common names (e.g., *munt / muntje*), as judged by trained research assistants.

For trials where all pictures were named correctly and which contained no hesitations or auto-corrections (hereafter, “fully correct trials”), we calculated two timing measures. *Onset latency* was defined as the time from stimulus onset to the first picture name onset. This reflects how long participants take to plan their speech before articulation, indexing the very beginning stages of speech planning. *Speech duration* was defined as the time between the onset of the first picture name and the offset of the sixth picture name. This reflects how long participants take to produce all stages of speech. These measures were both log-transformed because they were right skewed.

For these fully correct trials, we also examined how participants chunked or grouped their six responses. As described earlier, we coded responses that occurred with 200 ms or less between them as a single response chunk, as previous studies of spontaneous speech code durations larger than 200 ms as a silent pause (e.g., Campione & Véronis, 2002; Heldner, & Edlund, 2010; Walker & Trimboli, 1982). Two dependent measures were derived from this. *Total chunk number* refers to how many response chunks participants made in total, with a larger number of total response chunks meaning more separate planning units for production. *First chunk length* refers to how many names participants produced in their initial response, and illustrates how much information participants planned before starting to speak.

Accuracy, log-transformed onset latency, and log-transformed speech duration were analyzed with mixed-effect models implemented using the *lme4* package (Bates et al., 2015) in R version 3.6.1 (R Core Team, 2018). Predictors were name agreement (high NA / low NA) and representational similarity (Dutch / Chinese / Babble). Name agreement (high NA / low

[3] We also coded naming responses strictly such that only the first common names for the pictures given by MultiPic database (Duñabeitia et al., 2018) were correct. We found there were too few fully correct trials in the most difficult conditions in both experiments (51 trials in the low NA & Dutch condition for Exp1, 51 trials in the low NA & divided-attention condition for Exp2). Thus, we did not conduct the same version of analyses for these data.

NA) was contrast coded with (0.5, –0.5). For similarity, the first contrast was coded with (0.25, 0.25, –0.5) and compared the two language conditions (Dutch and Chinese speech) to language-like noise condition (eight-talker babble), and the second contrast was coded with (0.5, –0.5, 0) and compared Dutch with Chinese speech. The random effect structure in all models included random intercepts for participants and items. No random slopes were included because of convergence issues or evidence of model overfitting (high correlations between random terms). For the dependent measure of accuracy, a logistic mixed-effect model was fitted because of the binary nature of the responses. For the timing measures, separate linear mixed-effect models were fitted.

Because of the discrete nature of the total chunk number and first chunk length, these measures were analyzed with ordinal mixed models using the *clmm (culmulative link mixed model)* function in the package *ordinal* in R version 3.6.1 (R Core Team, 2018). The predictors were name agreement and representational similarity, contrast-coded as described above. The random effect structure in all models again only included random intercepts for participants and items.

We also conducted an additional set of analyses on a larger dataset which included all trials where participants made correct responses on the first picture, though the other pictures were not necessarily named correctly. This was done to test whether the analyses were underpowered due to the high error rates in some conditions. The results were largely comparable to the main analyses and are therefore only reported in Appendix B (see Table B1).

2.2.2 Results

Naming accuracy. Participants produced the intended responses on 65% of the naming trials. As shown in Tables 2.1 and 2.2, accuracy for high name agreement pictures was considerably higher than for low name agreement pictures ($\beta = 2.12$, $SE = 0.22$, $p < 0.001$), but did not vary by representational similarity. Name agreement and representational similarity did not interact.

Onset latency. As shown in Figure 2.1 (left), log-transformed onset latency was affected by name agreement and representational similarity. As supported by a linear mixed-effect model (see Table 2.2), it took participants reliably longer to plan names for low name agreement pictures than high name agreement pictures ($\beta = -0.12$, $SE = 0.03$, $p < 0.001$). Log-transformed onset latencies in the two language conditions (Dutch and Chinese) were reliably slower than in the eight-talker babble condition ($\beta = 0.15$, $SE = 0.02$, $p < 0.001$), and log-transformed onset latencies were reliably slower in the Dutch speech than Chinese speech conditions ($\beta = 0.09$, SE = 0.02, $p < 0.001$). Name agreement and representational similarity interacted on the first contrast ($\beta = 0.13$, SE = 0.05, $p < 0.01$), showing that log-transformed onset latencies in the two language conditions were slower than in the eight-talker babble condition for high name agreement pictures ($\beta = 0.08$, SE = 0.01, $p < 0.001$), but this difference was not observed for low name agreement pictures.

Table 2.1. Dependent measures in Experiment 1 by name agreement and representational similarity.

	High name agreement			Low name agreement		
	Dutch	Chinese	Babble	Dutch	Chinese	Babble
Accuracy (%)	81 (57-95)	84 (43-100)	86 (57-100)	44 (19-67)	46 (19-76)	45 (19-76)
Onset latencies (ms)	1231 (577)	1101 (495)	973 (378)	1332 (582)	1231 (546)	1184 (427)
Speech durations (ms)	5295 (1453)	4732 (1206)	4673 (1236)	5963 (1690)	5593 (1433)	5544 (1499)
Total chunk number	3.1(1.4)	2.8(1.4)	2.6(1.4)	3.5(1.5)	3.5(1.5)	3.2(1.5)
First chunk length	2.(1.7)	3.1(2.0)	3.4(1.9)	2.5(1.5)	2.3(1.7)	2.6(1.8)

Note. All timing and chunking measures reflect fully correct trials only. For accuracy, range follows in parentheses, for other measures, standard deviation follows in parentheses.

Speech duration. As shown in Figure 2.1 (right), log-transformed speech duration was affected by name agreement and representational similarity. As supported by a linear mixed-effect model (see Table 2.2), log-transformed speech duration was reliably longer for low name agreement pictures than high name agreement pictures ($\beta = -0.13$, SE = 0.02, $p < 0.001$). Log-transformed speech durations in the two language conditions (Dutch and Chinese) were reliably longer than in the eight-talker babble condition ($\beta = 0.08$, SE = 0.02, $p < 0.001$), and log-transformed speech duration was reliably longer in the Dutch speech than Chinese speech condition ($\beta = 0.08$, SE = 0.01, $p < 0.001$). Name agreement and representational similarity did not interact[4].

Total chunk number. As shown in Figure 2.2 (left) and Table 2.1, total chunk number was affected by name agreement and representational similarity. As supported by an ordinal mixed model (see Table 2.2), participants grouped their responses in more small chunks for low name agreement pictures than high name agreement pictures (β = –0.09, SE = 0.02, $p < 0.001$). Total chunk number was greater in the two language conditions (Dutch and Chinese) than in the eight-talker babble condition (β = 0.37, SE = 0.11, $p < 0.001$), but no difference between the Dutch and Chinese conditions was observed. Name agreement and representational similarity did not interact.

First chunk length. As shown in Figure 2.2 (right) and Table 2.1, first chunk length was affected by name agreement and representational similarity. As supported by an ordinal mixed model (see Table 2.2), participants planned, on average, fewer names in their first response chunk for low name agreement pictures than high name agreement pictures (β = 0.35, SE = 0.07, $p < 0.001$), as they made fewer responses with maximal first chunks (i.e. chunk length = 6) in the low name agreement than in the high name agreement conditions (see Figure 2.2 (right)). The first chunk length for pictures in the two language conditions (Dutch and Chinese) was shorter, on average, than in the eight-talker babble condition (β = –0.32, SE = 0.08, $p < 0.001$). Collapsed across name agreement, participants made more responses with minimal first chunks (i.e. chunk length = 1) and fewer responses with maximal first chunks (chunk length = 6) in the language conditions than in the babble condition (see Figure 2.2 (right)). There was no difference in first chunk length for the Dutch and Chinese speech conditions. However, name agreement and representational similarity did interact on the second contrast (β = –0.45, SE = 0.14, $p < 0.001$), which showed that while there was no main effect of Dutch versus Chinese speech, this main effect was qualified by name agreement such that participants produced more names in their first response chunk in the Dutch speech than in the Chinese speech conditions for high name agreement pictures (β = –0.21, SE = 0.08, $p < 0.05$) but not for low name

[4] To explore planning done between producing chunks of words, a linear mixed-effect model was also fitted on the measure of log-transformed total pause time. Total pause time was defined as the sum of all within-utterance pauses with minimal durations of 200 ms. The results for this variable patterned in the same way as speech duration. Log-transformed total pause time was affected by name agreement (β = –0.67, SE = 0.18, $p < 0.001$) and representational similarity (language conditions vs. language-like noise condition (β = 0.75, SE = 0.19, $p < 0.001$); and Dutch speech vs. Chinese speech (β = 0.34, SE = 0.16, $p < 0.05$), with no reliable interactions.

agreement pictures.

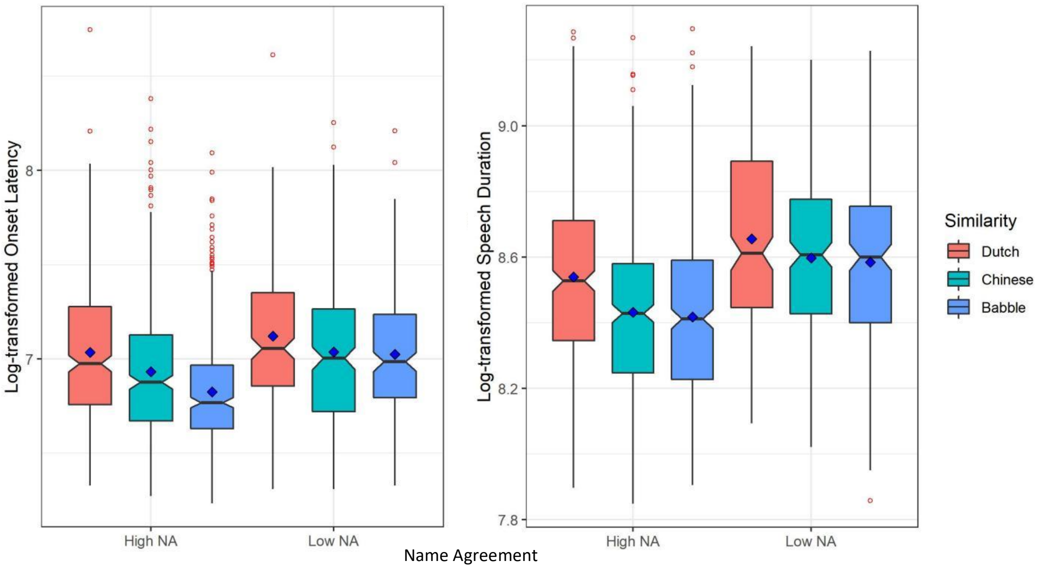

Figure 2.1. Log-transformed onset latencies (left) and log-transformed speech durations (right) in Experiment 1 split by representational similarity (Dutch speech, Chinese speech, eight-talker babble) and name agreement (NA; high, low). Blue squares represent condition means and red points reflect outliers. All measures reflect fully correct naming trials only.

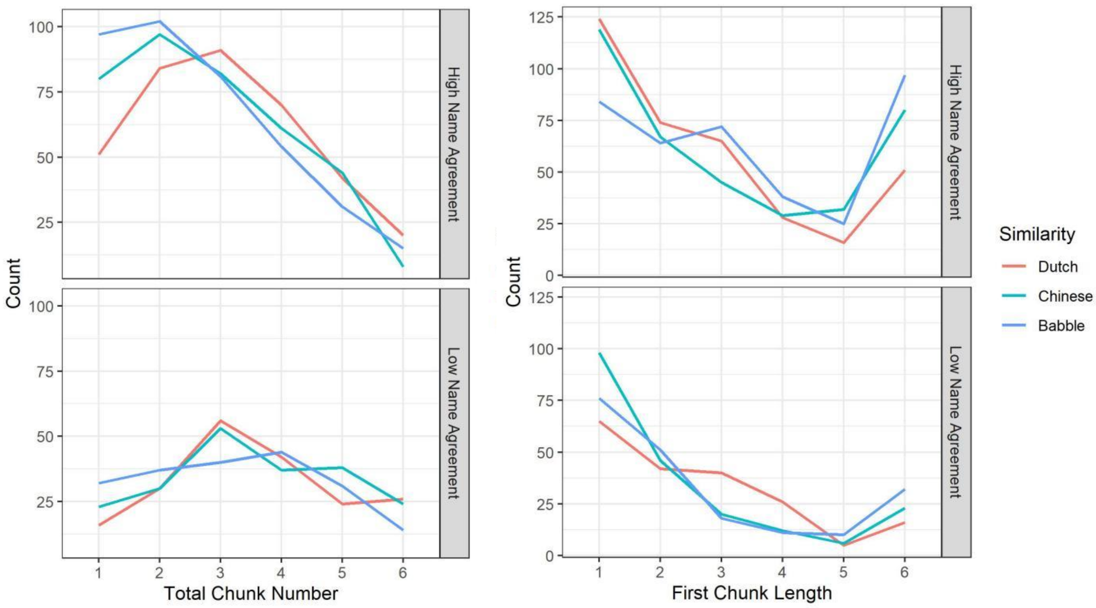

Figure 2.2. Total chunk number (left) and first chunk length (right) in Experiment 1 split by representational similarity (Dutch speech, Chinese speech, eight-talker babble) and name agreement (high, low). All measures reflect fully correct naming trials only.

Table 2.2. Mixed-effect models for log-transformed onset latencies (Log-Onset), log-transformed speech durations (Log-Duration), accuracy, and chunk measures in Experiment 1.

	Fixed effects	Estimate	SE	*t* value	*p*
Log-Onset	Intercept	7.00	0.04	197.934	< 0.001
	NA (High vs. Low)	-0.12	0.03	-4.525	< 0.001
	Similarity ((Dutch & Chinese) vs. Babble)	0.15	0.02	6.722	< 0.001
	Similarity (Dutch vs. Chinese)	0.09	0.02	4.626	< 0.001
	NA×Similarity ((Dutch & Chinese) vs. Babble)	0.13	0.05	2.826	< 0.01
	NA×Similarity (Dutch vs. Chinese)	0.03	0.04	0.711	0.477
Log-Duration	Intercept	8.54	0.03	302.136	< 0.001
	NA (High vs. Low)	-0.13	0.02	-6.599	< 0.001
	Similarity ((Dutch & Chinese) vs. Babble)	0.08	0.02	4.792	< 0.001
	Similarity (Dutch vs. Chinese)	0.08	0.01	5.827	< 0.001
	NA×Similarity ((Dutch & Chinese) vs. Babble)	0.03	0.03	1.041	0.298
	NA×Similarity (Dutch vs. Chinese)	0.04	0.03	1.586	0.113
	Fixed effects	Estimate	SE	*z* value	*p*
Accuracy	Intercept	0.83	0.15	5.411	< 0.001
	NA (High vs. Low)	2.12	0.22	9.771	< 0.001
	Similarity ((Dutch & Chinese) vs. Babble)	-0.18	0.14	-1.272	0.203
	Similarity (Dutch vs. Chinese)	-0.18	0.12	-1.530	0.126
	NA×Similarity ((Dutch & Chinese) vs. Babble)	-0.40	0.28	-1.415	0.157
	NA×Similarity (Dutch vs. Chinese)	-0.14	0.23	-0.581	0.561
Total chunk number	NA (High vs. Low)	-0.09	0.02	-5.904	< 0.001
	Similarity ((Dutch & Chinese) vs. Babble)	0.37	0.11	3.344	0.001
	Similarity (Dutch vs. Chinese)	0.02	0.10	0.198	0.843
	NA×Similarity ((Dutch & Chinese) vs. Babble)	-0.01	0.03	-0.285	0.775
	NA×Similarity (Dutch vs. Chinese)	0.04	0.02	1.888	0.059

First chunk length	NA (High vs. Low)	0.35	0.07	4.825	< 0.001
	Similarity ((Dutch & Chinese) vs. Babble)	-0.32	0.08	-4.106	< 0.001
	Similarity (Dutch vs. Chinese)	0.01	0.07	0.191	0.848
	NA×Similarity ((Dutch & Chinese) vs. Babble)	-0.21	0.16	-1.322	0.186
	NA×Similarity (Dutch vs. Chinese)	-0.45	0.14	-3.302	<0.001

Note. All measures reflect fully correct naming trials only. NA refers to name agreement, similarity refers to representational similarity.

Trials with correct first responses. For the larger dataset using all responses where at least the first picture name was produced accurately (see Appendix B, Table B1), one additional interaction was found on total chunk number, such that the representational similarity effect (Dutch vs. Chinese) was larger for high name agreement pictures than for low name agreement pictures.

2.2.3 Discussion

This experiment was designed to test how representational similarity impacted linguistic dual-task interference. Representational similarity had large effects on naming performance: we found differences between linguistic (Dutch and Chinese) and language-like noise (eight-talker babble) listening conditions on all measures except accuracy, and a difference between the two language conditions (Dutch and Chinese) on onset latency and speech duration. These results indicate that increased overlap in representations between simultaneous planning and listening leads to increased interference because of heightened code conflict, consistent with earlier work (e.g., Fairs et al., 2018; Fargier & Laganaro, 2016). This provides evidence that representational similarity plays an important role in simultaneous speaking and listening.

While representational similarity certainly affected the degree of overlapping representations recruited for speech planning and listening, it might also have affected attention demand because native language words might capture attention more effectively than non-native words or multi-talker babble. Hence, more attention may have been needed to suppress the Dutch input than the Chinese or eight-talker babble, which in turn affected the processing resources available for speech planning. This means that we cannot solely attribute the effects of representational similarity to domain-specific sources of interference; instead depletion of attention may also have played a role. Both factors are likely to play important roles in real-world conversations.

We also manipulated name agreement, a production-internal source of difficulty. This affected all five dependent measures, showing that speakers were less accurate, took longer to plan names for pictures with low name agreement, and produced fewer picture names at a time than pictures with high name agreement. This is consistent with name agreement effects in earlier work (e.g., Alario et al., 2004; Shao et al., 2014; Vitkovitch & Tyrrell, 1995).

Evidence for interaction between name agreement and representational similarity appeared on onset latency, showing that participants took more time to plan before articulation for high name agreement pictures in the language conditions than in the babble condition. The interaction was also found on the first chunk length, showing that participants reduced the scope of advance planning in utterance generation for high name agreement pictures in the Dutch speech condition than in the Chinese speech condition. The results suggest that representational similarity influences lexical selection in production. Note that this pattern opposes our prediction that greater representational similarity effects should be found for low name agreement pictures than high name agreement pictures. This may be because planning difficult picture names requires speakers to concentrate harder, making their locus of attention more steadfast and causing them to process the background information less (Halin et al., 2014; Halin et al., 2015). This attention enhancement mechanism might diminish the effects of representational similarity for low name agreement pictures. We discuss this further in the General Discussion.

To further explore the role of attention in concurrent speech planning while listening and to disclose how capacity limitation contributes to linguistic dual-task interference, Experiment 2 manipulated name agreement alongside the attention demand of comprehension. Varying how much attention is allocated to comprehension might also cause participants to more or less strongly activate a set of linguistic representations that can then cause competition during planning. The implication in either case is interference in production, whether from domain-general or domain-specific sources.

2.3 Experiment 2

In this experiment, we manipulated the attention demand of comprehension by asking participants to name pictures in Dutch while either ignoring Dutch speech (focused-attention condition) or trying to remember the Dutch words for a later memory test (divided-attention condition). Consistent with the capacity limitation account of interference in linguistic dual-tasking, we predicted that more interference should be observed in the divided-attention condition than in the focused-attention condition. To assess the role of attention demand in lexical selection, we also varied the name agreement (high, low) of to-be-named pictures. We predicted an interaction between attention demand and name agreement, such that a stronger effect of attention demand would be observed for low name agreement pictures than high name agreement pictures. This is because low name agreement pictures activate multiple target names, and attention is required to select among them. This is not the case for high name agreement pictures, which only activate one dominant name. Thus, the additional attentional load should affect naming more in the low than in the high name agreement conditions.

2.3.1 Method

Participants

We recruited 40 native Dutch speakers (31 females, M_{age} = 22 years, range: 18 - 29 years) from the Max Planck Institute for Psycholinguistics' database. This sample size was selected based on power simulations which showed that 40 participants and 24 items (allowing for trial inclusion rates of up to 60% of the total item number) would allow observation at 97% power to measure a plausibly-sized interaction between attention demand and name agreement on the onset latency measure. The interaction effect size used in these simulations involved a name agreement effect of 50 ms or smaller (SD = 100 ms) in the focused-attention condition, but 100 ms or larger (SD = 100 ms) in the divided-attention condition. All participants reported normal or corrected-to-normal vision as well as no speech or hearing problems. They signed an informed consent and received a payment of 6 € for their participation. The study was approved by the ethics board of the Faculty of Social Sciences of Radboud University.

Apparatus

The same apparatus was used as in Experiment 1.

Materials

Visual stimuli. A subset of the pictures (40 of the original 42 picture grids) from Experiment 1 was selected to yield 120 high name agreement items (100%) and 120 low name agreement items (50% - 87%). Independent *t*-tests revealed that the two sets of items differed significantly in name agreement, but not in any of the 10 psycholinguistic attributes described in Experiment

1 (i.e. visual complexity, AoA, WF, number of phonemes, number of syllables, word prevalence, PNF, PNS, ONF, and ONS). These pictures were divided into two subsets for the two blocks; both subsets were matched on all above-mentioned 10 properties including name agreement.

Trials were set up as in Experiment 1, with six pictures in a 2 × 3 grid (20 cm ×30 cm) that were neither semantically nor phonologically related. There were 20 picture grids per block, resulting in 40 trials in total, plus eight practice trials (containing 48 additional pictures), four presented before each experimental block.

Auditory Dutch Speech. We created 40 lists of Dutch nouns to pair with the 40 picture grids. These were comprised of the 14 lists of Dutch nouns (252 nouns) from Experiment 1 and 26 more lists made from 468 additional nouns (see Appendix C, Table C1) that were selected from the MultiPic database (Duñabeitia et al., 2018) and the Dutch Lexicon Project 2 (Brysbaert et al., 2016) in order to provide Dutch auditory stimuli for all trials with no repetition. All 40 lists were matched on five psycholinguistic variables: AoA, WF, number of phonemes, number of syllables, and word prevalence. The 40 lists were then divided into two subsets for the two blocks (360 Dutch nouns in each) matched on the same above-mentioned five variables. Items were arranged to avoid semantic and phonological overlap in the same way as described in Experiment 1. The 40 picture grids and 40 word lists were paired in a fixed way to make up trials that were presented in a unique random order for each participant. Finally, 110 additional Dutch nouns were also selected from the same database to make 8 word lists for practice trials.

All of the 48 word lists were recorded by a female native Dutch speaker in neutral prosody[5]. As in Experiment 1, each list was then edited to make an audio file lasting 12 seconds by deleting initial and final silences and compressing the trial duration by a small amount if necessary (up to 9.5%). All auditory files were also matched on intensity (80dB) using Praat.

Memory Task. To create the memory task used in the focused-attention blocks, 40 target words appearing in the 4th to 13th position in each word list were selected, corresponding to the hypothesized interval in which the participant would be speaking. An additional 40 foil words were selected from the Dutch Lexicon Project 2 (Brysbaert et al., 2016) to be used in invalid trials; these items did not appear in any word list. Items presented in valid and invalid trials were also matched on the five above-mentioned psycholinguistic variables.

[5] This was a different speaker than in Experiment 1.

Across lists, picture grids were assigned to have a valid or invalid memory probe. This was counterbalanced so that each participant received an equal number of valid and invalid trials; across participants, each item was paired with both valid and invalid memory trials. Two additional target words and two additional foil words were selected for practice trials. All words were recorded by the same female native Dutch speaker as the auditory conditions in neutral prosody and were also matched in intensity using Praat.

Design

The difficulty of lexical selection in planning (Name agreement: high, low) and attention demand of comprehension (focused-attention, divided-attention) were both treated as within participant variables. Name agreement was randomized across trials and blocks and counterbalanced across participants. The focused-attention block always preceded the divided-attention block for all participants. This makes Experiment 1 and Experiment 2 more comparable, and prevents a response strategy where participants continue allocating their attention to listening even in the focused-attention condition because they performed the divided-attention block first. Items assigned to the focused- and divided-attention conditions were counterbalanced across participants, and unlike Experiment 1, each item was shown only once during the experiment.

Procedure

Participants were tested individually in a soundproof room. The experiment was divided into two blocks of 20 trials each (focused-attention, followed by divided-attention), each preceded by four practice trials. Participants took a short break after finishing the first block, and the whole experiment lasted 20 minutes.

In the focused-attention condition (Block 1), trials began with a fixation cross that was presented for 500 ms, followed by a blank screen for 300 ms. Then a 2 × 3 grid appeared on the screen in which six pictures were presented while a 12 second long sound file played. Participants were asked to name the pictures one by one in order (first row, followed by second row) as quickly and accurately as possible while ignoring the Dutch speech. Finally, a blank screen was presented for 1500 ms before the start of the next trial.

In the divided-attention condition (Block 2), trials began with a fixation cross that was presented for 500 ms, followed by a blank screen for 300 ms. Then a 2 × 3 grid appeared on the screen in which six pictures were presented while a 12 second long sound file played.

Participants were again asked to name the pictures one by one in order (first row, followed by second row) while listening to the Dutch speech. Next a blank screen was presented for 700 ms, followed by an auditory word. Participants needed to decide whether this word appeared in the Dutch speech stream they just heard by pressing the left or right button on a button box; assignment of the buttons to yes/no responses was counterbalanced across participants. Then a blank screen was presented for 1500 ms before the start of the next trial.

Analysis

Onset latency and speech duration were again log-transformed. Data were analyzed with linear mixed-effect and ordinal mixed models including the predictors of name agreement and attention demand. Name agreement was contrast-coded as in Experiment 1 (high NA = 0.5; low NA = –0.5), and attention demand (focused-attention / divided-attention) was contrast coded as (0.5, –0.5). All models included random intercepts for participants and items, but random slopes were again not included because of convergence issues and / or evidence of model overfitting. Separate analyses were performed on the same five dependent measures as in Experiment 1. As in Experiment 1, all trials were submitted to analyses of production accuracy. In addition, all fully correct trials were submitted to the response timing and chunking analyses, regardless of memory task accuracy.

As in Experiment 1, to examine whether the results were influenced by the high error rate in naming responses, we also performed a secondary set of analyses on a larger data set comprised of trials with correct first name responses, regardless of the accuracy in the rest of the trial. We also conducted all analyses on trials with correct name responses and correct memory responses to test whether the accuracy of the memory task influenced the effects of name agreement or attention demand on speech planning. These are reported in Appendix D.

2.3.2 Results

Naming accuracy. Participants produced the intended names of all six pictures on 63% of naming trials. As shown in Table 2.3, naming accuracy was affected by both name agreement and attention demand. As supported by a logistic mixed-effect model (see Table 2.4), accuracy for high name agreement pictures was reliably higher than for low name agreement pictures (β = 2.23, SE = 0.23, $p < 0.001$), and accuracy in the focused-attention condition was reliably higher than in the divided-attention condition (β = 0.33, SE = 0.13, $p < 0.01$). Name agreement and attention demand also interacted (β = 0.51, SE = 0.26, $p < 0.05$), showing that accuracy for high name agreement pictures was higher in the focused-attention than in the divided-attention condition (β = 0.59, SE = 0.20, p < 0.01), with no such difference for low name agreement pictures.

Memory task accuracy. In the divided-attention condition, accuracy for the memory task was 67% overall (range: 45% - 90%), and was equal across the high name agreement (67%, range: 40% - 100%) and low name agreement conditions (also 67%, range: 40% - 90%). Participants tended to more often correctly reject invalid memory probes than correctly accept valid ones in both high name agreement (78% for invalid, 56% for valid) and low name agreement conditions (82% for invalid, 52% for valid).

Table 2.3. Dependent measures in Experiment 2 by name agreement and attention demand.

	High name agreement		Low name agreement	
	Focused-attention	Divided-attention	Focused-attention	Divided-attention
Accuracy (%)	86 (20-100)	79 (10-100)	44 (10-90)	42 (0-80)
Onset latencies (ms)	1083 (386)	1132 (442)	1367 (574)	1357 (494)
Speech durations (ms)	4587 (951)	4832 (1241)	6026 (1286)	6102 (1383)
Total chunk number	2.4 (1.3)	2.7 (1.4)	3.9 (1.3)	3.9 (1.5)
First chunk length	3.5 (1.9)	3.1 (1.8)	2.2 (1.3)	2.1 (1.5)

Note. All timing and chunking measures reflect fully correct naming trials only. For accuracy, range follows in parentheses, for other measures, standard deviation follows in parentheses.

Onset latency. As shown in Figure 2.3 (left), log-transformed onset latency was affected by name agreement only. As supported by a linear mixed-effect model (see Table 2.4), it took reliably longer for participants to plan names for low name agreement pictures than high name agreement pictures (β = –0.18, SE = 0.04, p < 0.001). No attention demand effect was observed, and name agreement and attention demand did not interact.

Speech duration. As shown in Figure 2.3 (right), log-transformed speech duration was affected by name agreement and attention demand. As supported by a linear mixed-effect model (see Table 2.4), it took reliably longer for participants to plan names for low name agreement pictures than high name agreement pictures (β = –0.26, SE = 0.02, p < 0.001). Log-transformed speech duration in the divided-attention condition was reliably longer than in the focused-attention condition (β = –0.03, SE = 0.01, $p < 0.05$). Name agreement and attention demand did not interact[6].

Total chunk number. As shown in Figure 2.4 (left) and Table 2.3, total chunk number was affected by name agreement and attention demand. As supported by an ordinal mixed model (see Table 2.4), participants grouped their responses in more small chunks for low name agreement pictures than high name agreement pictures (β = –1.27, SE = 0.11, $p < 0.001$). Participants also grouped their responses in more small chunks in the divided-attention than in the focused-attention conditions (β = –0.15, SE = 0.07, $p < 0.05$). Name agreement and attention demand interacted (β = –0.32, SE = 0.14, $p < 0.05$) such that participants grouped the high name agreement pictures into more small chunks in the divided-attention condition than in the focused-attention condition (β = –0.31, SE = 0.08, $p < 0.001$), with no difference for low name agreement pictures.

First chunk length. As shown in Figure 2.4 (right) and Table 2.3, first chunk length was also affected by name agreement and attention demand. As supported by an ordinal mixed model (see Table 2.4), participants planned, on average, fewer names in their first response chunk for low name agreement than high name agreement pictures (β = 0.87, SE = 0.13, $p < 0.001$), as they made fewer responses with maximal first chunks (i.e. chunk length = 6) in the low name agreement than in the high name agreement conditions (see Figure 2.4 (right)). The first chunk length was also shorter, on average, in the divided-attention condition than in the focused-attention condition (β = 0.17, SE = 0.08, $p < 0.05$), as participants made more responses with maximal first chunks (i.e. chunk length = 6) in the focused-attention than in the divided-attention conditions (see Figure 2.4 (right)). Name agreement and attention demand did not interact.

---

[6] As in Experiment 1, we also performed analyses on log-transformed total pause time. Log-transformed total pause time was only affected by name agreement (β = –1.40, SE = 0.22, $p < 0.001$), suggesting that it took longer for participants to plan names for low name agreement than high name agreement pictures. Attention demand did not affect log-transformed total pause time, and it also did not interact with name agreement.

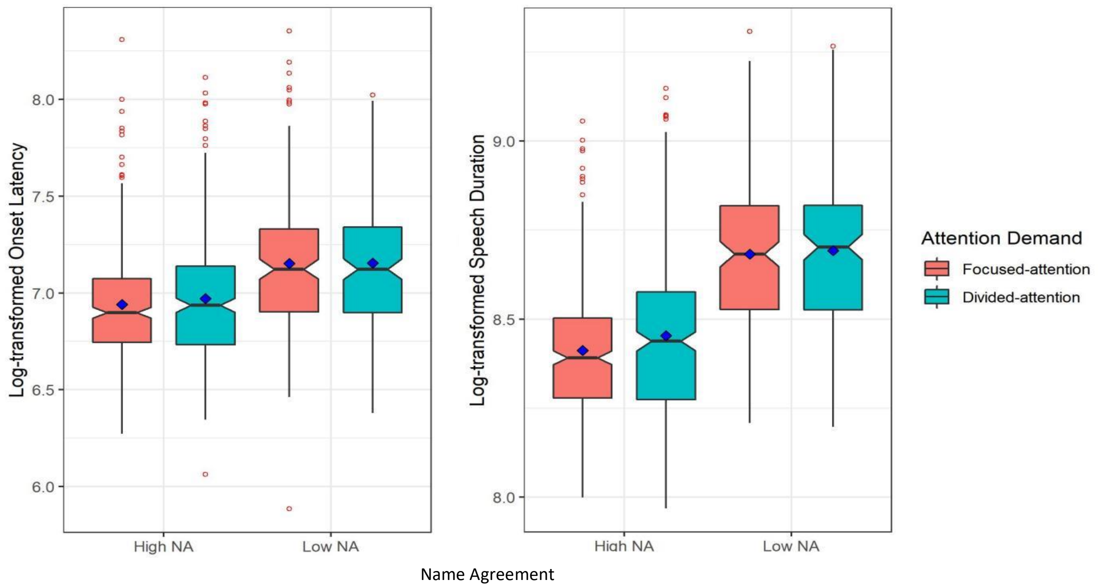

Figure 2.3. Log-transformed onset latencies (left) and log-transformed speech durations (right) in Experiment 2 split by attention demand (focused-attention, divided-attention) and name agreement (NA; high, low). Per condition, blue squares represent means and red points reflect outliers. All measures reflect fully correct naming trials only.

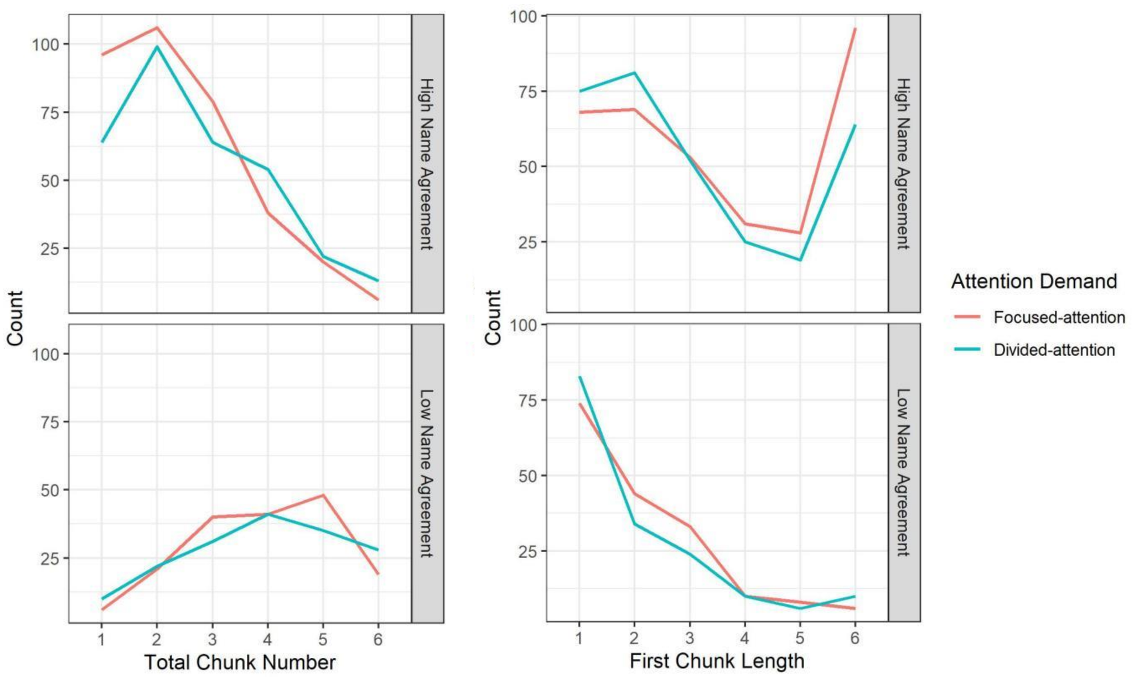

Figure 2.4. Total chunk number (left) and first chunk length (right) in Experiment 2 split by attention demand (focused-attention, divided-attention) and name agreement (NA; high, low). All measures reflect fully correct naming trials only.

Table 2.4. Mixed-effect models for log-transformed onset latencies (Log-Onset), log-transformed speech durations (Log-Duration), accuracy, and chunk measures in Experiment 2.

	Fixed effects	Estimate	SE	*t* value	*p*
Log-Onset	Intercept	7.06	0.03	207.111	< 0.001
	NA (High vs. Low)	-0.18	0.04	-4.563	< 0.001
	Attention Demand (Focused vs. Divided)	-0.03	0.02	-1.857	0.064
	NA × Attention Demand	-0.01	0.04	-0.182	0.856
Log-Duration	Intercept	8.57	0.02	405.177	< 0.001
	NA (High vs. Low)	-0.26	0.02	-11.572	< 0.001
	Attention Demand (Focused vs. Divided)	-0.03	0.01	-2.295	< 0.05
	NA × Attention Demand	-0.04	0.02	-1.594	0.111
	Fixed effects	Estimate	SE	*z* value	*p*
Accuracy	Intercept	0.75	0.17	4.440	< 0.001
	NA (High vs. Low)	2.23	0.23	9.765	< 0.001
	Attention Demand (Focused vs. Divided)	0.33	0.13	2.596	< 0.01
	NA × Attention Demand	0.51	0.26	2.008	< 0.05
Total chunk number	NA (High vs. Low)	-1.27	0.11	-11.462	< 0.001
	Attention Demand (Focused vs. Divided)	-0.15	0.07	-2.057	< 0.05
	NA × Attention Demand	-0.32	0.14	-2.214	< 0.05
First chunk length	NA (High vs. Low)	0.87	0.13	6.980	< 0.001
	Attention Demand (Focused vs. Divided)	0.17	0.08	2.249	< 0.05
	NA × Attention Demand	0.23	0.15	1.526	0.127

Note. All measures reflect fully correct naming trials only. NA refers to name agreement.

Trials with correct first responses. As shown in Appendix D (see Table D1), patterns differed slightly between the conservatively coded data set and the larger data set including all trials in which at least the first word was named accurately. The attention demand effect disappeared on accuracy but appeared on onset latency, and the interaction between name agreement and attention demand disappeared on accuracy but appeared on the measures of speech duration and first chunk length. However, all patterns were in the same direction and were broadly consistent with similar sources of interference in linguistic dual-tasking.

Correct memory trials. As shown in Appendix D (see Table D2), the pattern of results that took only the correct trials from the divided-attention condition, and all trials from the focused-attention condition, was highly comparable to the main analysis. The only difference was that an additional interaction between name agreement and attention demand appeared on speech duration, showing a divided-attention effect only for high name agreement pictures. This suggests that similar levels of interference arose regardless of whether participants were successful in the memory task.

2.3.3 Discussion

In this experiment, participants were either asked to focus on the speech planning task or divide their attention between speech planning and trying to remember the spoken words for a later memory test. This difference in the listening task affected all dependent measures except onset latency, which indicates that the increasing attention demand of listening increased interference during production. This is consistent with a capacity limitation account of interference in dual-tasking (Pashler, 1994; Ruthruff et al., 2003). However, it is also consistent with code conflict in dual-tasking because the linguistic representations of the spoken words may have been activated more strongly when the participants tried to memorize them than when they tried to ignore them.

As in Experiment 1, we manipulated the name agreement of to-be-named pictures in order to assess the role of interference on lexical selection for production. We replicated the name agreement effects found in Experiment 1 on all dependent measures, demonstrating again that competitive lexical selection slows speech planning and reduces the planned utterance units in each response for multiple-object naming.

While name agreement and attention demand did not interact on the timing measures, we did observe an interaction between the two factors on accuracy and total chunk number. This suggests that when the attention demand for the comprehension task was high, individuals grouped high name agreement pictures into more chunks — coordinating the planning and articulation of the picture names more sequentially — and were reliably less accurate than when attention demand was low, but the effect was not found for low name agreement pictures.

Similar to what we observed in Experiment 1, this pattern is opposite of what we predicted. We discuss this further in the General Discussion.

2.4 General Discussion

In two experiments, we explored how two factors linked to interference in dual-tasking, representational similarity and attention demand, influenced the dual task of speaking while listening, with a focus on their impact on lexical selection in speech planning. Experiment 1 tested the role of representational similarity in dual-task interference. We found that high representational overlap between what participants produced and what they listened to increased interference. Linguistic stimuli (Dutch and Chinese speech) interfered more with concurrent speech planning than language-like noise (eight-talker babble) did, and the linguistic stimuli with the largest overlap with the production task (Dutch speech) caused the most interference. Experiment 2 assessed the role of capacity allocation in dual-task interference. Increased attention demand for comprehension also increased interference, such that naming performance was worse in the divided-attention condition than in the focused-attention condition. Combined, the results from both experiments show that representational similarity and capacity limitation play important roles in the dual-tasking interference that results from simultaneously speech planning and listening.

In both experiments, we also manipulated name agreement. Low name agreement increases competition during lexical selection for production. We found large effects of name agreement in both experiments, showing that increased competition during lexical selection decreased the accuracy of production, decreased planning speed, and reduced the planned utterance units in each response for multiple-picture naming.

Name agreement interacted with representational similarity and attention demand in unpredicted ways. In Experiment 1, representational similarity interacted with name agreement on the measure of onset latency and first chunk length, suggesting that representational similarity modulated planning time and the scope of planned utterances before speech onset for high name agreement pictures. Contrary to our predictions, the results indicate that only planning pictures with low selection demand (i.e. high name agreement pictures) is influenced by overlapping representations from comprehension. In Experiment 2, attention demand interacted with name agreement on the measure of accuracy and total chunk number, modulating the accuracy and the planned utterance units in each response for high name agreement pictures only. These patterns suggest that speakers may actively manage how much interference they are susceptible to in linguistic dual-tasking by changing the way that they coordinate speech planning and articulation of successive words, as we discuss further below.

2.4.1 Lexical selection of planning in continuous speaking and listening

The largest effect across both experiments was the effect of name agreement, which influenced speech production as measured by each dependent measure in each experiment. Compared to high name agreement pictures, speakers took longer to plan the names of low name agreement pictures and made more errors. This finding is consistent with earlier studies using single picture naming in a variety of languages, including English (Cheng et al., 2010; Snodgrass & Yuditsky, 1996; Vitkovitch & Tyrrell, 1995), Welsh (Barry et al., 1997), French (Alario et al., 2004; Bonin et al., 2002), Spanish (Cuetos et al., 1999), and Italian (Dell′Acqua et al., 2000), where low name agreement pictures elicited slower response latencies and lower accuracy. Pictures can differ in name agreement because speakers misidentify objects or because they need to select among several appropriate names activated by the depicted objects (Vitkovitch & Tyrrell, 1995). Our items were designed to elicit multiple names, and since we excluded naming responses which were neither the first nor second most common names from analysis, the name agreement effect in our study likely arose because of varying degrees of competition between candidate names. Pictures with low name agreement evoked more lexical candidates, and it took participants longer to eliminate competitors and select a name (e.g., Alario et al., 2004).

Novel to the current work are effects of name agreement on the measures of speech duration and response chunking. Multiple-object naming requires the retrieval of names of simultaneously presented objects in quick succession and in the correct order. The name agreement effects on speech duration mean that it took speakers longer to articulate the sequences of object names in the low name agreement than in the high name agreement condition. As the object names in the two conditions were matched for length in number of syllables and phonemes, the name agreement effects most likely reflect on the time required to plan the names, rather than any phonetic properties of the names. Thus, the results show that speakers retrieve object names during the whole process of planning the sequence of picture names, which supports the claim that speakers plan speech incrementally (e.g., Levelt, 1989; Levelt et al., 1999; Roelofs, 1998; Wheeldon & Lahiri, 1997).

More interestingly, the response chunking analysis found that speakers planned names of low name agreement pictures in a larger number of shorter chunks compared to high name agreement pictures. As explained in the *Introduction*, in order to produce two object names as part of one chunk, i.e. without an intervening pause, the planning processes for the second object name must begin well before the end of the first object name. The finding that sequences of low name agreement names featured shorter chunks (i.e. more pauses) than sequences of high name agreement names may indicate that speakers were less successful in achieving this tight coordination between articulations and planning. Alternatively, they may have chosen to use smaller planning chunks. As the chunks were defined by intervening pauses (and not, for instance, by reference to prosodic properties of the utterances) we cannot distinguish between these options. However, either way the results indicate that the difficulty of lexical selection not only influences the accuracy and planning time, but also the planned utterance units in each response.

2.4.2 Representational similarity in concurrent production and comprehension

In Experiment 1, we manipulated the representational similarity between production and comprehension by varying the type of auditory information that participants needed to ignore while speaking (Dutch speech, Chinese speech, eight-talker babble). As expected, we observed more interference in the two linguistic conditions compared to the language-like noise condition (eight-talker babble) on all dependent measures except accuracy. This suggests that listening to concurrent linguistic input creates more interference during speech planning, such that it affects the speakers' naming accuracy, speed of production, and the grouping of words into chunks.

Our results show that activated linguistic representations for Dutch speech led to code conflict with what was being concurrently planned, impairing naming performance. In contrast, Chinese speech may only activate some phonemic or phonetic representations, leading to little interference. The results fit with the representational similarity account (Navon & Miller, 1987; Pashler, 1994): activated representations of irrelevant auditory information are incompatible with the representations that needed to be engaged for speech planning, creating conflict and impairing naming performance.

However, Fairs (2019) found that additional interference in picture naming caused by a secondary linguistic task (syllable identification) disappeared when the acoustic complexity of the secondary task was controlled, suggesting that acoustic differences between auditory stimuli may also play a role in dual-task interference. This provides an alternate explanation for the differences between linguistic and language-like noise conditions. The Dutch and Chinese speech conditions were segmented by pauses between two adjacent nouns, while the eight-talker babble was continuous, which could have led to less disruption in picture naming. However, a post-hoc comparison between the Chinese and eight-talker babble conditions argues against this possibility. In this analysis, there were differences between Chinese speech and eight-talker babble only on log-transformed onset latency ($\beta = 0.07$, SE = 0.02, $p < 0.001$) and first chunk length ($\beta = -0.25$, SE = 0.07, $p < 0.001$), showing that the Chinese speech led to more interference than eight-talker babble before articulation, but that both conditions led to similar amounts of interference once speaking was initiated. If the interference effect was primarily driven by differences between conditions in phonological segmentation, we should instead observe differences between Chinese speech and eight-talker babble on measures reflecting processing *during* planning (e.g., speech duration, total chunk number). Therefore, our results are more consistent with the idea that interference between the language and eight-talker babble conditions is attributable to conflict from overlapping linguistic representations.

While there were robust main effects of representational similarity on interference, we found evidence of interaction between representational similarity and name agreement on the measures of onset latency and first chunk length, such that speakers took more time to plan high name agreement pictures before articulation in the two language conditions than the language-like noise condition, and they also planned less in their first response for high name agreement pictures in the Dutch condition than in the Chinese condition. The results suggest that representational similarity modulates lexical selection in terms of initial planning time and the amount of advance planning in utterance generation. However, no such difference was found for low name agreement pictures, which opposed our prediction that greater representational similarity effect would be observed for low name agreement pictures because interference arises from both comprehension and production constraints in this condition.

This unexpected direction of the interaction between name agreement and representational similarity might be for trivial reasons. One possibility is that because of low accuracy in the low name agreement condition, there were too few observations for the low name agreement pictures to observe an interaction with representational similarity. To assess this possibility, we conducted all analyses in a larger data set with all correct first name responses (see Appendix B, Table B1). In this data set, more interactions between name agreement and representational similarity were present (i.e. on the dependent measures of onset latency, total chunk number, and first chunk length), but the pattern was always the same: the effect of representational similarity was larger for high name agreement pictures, indicating that naming simple pictures was modulated by concurrent auditory information but naming difficult pictures was not. This suggests against a power issue in leading to this unexpected interaction.

Another interpretation is that naming low name agreement pictures was so hard that participants had to strategically allocate more attention to them, meaning that they were less likely to process auditory information sufficiently deeply to cause interference. The implication is that representational similarity is only one important source for interference between concurrent planning and listening, as we have aimed to highlight throughout the paper. This is consistent with the proposal by Halin et al. (2014) that when people concentrate harder, they are less likely to notice irrelevant information and there is attenuated processing of background information. This hypothesis suggests that speakers may have strategies available for managing conflict in linguistic dual-tasking situations like conversation, potentially leading to less interference between production and comprehension when they focus on their speech planning task. Investigating the strategic allocation of attention in conversation would therefore be a fruitful direction for future research.

2.4.3 Attention demand of comprehension influences concurrent production

In Experiment 2, we manipulated the attention demand of the comprehension task by asking participants to ignore Dutch speech (focused-attention condition) or attend to it in preparation for a memory task (divided-attention condition). Indeed, naming performance was significantly worse in the divided-attention condition in terms of accuracy, speech duration, total chunk number, and first chunk length. This supports a key prediction of the capacity limitation account (Kahneman, 1973; Navon & Gopher, 1979): the more attentional resources required by one task, the worse performance should be observed on the other task.

Importantly, we again cannot exclude the possibility that dual-task interference might also be caused by activated competing linguistic representations when attention demand was high. As discussed above, when participants allocate more attention to listening in the divided-attention condition, linguistic representations for comprehension might be more activated, creating additional code conflict and causing interference. This further suggests that the effect of attention demand on speech planning is tightly connected with interference from overlapping linguistic representations. A fruitful direction for future work would be to disentangle the unique contribution of each source of interference in linguistic dual-tasking. Note that in everyday conversation, the same “confound” is likely to exist: When speakers plan utterances while others are speaking, more interference should arise as speakers attend more to this input, both because capacity is directed away from speech planning and because linguistic representations from the input become more strongly activated. Alternatively, speakers may stop paying careful attention to their interlocutor once they start planning a response, but the interference from speech input on speech planning may also arise due to involuntary attention capture and / or shared linguistic information.

Despite the overall pattern of interference from increased attention demand, the attention demand effect was not found on the measure of onset latency. One possible reason for this is that speakers may trade off between how much speech they plan and how long they spend planning before articulation: participants did plan reliably fewer words in their first response chunk in the divided-attention condition than the focused-attention condition, which could have potentially minimized any differences in onset latency. However, a follow-up analysis disconfirmed this notion. We found a significant negative, rather than a positive correlation in Experiment 2 between the first chunk length and log-transformed onset latency ($r = -0.14$, $p < 0.001$, $n = 1003$), showing that the more words were planned in the first chunk, the shorter the onset latency. This pattern also obtains for Experiment 1 ($r = -0.16$, $p < 0.001$, $n = 1707$), which clearly argues against the trade-off interpretation. Instead it suggests that onset latency and first chunk length were affected in the same way by certain variables: On easier trials, speakers began to talk earlier than on harder trials and generated a longer first chunk.

Another plausible interpretation for the finding that attention demand did not affect onset latency is that participants might focus on speech planning before articulation no matter whether they were asked to attend to the listening or not. Performance on the secondary memory task is somewhat consistent with this. We found that the memory accuracies were at chance level on earlier items (e.g., the 4th, 5th, and 6th probes) that corresponded roughly to the time window of planning of the first two picture names (see Figure E1 in Appendix E). This suggests that participants might be more engaged in speech planning and might pay less attention to listening in the initial stage of the speaking-listening task, even though they were asked to attend to speech input.

The lack of an attention demand effect on the measure of onset latency could also be because that we had low power to observe any differences, given the few fully correct trials available for analysis (focused-attention: 520 total trials, divided-attention: 483 trials). To test this question, we analyzed all of the data with correct first naming responses regardless of whether the rest of the trial was correct (see Appendix D, Table D1). In this analysis, we indeed found a reliable attention demand effect on onset latency, such that it took longer to begin to name pictures in the divided-attention condition than in the focused-attention condition. This result suggests that the lack of an onset latency effect in the main analyses could be due to low experimental power.

In general, it was clear that while attention demand may or may not have affected onset latency, it did have a clear effect on other measures of interference, including accuracy, speech duration, total chunk number, and first chunk length. These effects are consistent with the finding that speech production requires attention (e.g., Ferreira & Pashler, 2002; Jongman et al., 2015; Mädebach et al., 2011) and show how taking away attentional resources impairs speech planning. When participants had to allocate more attention to listening, speech planning took longer and became more sequential. This strongly supports a role of capacity limitation in the interference that arises in speaking-while-listening.

One caveat in thinking about the effects of the experimental manipulation in Experiment 2 is that the focused-attention condition always preceded the divided-attention condition. This means that fatigue could have contributed to the effects we ascribe to divided attention. However, each test block only took about five minutes to complete and participants were invited to take a break between blocks. Thus, we think that any effects of fatigue were likely to be quite small.

We also found interactions between name agreement and attention demand on overall accuracy and total chunk number, such that speakers made more errors and grouped their responses into more chunks when they retrieved the names of high name agreement pictures in the divided-attention condition than in the focused-attention condition, with no attention demand effect present for low name agreement pictures. This finding opposed our prediction that a greater effect of attention demand would be found for low name agreement pictures than high name agreement pictures. This could again be for several possible reasons.

One possibility is again that the few fully correct observations for low name agreement pictures prohibited us from observing an attention demand effect in the low name agreement trials. To test this, we performed an analysis on a larger data set containing responses where the first word was correct (see Appendix D, Table D1). Again, high name agreement pictures led to differences between the focused-attention and divided-attention condition, with no difference for low name agreement pictures. This suggests against a power issue in explaining the unexpected interaction direction.

An alternative interpretation is that naming low name agreement pictures was quite difficult, meaning that speakers always tended to produce very few picture names in each response chunk, even when they had sufficient attentional resources. When attentional resources were diminished, the planning scope was still at the same low level for low name agreement pictures. Consistent with the hypothesis we discussed above that low name agreement leads to a more steadfast locus of attention, the attention demands of comprehension may make it so that speakers tend to produce more picture names in each chunk only when they have the extra attentional resources to do so.

2.4.4 Outlook

Speakers often talk while hearing others talk at the same time. This situation arises, for instance, when people talk simultaneously in an animated discussion, or when they talk in busy offices or restaurants. Although speaking while others are talking is common, it has rarely been studied in the lab. We presented the results of two experiments using a novel paradigm to do so. The paradigm builds on the well-established picture naming paradigm and requires participants to name multiple pictures while being exposed to continuous speech input. This takes a step towards an ecologically valid way of studying interference in simultaneous speech production and comprehension while preserving experimental control. We showed that indicators of naming accuracy, speed, and fluency were sensitive to effects of different types of speech input, and to variations in the difficulty of the speaking task and the focus of attention. The results, though not fully in line with our expectations, yielded meaningful patterns. They indicate that the paradigm may be fruitfully used in further work.

A number of lines of work suggest themselves. First, as already indicated, we could not separate the effects of diverting attention away from speech planning from the effects of directing attention towards listening. This separation might be achieved in further work by including conditions where participants are asked to listen more or less attentively to non-linguistic as well as linguistic stimuli. Second, we could not determine whether differences in chunking were caused directly by differences in task difficulty or by deliberate changes in participants' planning strategies. This issue might be addressed in further work by more tightly constraining the task (stressing fluency or prescribing the chunk size) or by asking participants to produce sentences instead of lists, where prosody might help to distinguish between pauses between planning chunks from pauses due to unplanned delays in word planning. Finally, presenting participants with spoken sentences rather than word lists would be a way of assessing how sentence understanding is affected by attention and how sentence meaning can affect planning. This would inform theories of sentence production and processing, and would contribute to a better understanding of how people plan speech in conversation.

2.5 Conclusion

Two experiments using a novel linguistic dual-tasking paradigm involving multiple picture naming showed that representational similarity and attention demand caused interference in speech production. This interference affects the amount of time spent at the initial planning stage, the amount of planning done while speaking, and the planned utterance units in each response. Representational similarity interacted with lexical selection during the initial planning before articulation, while attention demand interacted with lexical selection difficulty in how much speakers chose to plan at a time. These results indicate that representational similarity and capacity limitation play important roles in dual-task interference arising from planning while listening, and show how speakers can reduce this interference by changing their planning units in utterance generation. The implication is that while the dual-task nature of conversation leads to interference, individuals may be able to manage this interference by changing when and how they plan their speech.

Acknowledgments

We would like to thank Annelies van Wijngaarden, Sophie Slaats, and Xin Liu for recording materials, Annelies van Wijngaarden and Joe Rodd for translations, Caitlin Decuyper and Maarten van den Heuvel for programming advice, Carlijn van Herpt, Esther de Kerf, Inge Pasman, Marjolijn Dijkhuis, Merel Wolf, Jiska Koemans, Vera van't Hoff, Dophne Jansen, Sjaroes Salimian, and Dennis Joosen for data coding.

Appendices

Appendix A: Stimuli in Experiment 1

Table A1. 252 pictures used in Experiment 1.

Picture Grid	*Picture 1*	*Picture 2*	*Picture 3*	*Picture 4*	*Picture 5*	*Picture 6*
Pictures with high name agreement						
1	bokser	koelkast	dolfijn	pijl	gevangenis	helikopter
2	harp	radio	driehoek	tomaat	leeuw	kruiwagen
3	trap	vlinder	knie	batterij	cactus	paprika
4	zaag	bezem	vliegtuig	waaier	schaap	kiwi
5	handschoen	doedelzak	ster	baard	glas	konijn
6	berg	pijp	eekhoorn	duim	hamer	keuken
7	banaan	slager	orkest	anker	kwal	vuist
8	riem	koning	toetsenbord	microfoon	stier	bloem
9	trechter	kokosnoot	egel	gitaar	roos	steen
10	rug	ballon	weegschaal	kroon	honing	slak
11	ananas	tandarts	spiegel	drumstel	muis	parachute
12	zaklamp	broek	schilderij	kangoeroe	tunnel	robot
13	ezel	rechter	sleutel	arm	dobbelsteen	ketting
14	diamant	wolk	zebra	stopcontact	aardbei	kapper
15	kraan	eiland	schildpad	clown	bril	puzzel
16	geit	pompoen	vlieger	schaduw	kompas	horloge
17	aardappel	kaars	skelet	heks	vleermuis	boog
18	vlag	wasmachine	kikker	aansteker	lepel	fruit
19	vork	bus	trompet	fabriek	papegaai	sok
20	masker	schaar	rups	bijbel	kanon	zwembad
21	grasmaaier	boek	vuurtoren	paraplu	snavel	cowboy
Pictures with low name agreement						
1	trui	baksteen	schedel	lade	klauw	jager
2	duif	melk	foto	nagel	kerkhof	speer

3	brievenbus	engel	snoepje	troon	kasteel	viool
4	bank	walrus	parel	vogelkooi	kerk	schoolbord
5	armband	soldaat	rimpels	gorilla	kruk	vis
6	ijsje	paus	spuit	badkuip	kogel	hagedis
7	wasbak	varken	broekzak	koekje	schrift	naald
8	sigaret	handdoek	kwast	worst	gymzaal	leraar
9	hersenen	soep	ijsberg	koningin	museum	druif
10	knuffel	trein	antenne	buik	olie	piano
11	gang	litteken	planeet	komkommer	motor	badkamer
12	elf	wortels	domino	schatkist	koffie	put
13	prullenbak	schelp	ridder	kaarten	meloen	haven
14	staart	tuinslang	kegel	inktvis	herder	perzik
15	magneet	pion	hengel	brug	driewieler	gevangene
16	hoorn	raam	blad	zanger	plas	jurk
17	rivier	goochelaar	monster	bliksem	chocolade	vinger
18	mossel	garage	balkon	rugzak	cirkel	schep
19	sneeuw	camping	ballerina	kleed	garnaal	pruik
20	munt	tram	doodskist	strand	lamp	kameel
21	park	badjas	regen	walvis	tovenaar	goal

Table A2. 252 spoken Dutch words used in Experiment 1.

List 1	*List 2*	*List 3*	*List 4*	*List 5*	*List 6*	*List 7*
computer	sprinkhaan	molen	pet	zwanger	kas	schorpioen
kam	ambulance	kruik	kopje	kabel	vogel	priester
veter	bot	doolhof	vliegveld	bord	wortel	concert
medaille	kast	borstel	doos	spook	matras	eend
boot	prinses	long	nijlpaard	zwaard	adelaar	rok
veer	gier	theepot	wol	beker	goud	voetbal
postzegel	watermeloen	pop	map	voet	roer	schilder
televisie	schouder	bizon	pistool	lantaarnpaal	lippen	kano
grot	plant	vulkaan	kreeft	wieg	trommel	boon
fee	uil	houthakker	brandblusser	klaver	vuur	gans
salade	accordeon	rekenma-chine	liniaal	vlieg	koe	visser
raket	stoel	tas	parfum	zwemmer	loodgieter	supermarkt
vierkant	oog	krokodil	citroen	eikel	knoop	piloot
scheermes	kaas	hek	zeep	aquarium	ooievaar	appel
borst	tamboerijn	telefoon	potlood	peer	serveerster	golf
tafel	brandweer-man	slang	rook	wenkbrauw	ventilator	kurk
neushoorn	stinkdier	ontbijt	wesp	verwarming	tank	haas
kaal	ring	billen	onderbroek	portemonnee	draak	kapitein
List 8	*List 9*	*List 10*	*List 11*	*List 12*	*List 13*	*List 14*
broodrooster	kers	Schroevend-raaier	kalkoen	kameleon	bijl	sjaal
ader	snor	politieagent	schort	cel	kraai	beer
koffer	graan	tak	lift	enkel	cadeau	pil
fotograaf	haak	luipaard	stropdas	pot	graf	hoek
spijker	vleugel	kar	bos	oorbel	sleutelgat	zak
roofvogel	lijst	bier	telescoop	hert	kettingzaag	standbeeld
zuster	tong	tepel	boter	bloemkool	vlees	luier

boom	camera	rits	wandelstok	kleerhanger	brief	walnoot
krijt	paard	bom	sla	slee	kubus	mug
olijf	steel	olifant	kip	neus	vos	slot
spin	kalender	dienblad	strijkijzer	woestijn	sinaasappel	vogel-verschrikker
geweer	dinosaurus	piraat	boerderij	postbode	koor	apotheek
druppel	thermometer	zout	oven	tijger	haai	bal
pinda	stokbrood	jas	spons	schoen	draad	vingerafdruk
nest	wolf	duivel	beul	druiven	lerares	kever
kies	fakkel	slaapkamer	zwaan	zeehond	vrachtwagen	schild
zonnebloem	krant	prins	boomstam	stoeprand	mand	gordijn
ham	indiaan	krab	pelikaan	hak	blinde	zeemeermin

Table A3. 252 spoken Chinese words used in Experiment 1.

List 1	*List 2*	*List 3*	*List 4*	*List 5*	*List 6*	*List 7*
仙女	奶酪	茶壶	尺子	剑	老鹰	渔夫
奖章	枭	火山	灭火器	苍蝇	火	超市
船	植物	袋子	柠檬	梨子	鼓	足球
秃头	救护车	磨坊	香水	三叶草	纽扣	苹果
桌子	眼睛	鳄鱼	铅笔	奖杯	龙	牧师
犀牛	手鼓	洋娃娃	黄蜂	暖气	风扇	鸭子
计算机	臭鼬	肺	烟	鬼	胡萝卜	波浪
火箭	戒指	盆子	帽子	电缆	坦克	飞行员
羽毛	西瓜	蛇	文件夹	脚	奶牛	野兔
沙拉	蚱蜢	栅栏	河马	摇篮	服务员	船长
鞋带	公主	计算器	手枪	水族馆	黄金	音乐会
正方形	椅子	迷宫	毛线	灯柱	床垫	鹅
洞穴	肩膀	野牛	机场	橡子	鸟	独木舟
邮票	手风琴	早餐	杯子	游泳者	水管工人	瓶塞
胸	秃鹫	樵夫	肥皂	盘子	船舵	画家
梳子	消防员	屁股	龙虾	钱包	嘴唇	蝎子
电视	衣柜	电话	盒子	怀孕	温室	豆
剃刀	骨	发刷	内裤	眉毛	鹳鸟	裙子
List 8	*List 9*	*List 10*	*List 11*	*List 12*	*List 13*	*List 14*
衣箱	谷物	魔鬼	围裙	变色龙	盲人	围巾
护士	印第安人	象	树干	监狱	锁孔	角落
巢	报纸	啤酒	天鹅	衣架	篮子	盾牌
牙齿	马	警察	海绵	雪橇	狐狸	药店
向日葵	火炬	拉链	望远镜	海豹	卡车	皮球
步枪	樱桃	盐	生菜	脚踝	斧头	核桃
猛禽	日历	树枝	刽子手	耳环	橙子	尿布
水滴	恐龙	王子	火鸡	菜花	乌鸦	蚊子

花生	手柄	螺丝刀	熨斗	老虎	老师	锁
面包机	钩	夹克	电梯	脚跟	电锯	美人鱼
蜘蛛	胡子	美洲豹	手杖	葡萄	肉	雕像
摄影师	狼	卧室	黄油	鹿	信件	药丸
火腿	面包	手推车	鸡	罐子	鲨鱼	熊
粉笔	温度计	螃蟹	领带	邮差	立方体	窗帘
血管	画框	海盗	森林	路缘石	合唱团	稻草人
树	舌头	托盘	烤箱	鞋	线	甲虫
钉子	相机	乳头	鹈	沙漠	坟墓	麻袋
橄榄	翅膀	炸弹	牧场	鼻子	礼物	指纹

Table A4. Twenty semantically anomalous sentences used for eight-talker babble in Experiment 1.

No.	*Semantically anomalous sentences*
1	Jouw saaie baken tilde onze taxi op.
2	Een kool zou zijn vermoeide dinsdag laten zinken.
3	Hun stoofpot was een nieuwsgierige gok aan het graven.
4	Mijn puppy zou hun fundamentele liter kunnen benadrukken.
5	De snelle emmer pikte haar tweeling.
6	Het verre budget is de slaperige pet aan het bakken.
7	Marieke zal bestaan uit een lauw teken en een varken.
8	Haar maisonette zou een dubieuze vrachtwagen onderwijzen.
9	Peter en zijn hoofdticket werden overkapt door hun bed.
10	Zijn vriendelijke pudding plakte een decennium op mijn keuze.
11	Haar compacte schrijver zou hun toga en secretarissen splitsen.
12	Het gewone lichaam hield ons hout.
13	De ultieme kapitein zal morgen de fles kraken.
14	Zijn gegrilde koekje leidde de baby af via een clausule.
15	De gevlochten gewoonte draaide haar duif in de segmenten.
16	Haar overvloedige zak cirkelt naar zijn knikker.
17	Mijn rooster zou haar mollige onderwerp met de feiten melken.
18	De gemalen bagage miste de frisdrank kort.
19	De theorie zou haar huis in de oceaan moeten slepen.
20	Onze zeldzame toekomst legde een sprong voor aan de juryleden.

Appendix B: Experiment 1 Supplemental Analyses

Table B1. Mixed-effect models for log-transformed onset latencies (Log-Onset), log-transformed speech durations (Log-Duration), accuracy, and chunk measures in Experiment 1. All measures reflect trials with correct first naming responses.

	Fixed effects	Estimate	SE	*t* value	*p*
Log-Onset	Intercept	7.00	0.03	201.070	< 0.001
	NA (High vs. Low)	-0.12	0.02	-5.514	< 0.001
	Similarity ((Dutch & Chinese) vs. Babble)	0.14	0.02	7.742	< 0.001
	Similarity (Dutch vs. Chinese)	0.09	0.02	5.682	< 0.001
	NA×Similarity ((Dutch & Chinese) vs. Babble)	0.16	0.04	4.363	< 0.001
	NA×Similarity (Dutch vs. Chinese)	-0.01	0.03	-0.360	0.719
Log-Duration	Intercept	8.57	0.03	320.303	< 0.001
	NA (High vs. Low)	-0.17	0.02	-7.900	< 0.001
	Similarity ((Dutch & Chinese) vs. Babble)	0.07	0.01	4.976	< 0.001
	Similarity (Dutch vs. Chinese)	0.08	0.01	6.928	< 0.001
	NA×Similarity ((Dutch & Chinese) vs. Babble)	0.03	0.03	1.073	0.284
	NA×Similarity (Dutch vs. Chinese)	0.04	0.02	1.717	0.086
	Fixed effects	Estimate	SE	*z value*	*p*
Accuracy	Intercept	3.30	0.21	15.856	< 0.001
	NA (High vs. Low)	1.79	0.32	5.513	< 0.001
	Similarity ((Dutch & Chinese) vs. Babble)	-0.34	0.32	-1.048	0.295
	Similarity (Dutch vs. Chinese)	-0.43	0.25	-1.745	0.081
	NA×Similarity ((Dutch & Chinese) vs. Babble)	-0.58	0.65	-0.891	0.373
	NA×Similarity (Dutch vs. Chinese)	0.86	0.50	1.745	0.081
Total chunk number	NA (High vs. Low)	-0.11	0.01	-7.692	< 0.001
	Similarity ((Dutch & Chinese) vs. Babble)	0.32	0.08	4.107	< 0.001
	Similarity (Dutch vs. Chinese)	0.06	0.07	0.807	0.420
	NA×Similarity ((Dutch & Chinese) vs. Babble)	-0.01	0.02	-0.307	0.759
	NA×Similarity (Dutch vs. Chinese)	0.04	0.02	2.228	< 0.05

First chunk length	NA (High vs. Low)	0.42	0.07	6.189	< 0.001
	Similarity ((Dutch & Chinese) vs. Babble)	-0.27	0.06	-4.341	< 0.001
	Similarity (Dutch vs. Chinese)	-0.03	0.05	-0.485	0.628
	NA×Similarity ((Dutch & Chinese) vs. Babble)	-0.24	0.13	-1.887	0.059
	NA×Similarity (Dutch vs. Chinese)	-0.33	0.11	-2.984	< 0.01

Note. NA refers to name agreement, similarity refers to representational similarity.

Appendix C: Stimuli in Experiment 2

Table C1. 468 additional spoken Dutch words (26 Dutch lists) used in Experiment 2.

List15	*List 16*	*List 17*	*List 18*	*List 19*	*List 20*
pakje	straat	parkeerplaats	toilet	slijm	schoonmaakster
zonsopgang	noten	schip	nagels	ondergoed	paal
vuurwerk	plein	rek	schot	papiertje	vlam
spoor	schuif	oprit	pakket	schuim	spieren
plaat	rechterhand	tekening	vleugels	regenboog	restaurant
scheur	plakband	poep	sokken	tasje	pap
taart	speeksel	shampoo	pony	voorhoofd	slaapplaats
vacht	verf	rozen	timmerman	post	tenen
strik	zolder	vijver	rijtje	spek	zwembroek
pudding	scheet	zaklantaarn	sprong	tractor	poesje
zetel	poot	postkantoor	puree	vest	saus
schoot	rol	strips	tweeling	station	vuilnis
woonkamer	salami	vrucht	veld	pols	pretpark
tent	voetstappen	wip	stoot	wekker	wonde
sandwich	stam	podium	wanten	spiegeltje	peper
rietje	waterval	rots	tanden	muziek	stoep
voeding	zaad	zwaai	ziekenwagen	zadel	zeeman
stofzuiger	tang	tuin	worp	stekker	struik
List 21	*List22*	*List 23*	*List 24*	*List 25*	*List 26*
bever	lasagne	beest	bek	bessen	dame
eitje	kachel	kamp	vuurtoren	paraplu	klink
kaarsen	park	regen	dierenarts	apparaat	muzikant
achterdeur	bonen	spier	fontein	spel	scherm
barst	achterwerk	fluit	gras	mosterd	walvis
geur	hokje	gordel	hotel	doek	armen
papier	kool	hoofdje	kap	grond	broodje

kont	feestje	klei	regenjas	kassa	halsband
rem	danseres	leger	altaar	limonade	inbreker
bocht	kleding	meubels	mist	poes	linkerkant
eieren	wagen	rolstoel	botten	tuinman	dorp
machine	pastoor	watje	pleister	friet	muntje
lakens	glimlach	plafond	kousen	klik	keel
badjas	speelgoed	afwas	lijf	brandweer	stempel
vruchten	beek	kostuum	spoorweg	appeltaart	vel
klap	mantel	terras	wolken	worm	zaal
tapijt	rijst	deksel	traan	kooi	pomp
snoep	kalf	borsten	meeuw	sap	frisdrank

List 27	*List 28*	*List29*	*List 30*	*List 31*	*List 32*
bloemen	dynamiet	laarzen	band	bestek	badpak
kraag	bakker	schuilplaats	ladder	kauwgom	kelder
stok	kerstboom	ballet	aardappelen	grasmaaier	schijf
drank	schop	kleren	kist	applaus	halsketting
gazon	gebit	cake	circus	donder	tovenaar
bui	handvat	duister	eetkamer	frietjes	draai
klop	jasje	hartslag	gelach	haard	muts
lintje	potje	gips	hoef	ijskast	inkt
jacht	buurvrouw	jojo	juweel	klim	kudde
kuil	drol	oren	tand	linkerhand	bibliotheek
verrekijker	knijp	knop	reep	wapen	lint
biefstuk	lippenstift	blikje	koets	muggen	gaatje
handschoenen	oerwoud	puntje	blok	brood	poort
portefeuille	vingers	vloer	voordeur	politieman	stof
zand	ziekenhuis	zonlicht	zweep	schaal	venster
bagage	bladzijde	lucifer	luis	vaas	zakdoek
kruis	kust	kijker	pakjes	kring	bruid

schoenen	streep	stroop	sirene	yoghurt	klok
List 33	*List 34*	*List 35*	*List 36*	*List 37*	*List 38*
basketbal	koord	begraafplaats	kerel	bumper	bakkerij
cowboy	beeld	kampvuur	snavel	gebak	ketel
achtertuin	goal	fluitje	benzine	kermis	geboorte
kaart	adem	dessert	dierentuin	schoorsteen	schreeuw
reus	telefoonboek	hooi	fornuis	agenda	laars
hol	deken	lichaam	kapsel	drankje	fiets
ijs	flits	boterham	lijm	handtas	cafetaria
boer	boord	mieren	mol	knal	mens
hok	gootsteen	kous	groenten	bak	onkruid
lap	stapel	plank	salon	lip	haren
danser	kleedkamer	jam	huiskamer	poster	vlek
klas	druppels	spinnen	poeder	veters	zilver
parels	leer	kleingeld	klets	zegel	luchthaven
spaghetti	pit	toeter	brand	kus	prik
teddybeer	gebouw	bil	woning	gordijnen	knip
vuilnisbak	tennis	winkel	trouwring	storm	blik
mannetje	riool	rotsen	kras	buis	strip
emmer	steentje	kistje	stal	nek	jeuk
List 39	*List 40*				
schuur	hoepel				
koekjes	boek				
bloedneus	aardbeien				
luik	kaak				
toren	dagboek				
chauffeur	gereedschap				
jungle	rund				
zucht	blondje				

gehakt	laken
voeten	maanlicht
pagina	slagroom
heuvel	bandje
kinderjuf	tandpasta
rand	kok
lach	paleis
suiker	voorruit
dwerg	wind
balletje	hoest

Note. We used the spoken Dutch words in lists 1 through 14 from Experiment 1, and the other 26 Dutch lists were shown in this table.

Appendix D: Experiment 2 Supplemental Analyses

Table D1. Mixed-effect models for log-transformed onset latencies (Log-Onset), log-transformed speech durations (Log-Duration), accuracy, and chunk measures in Experiment 2. All measures reflect trials with correct first naming responses.

	Fixed effects	Estimate	SE	*t* value	*p*
Log-Onset	Intercept	7.07	0.04	192.330	< 0.001
	NA (High vs. Low)	-0.18	0.04	-4.778	< 0.001
	Attention Demand (Focused vs. Divided)	-0.05	0.01	-3.387	< 0.001
	NA × Attention Demand	0.00	0.03	0.043	0.966
Log-Duration	Intercept	8.60	0.02	385.093	< 0.001
	NA (High vs. Low)	-0.28	0.02	-12.225	< 0.001
	Attention Demand (Focused vs. Divided)	-0.03	0.01	-2.602	< 0.01
	NA × Attention Demand	-0.04	0.02	-2.243	< 0.05
	Fixed effects	Estimate	SE	*z* value	*p*
Accuracy	Intercept	3.66	0.38	9.591	< 0.001
	NA (High vs. Low)	1.65	0.63	2.625	0.009
	Attention Demand (Focused vs. Divided)	0.42	0.23	1.787	0.074
	NA × Attention Demand	0.15	0.47	0.333	0.739
Total chunk number	NA (High vs. Low)	-1.26	0.10	-11.992	< 0.001
	Attention Demand (Focused vs. Divided)	-0.21	0.06	-3.804	< 0.001
	NA × Attention Demand	-0.27	0.11	-2.404	< 0.05
First chunk length	NA (High vs. Low)	0.87	0.12	7.013	< 0.001
	Attention Demand (Focused vs. Divided)	0.20	0.06	3.484	< 0.001
	NA × Attention Demand	0.24	0.12	2.063	< 0.05

Note. NA refers to name agreement.

Table D2. Mixed-effect models for log-transformed onset latencies (Log-Onset), log-transformed speech durations (Log-Duration), accuracy, and chunk measures in Experiment 2. All measures reflect fully correct naming trials and correct memory responses in the divided-attention condition.

	Fixed effects	Estimate	SE	*t* value	*p*
Log-Onset	Intercept	7.06	0.03	208.615	< 0.001
	NA (High vs. Low)	-0.19	0.04	-4.899	< 0.001
	Attention Demand (Focused vs. Divided)	-0.02	0.02	-1.213	0.226
	NA × Attention Demand	0.01	0.04	0.261	0.794
Log-Duration	Intercept	8.58	0.02	407.956	< 0.001
	NA (High vs. Low)	-0.26	0.02	-10.872	< 0.001
	Attention Demand (Focused vs. Divided)	-0.03	0.01	-2.575	0.01
	NA × Attention Demand	-0.05	0.03	-2.057	< 0.05
Accuracy	Fixed effects				
	Intercept	0.77	0.16	4.673	< 0.001
	NA (High vs. Low)	2.12	0.23	9.119	< 0.001
	Attention Demand (Focused vs. Divided)	0.28	0.14	1.976	< 0.05
	NA × Attention Demand	0.65	0.28	2.309	< 0.05
Total chunk number	NA (High vs. Low)	-1.22	0.12	-10.165	< 0.001
	Attention Demand (Focused vs. Divided)	-0.17	0.08	-2.121	< 0.05
	NA × Attention Demand	-0.40	0.16	-2.491	< 0.05
First chunk length	NA (High vs. Low)	0.84	0.13	6.313	< 0.001
	Attention Demand (Focused vs. Divided)	0.18	0.08	2.066	< 0.05
	NA × Attention Demand	0.29	0.17	1.739	0.082

Note. NA refers to name agreement.

Appendix E: Performance on secondary memory task

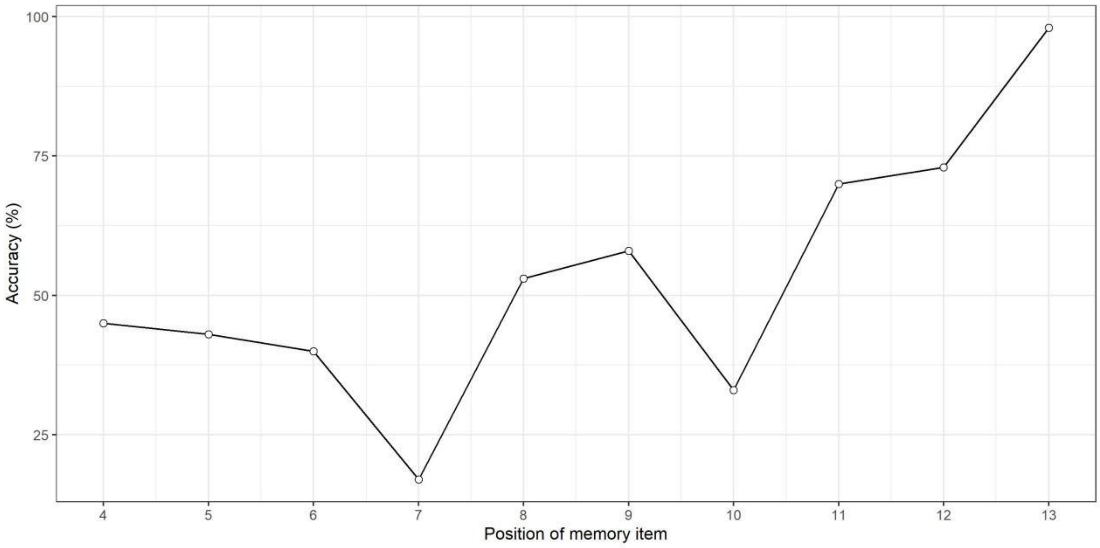

Figure E1. Accuracy in the memory task in the divided-attention condition per position of the memory item.

3 | Conducting language production research online: A web-based study of semantic context and name agreement effects in multi-word production[7]

Abstract

Few web-based experiments have explored spoken language production, perhaps due to concerns of data quality, especially for measuring onset latencies. The present study highlights how speech production research can be done outside of the laboratory by measuring utterance durations and speech fluency in a multiple-object naming task when examining two effects related to lexical selection: semantic context and name agreement. A web-based modified blocked-cyclic naming paradigm was created, in which participants named a total of sixteen simultaneously presented pictures on each trial. The pictures were either four tokens from the same semantic category (homogeneous context), or four tokens from different semantic categories (heterogeneous context). Name agreement of the pictures was varied orthogonally (high, low). In addition to onset latency, five dependent variables were measured to index naming performance: accuracy, utterance duration, total pause time, the number of chunks (word groups pronounced without intervening pauses), and first chunk length. Bayesian analyses showed effects of semantic context and name agreement for some of the dependent measures, but no interaction. We discuss the methodological implications of the current study and make best practice recommendations for spoken language production research in an online environment.

[7] Adapted from He, J., Meyer, A. S., Creemers, A., & Brehm, L. (2021). Conducting language production research online: A web-based study of semantic context and name agreement effects in multi-word production. *Collabra: Psychology*, *7*(1), 29935. https://doi.org/10.1525/collabra.29935

3.1 Introduction

The use of internet-based experiments for behavioral research has gained in popularity over the last few years, driven by the increasing ease and efficiency with which larger and more diverse samples of participants can be reached (e.g., Reimers & Stewart, 2015) and by the Covid-19 pandemic (e.g., Sauter et al., 2020). In psycholinguistics, web-based variants of sentence comprehension and word recognition experiments elicit good quality data in questionnaires or typed responses (e.g., Cooke et al., 2011; Schnoebelen & Kuperman, 2010). However, web-based experiments of spoken production are still uncommon. At the time of planning this study, there were two main concerns: one concerned the quality of speech recording made outside of a laboratory environment, the other concerned the precision of measurement of speech onset latencies due to potentially poor audiovisual synchrony. That is, it was not clear whether the timing of visual stimuli on the participant's screen and of the onset of the recording of their responses could be controlled precisely enough to obtain useful measures of speech onset latencies (see also Bridges et al., 2020). The current study therefore explored the usefulness of dependent measures that did not depended on this synchrony, but were derived from the durations and fluency of the participants' utterances. Meanwhile, recent speech production studies have shown that onset latencies can in fact be measured with good accuracy in web-based platforms (e.g., Fairs & Strijkers, 2021; Stark et al., 2022; Vogt et al., 2022). We review these studies in the Discussion section.

The present study measured utterance durations and utterance internal pauses (indexing speech fluency) offline during multiple-utterance production. Unlike speech onset latency, the precision of temporal characteristics within participants' audio recordings can be guaranteed sufficiently in web-based experiments: the interval between the recorded utterance onset and offset (i.e., utterance duration), or the interval between the offset of the first word and the onset of the second word (i.e., pause time) can be measured from the recording itself. These measures are limited by the quality of the participants' recording equipment and the researcher's speech analysis tools, but not by the issue of audiovisual synchrony, which means that regardless of how successful the audiovisual synchrony is, we should be able to obtain reliable measurements.

Language production work has typically exploited speech onset latency as the dependent variable, but variations in other characteristics of the utterance, such as utterance duration and speech fluency (indexed by pauses), are also promising measures for examining multi-word production (e.g., Ferreira & Swets, 2017; Kandel et al., 2021; Momma & Ferreira, 2019). This is because speakers do not necessarily fully plan multi-word utterances before beginning to speak, but rather often continue planning while articulating their utterance. The clearest evidence for this comes from studies recording participant's eye movements while they are describing scenes or events (e.g., Griffin & Bock, 2000; Konopka, 2019). Most relevant to the present study are multiple-object naming studies (e.g., Belke & Meyer, 2007; Meyer et al., 2012; Mortensen et al., 2008), which have showed that when speakers are asked to name sets of three or more objects, they usually fixate upon them in the order of mention, with the eyes running slightly ahead of the articulation of the object names. Speakers typically initiate their utterance after the shift of gaze to the second or third object. This pattern shows that speech planning continues after utterance onset. Since an upcoming word may be planned while another word is being articulated, the difficulty of word planning may be reflected in the time elapsed between word onsets, where speakers may either stretch words or insert pauses between them. Consequently, variation in the difficulty of planning processes can manifest itself not only in onset latencies, but also in utterance durations and speech fluency (see also Lee et al., 2013).

To investigate how speech production research can be done outside of the laboratory by measuring utterance durations and speech fluency, we created a modified blocked-cyclic naming paradigm to examine two previously studied phenomena related to lexical selection: semantic context and name agreement effects. The design of the modified blocked-cyclic paradigm was inspired by work of Belke and Meyer (2007), who explored semantic context effects in picture naming. The semantic context effect is the finding that it is more difficult to name multiple objects from the same semantic category (a homogeneous context) than from different semantic categories (a heterogeneous context). In most semantic context experiments, one picture is presented per trial and onset latencies are measured (e.g., Damian & Als, 2005; Damian et al., 2001). However, Belke and Meyer (2007, Experiment 1b), explored semantic context effects during multiple object naming in young (college-aged) and older (52-68 years) speakers. On each trial four objects belonging to the same or different semantic categories were presented simultaneously on the screen and had to be named. The authors found small but significant semantic context effects on word durations for both groups of speakers, and a significant semantic context effect on pause rate for the older speakers. This indicates that semantic context effects can be obtained on measures such as utterance durations and speech fluency. These measures should remain reliable in web-based research because they are derived from the participants' speech alone rather than the timing of their speech relative to a stimulus.

The paradigm used in the current study was further inspired by studies on rapid automatized naming (RAN), used primarily in neuropsychological work. In a RAN task, a set of familiar items (e.g., five objects or digits) repeated multiple times across rows of a grid is named as quickly as possible, and the total naming time of the grid is measured (Denckla & Rudel, 1976). There are large individual differences in total naming times. Moreover, total naming times depend also on properties of the materials such as the word frequency and phonological neighborhood density of the object names (Araújo et al., 2020). This implies that when objects are repeatedly named in a grid, variation in the difficulty of speech production can be reflected in total naming times.

Inspired by these two lines of work, we created a modified blocked-cyclic naming paradigm suitable for web-based research. On each trial, participants were asked to name sixteen pictures that were simultaneously presented in a 4 × 4 grid. Each set of sixteen pictures consisted of repetitions of four pictures which belonged either to the same semantic category or to different semantic categories, quadrupling the number of pictures named per trial in Belke and Meyer (2007). Orthogonally, name agreement for the pictures was varied. We measured five main dependent variables: accuracy, utterance duration, total pause time, total chunk number, and first chunk length. A chunk was defined as a group of words produced without intervening pause longer than 200 ms (for details, see Methods). While we were not entirely confident about the reliability of onset latencies, we also measured them, allowing us to make a rough comparison with lab-based studies.

The modified blocked-cyclic naming paradigm was used to examine whether effects of semantic context and name agreement would be obtained on dependent variables that can be measured reliably on web-based experimental platforms. We selected these variables because they were deemed likely to affect lexical selection in different ways. As noted earlier, the semantic context effect is the finding that speakers are slower and less accurate to repeatedly name small sets of objects in homogeneous contexts than in heterogeneous contexts (e.g., Belke & Meyer, 2007; Damian & Als, 2005; Damian et al., 2001). The semantic context effect has been attributed to the selection of lexical-semantic entries (i.e., lemmas): selecting a target lexical representation is more difficult in the context of semantically related than unrelated items (Damian et al., 2001). Importantly, the semantic context effect takes some time to build up: Typically, participants show either no semantic interference effect or a semantic facilitation effect when they name the pictures for the first time, but from the second cycle onward, they display a stable semantic interference effect (Belke, 2017; Belke et al., 2005; Damian & Als, 2005). Given that semantic context effects were mainly found on word durations in multiple object naming (Belke & Meyer, 2007), we predicted that in our paradigm semantic context effects would start to emerge, especially on the measure of utterance durations, when participants began to name the second row of objects.

Name agreement is the extent to which participants agree on the name of a picture. The name agreement effect refers to the finding that naming a picture with high name agreement (e.g., a picture of a *banana*) is faster and more accurate than naming a picture with low name agreement (e.g., a picture of a piece of furniture which could be called *sofa*, *couch*, or *settee*; Alario et al., 2004; Vitkovitch & Tyrrell, 1995). Name agreement effects come from multiple sources. The name agreement effect is found for objects that are often incorrectly named (e.g., *celery*, which is commonly misidentified as *rhubarb*, *Chinese leaves*, or *cabbage*), reflecting difficulty in object recognition. The effect has also been obtained for objects with multiple plausible names (e.g., a *jumper* is also called *sweater*, *pullover*, *jersey*, or *sweatshirt*), reflecting difficulty at the lexical selection stage of spoken language production (Alario et al., 2004; Shao et al., 2014; Vitkovitch & Tyrrell, 1995). The present study focused on the latter effect: these low name agreement pictures evoke more lexical candidates than pictures with high name agreement, and hence, it takes longer to eliminate candidates and select one name. Thus, we predicted that name agreement would affect utterance durations, total pause time and chunk measures.

The effects of semantic context and name agreement are interesting to investigate in tandem because their relationship can provide some insight into how lexical selection is achieved in speech production. Existing models proposed to account for semantic context effects (e.g., Abdel Rahman & Melinger, 2009; Howard et al., 2006; Oppenheim et al., 2010) disagree on whether lexical selection during spoken language production is competitive or not. This disagreement means that these models make different predictions about whether increasing the number of activated lemmas during lexical selection will increase semantic context effects. Models with lexical competition (e.g., Abdel Rahman & Melinger, 2009; Howard et al., 2006) predict that semantic context should interact with name agreement, such that the semantic context effects would be stronger for low name agreement pictures than high name agreement pictures. By contrast, models not assuming lexical competition predict that semantic context effects should not be influenced by name agreement (e.g., Oppenheim et al., 2010).

3.2 Methods

Participants

We recruited 41 native Dutch speakers (36 females, M_{age} = 22 years, range: 19 - 26 years) from the participant pool at the Max Planck Institute for Psycholinguistics. This is about twice the sample size used in most semantic context experiments (e.g., 16 participants in Belke & Meyer, 2007; 24 participants in Damian & Als, 2005) and seemed appropriate for an exploratory study. All participants reported normal or corrected-to-normal vision and no speech or hearing problems. They signed an online informed consent form and received a payment of €6 for their participation. The study was approved by the ethics board of the Faculty of Social Sciences of Radboud University.

Apparatus

The experiment was implemented on FRINEX (FRamework for INteractive EXperiments), a web-based platform developed by the technical group at the Max Planck Institute for Psycholinguistics (for details, see Withers, 2017). It was displayed on the participants' own laptops; we restricted participation to 14 or 15.6 inch laptops with Google Chrome, Firefox, or Safari web browsers. Participants' speech was recorded by a built-in voice recorder of the web browser. WebMAUS Basic was used for phonetic segmentation and transcription (https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic).
Praat (Boersma & Weenink, 2009) was then used to extract the onsets and offsets of all segmented responses.

Materials

Thirty-two pictures with one- or two- syllable primary names (see Appendix A, Table A1) were selected from the MultiPic database of 750 single-object drawings (Duñabeitia et al., 2018), which provides language norms (e.g., name agreement, visual complexity) in standard Dutch. Of these, sixteen were high name agreement pictures, all with name agreement percentage of 100%, and sixteen were low name agreement pictures, with name agreement percentages between 50% and 85% (M = 64%, SD = 11%). Independent *t*-tests revealed that the two sets of pictures differed significantly in name agreement, but not in any of ten other psycholinguistic attributes[8]. For all low name agreement pictures, their first and second modal names in the MultiPic database share the same semantic features (e.g., *kat 'cat'* and *poes 'cat'*), as judged by a native speaker of Dutch.

The sixteen high name agreement and sixteen low name agreement pictures were selected from four semantic categories (animal, body part, clothing, and tool), with four of each semantic category. Each set of sixteen pictures was used to make a matrix of 4 × 4 picture grids such that the rows corresponded to the categories and thus formed homogeneous stimulus sets of four pictures each, whereas columns formed the sets for heterogeneous condition of the same size. Two picture names in each row and in each column were monosyllabic and two were bisyllabic. Pictures were selected to minimize within-category visual similarity and avoid shared initial phoneme or letter.

To equate the semantic similarity between the high and low name agreement conditions, we calculated the semantic similarity of all six pairs within each four-picture set by using *sub2vec* (van Paridon & Thompson, 2021). In homogeneous sets, semantic similarities of the pairwise combinations of all pictures per set were matched across semantic categories by name agreement. Independent sample *t*-test showed that there was no difference in semantic similarity between high and low name agreement pictures ($t_{(46)} = 0.004$, $p = 0.997$). In heterogeneous sets, semantic similarities for the pairwise combinations of all pictures per set were also matched. Independent *t*-test also showed that there was no difference on semantic similarities between high and low name agreement pictures in each heterogeneous set ($ts < -0.6$, $ps > 0.01$).

On each trial, a 4 × 4 picture grid was presented from the matrix described above. There were eight picture grids (four for homogeneous trials, four for heterogeneous trials) for each name agreement condition, resulting in sixteen picture grids in total (i.e., 16 trials). Each picture grid was shown three times in different test blocks, which results in 48 trials in total. This means each individual picture was repeated six times (twice per block: once for a homogeneous picture grid, and once for a heterogeneous picture grid) during the experiment. Sixteen

[8] Ten matched variables: visual complexity, age-of-acquisition, word frequency, number of syllables, number of phonemes, word prevalence, phonological neighborhood frequency, phonological neighborhood size, orthographic neighborhood frequency, and orthographic neighborhood size.

additional pictures (combined into four picture grids) were selected from the same database as practice stimuli, resulting in four practice trials.

Design

Semantic context (homogeneous, heterogeneous) and name agreement (high, low) were both treated as within-participant variables; both were randomized within experimental blocks and counterbalanced across participants. The same four pictures per homogeneous or heterogeneous set were presented in a different arrangement across blocks and participants with a Latin square design so that each item appeared in each ordinal position. Within a picture grid, note that the same items did always follow each other (e.g., *leeuw 'lion'* always followed *muis 'mouse'*). A unique order of displays was created for each participant with the Mix program (van Casteren & Davis, 2006), with the constraints that homogeneous and heterogeneous trials alternated, trials from the same semantic category were not presented consecutively, and the last picture on a trial was not the same as the first picture on the next trial.

Procedure

Participants were tested on the web[9] with the instructions that they should perform this experiment in a quiet room with the door shut and with potentially distracting electronic equipment turned off. They were told to imagine that they were in a laboratory during the experiment. We asked for permission to record before the test began. At the beginning of the test, participants were asked to familiarize themselves with all pictures and name them quickly in Dutch. Familiarization trials began with a fixation cross presented for 500 ms, followed by a blank screen for 500 ms. Then, a picture appeared on the screen for a 2-second period during which participants were asked to name the picture in Dutch as quickly and accurately as possible. Finally, a blank screen was presented for 1500 ms before the start of the next trial.

A practice session of four trials was followed by the three blocks of experimental trials. Participants took a short break after each block of sixteen trials. The whole experiment lasted 30 minutes. Practice and experimental trials began with a fixation cross presented for 500 ms, followed by a blank screen for 500 ms. Then a 4 × 4 picture grid appeared on the screen in which sixteen pictures were presented simultaneously for up to 30 seconds. Participants named the sixteen pictures one by one in order from left to right starting with the first row as quickly

---

[9] Here is an example of the experiment for one participant: https://frinexproduction.mpi.nl/image_naming_experiment/?stimulusList=List1

and accurately as possible. They ended the trial with a mouse click. If they had not finished within 30 seconds, the picture grid disappeared automatically. A blank screen was presented for 1500 ms before the onset of the next trial. An example of a trial is shown in Figure 3.1.

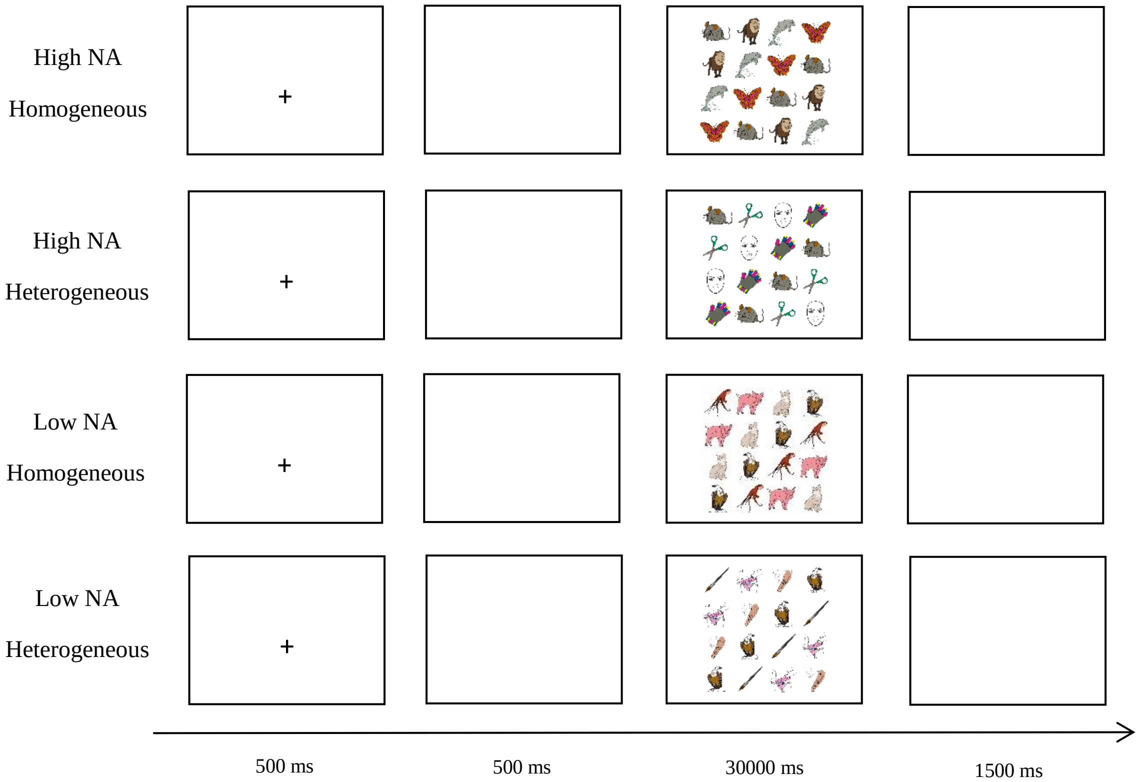

Figure 3.1. Trial examples of four conditions, NA refers to name agreement. Heterogeneous sets include one picture from each homogeneous set.

Analysis

Five main dependent variables were coded to index naming performance. Production *accuracy* indexes the proportion of trials where all sixteen pictures were named correctly. Participants were not presented with the expected names of the pictures in the familiarization stage, as it was impossible to give them timely feedback on their naming responses and we did not want to ask them to use picture names they would not spontaneously use. Therefore, we later coded any reasonable naming responses as correct. Picture names were coded as correct if they matched any of the multiple names given to the picture in the MultiPic database (Duñabeitia et al., 2018); if they were diminutive versions of the multiple names (e.g., *big 'piglet'* was named as *biggetje ‘little piglet’*), or if they were judged reasonable by trained research assistants (e.g., *gier ‘vulture’* was named as *havik ‘hawk’*).

For trials where all pictures were named sensibly and without hesitations or self-corrections (hereafter, “fully correct trials”), we calculated two main time measures. *Utterance duration* was defined as the time interval between the utterance onset of the first picture name and the utterance offset of the sixteenth picture name. This reflects how long participants took to produce all sixteen picture names. *Total pause time* was defined as the sum of all pauses between picture names. This reflects the planning done between producing responses.

For these fully correct trials, we also examined how participants chunked or grouped their sixteen responses. Since earlier studies of spontaneous speech coded silent durations longer than 200 ms as silent pauses (e.g., Heldner & Edlund, 2010), we coded the responses that occurred with 200 ms or less between them as a single response chunk. *Total chunk number* refers to how many response chunks participants made on one trial, with a larger number of total response chunks meaning more separate planning units for production. *First chunk length* refers to how many names participants produced in their initial response, and provides a measure of how much information participants planned before starting to speak. In addition to the five primary measures of interest, we also measured *onset latency*, defined as the interval from the onset of stimulus presentation to the onset of the utterance, which indexes the beginning stages of speech planning.

Bayesian mixed-effect models were conducted to assess the likely magnitude of the effects and quantify the size of parameters and the uncertainty around them (Nicenboim & Vasishth, 2016). Bayes factors were computed to evaluate the evidence in favor of or against the effects. For these analyses, we used R version 4.0.3 (R Core Team, 2020) with the package *brms* (version 2.14.4; Bürkner, 2018).

Bayesian mixed-effect models. For all Bayesian mixed-effect models, predictors were name agreement (high / low) and semantic context (homogeneous / heterogeneous), which were both contrast coded with (0.5, -0.5). The random effect structure for the models included random intercepts for participants and items, and did not include any random slopes because of the small number of observations (four per block) for each condition of each participant (Barr et al., 2013). Separate models were fitted for each dependent measure. All models had four chains and each chain had 4000 to 7000 iterations depending on model convergence (listed in model output tables). We used a warm-up (or burn-in) period of 1000 iterations in each chain, which means we removed the data based on the first 1000 iterations in order to correct the initial sampling bias.

All models used weak, widely spread priors that would be consistent with a range of null to moderate effects. The model of accuracy used family *bernoulli* combined with a *logit* link, and the model used a student-*t* prior with 1 degree of freedom and a scale parameter of 2.5. The model of log-transformed utterance duration used weak normal priors with an SD of 0.2, and the model of log-transformed total pause time had a weak normal prior with an SD of 1. Both were performed using the family *gaussian* combined with *identity* link. For chunk measures (i.e., total chunk number, first chunk length), the models had weak normal priors centered at zero with an SD of 3, and used the family *poisson* combined with the *log* link. In addition, the model of log-transformed onset latency used weak normal priors with an SD of 0.2, and used the family *gaussian* combined with *identity* link. All models were run until the R hat value for each estimated parameter was 1.00, indicating full convergence. Analyses of posterior distributions given different prior distributions indicate that these priors were appropriate (see https://osf.io/6jg4p/ for details).

For these models, the size of reported betas reflects estimated effect sizes, with larger absolute values of betas reflecting larger effects. We reported the parameters for which 95% Credible Intervals (hereafter, Cr.I) do not contain zero, which is analogous to the frequentist null hypothesis significance test: the parameter has a non-zero effect with high certainty. We also reported any parameters for which the point estimate for the beta is about twice the size of its error, as this also provides evidence for an effect: the estimated effect is large compared to the uncertainty around it. We also reported the posterior probability of the weak effects, indicating the proportion of samples with a value equal to or above the beta estimate.

Bayes factors. Bayes factors provide a way to quantify the evidence a data set provides in favor of one model over another. Although Bayes factors are defined on a continuous scale, several researchers have proposed to subdivide the scale in discrete evidential categories (e.g., Lee & Wagenmakers, 2014; Schönbrodt & Wagenmakers, 2018), which we report below. To obtain Bayes factors, we computed a series of reduced models eliminating each effect of interest one at a time, and then compared the reduced and full model using bridge sampling. These models used the same priors as the Bayesian mixed-effect models, but with a higher number of iterations, i.e., 20000. Sensitivity analyses suggest that the priors we selected were reasonable for this analysis, though they did have a moderate effect on the Bayes factor for the name agreement effect on log-transformed utterance duration (see https://osf.io/6jg4p/ for details).

Analyses without the first row. Before data collection, we also planned to conduct an additional set of analyses where four dependent variables (i.e., accuracy, utterance duration, total pause time, and total chunk number) were calculated from the onset of naming the fifth picture (i.e., from the second row). This was done because the semantic context effect often arises from the second cycle (analogous to the second row of pictures in our study) and stays stable over subsequent cycles (Belke, 2017; Belke et al., 2005; Damian & Als, 2005).

3.3 Results

One participant was removed from further analyses because their responses were not recorded. The data from the remaining 40 participants was checked for errors, removing from analysis any trials with implausible names (e.g., *handschoen ‘glove’* misnamed as *jas ‘coat’*), hesitations (e.g., *sok ‘sock’* named as *sss...sok*), self-corrections (e.g., *oor ‘ear’* named as *neus...oor ‘nose...ear’*) and any trials where objects were omitted or named in the wrong order. Two more participants were then excluded because of high error rates (> 60%), following exclusion criteria we set before data collection. For the remaining 38 participants, the exclusion of inaccurate trials resulted in a loss of 12.17% of the data (range by participants: 0 - 37.5% of removed trials). Finally, any data points that were more than 2.5 standard deviations below or above the participant mean were removed for time measures (0.12 % for log-transformed utterance duration, 2.31% for log-transformed total pause time, and 0.81% for log-transformed onset latency). Descriptive statistics of all dependent variables are shown in Table 3.1.

Accuracy. Participants produced the intended responses on 88% of the naming trials. As shown in Tables 3.1 and 3.2, a Bayesian mixed-effect model showed that accuracy was not influenced by name agreement, but it was considerably lower in the homogeneous condition than in the heterogeneous condition (β = -0.379, SE = 0.188, 95% Cr.I = [-0.753, -0.015]). Name agreement and semantic context did not interact. However, as shown in Table 3.3, Bayes factors showed only weak evidence in favor of the name agreement effect (BF = 1.75), and presented moderate evidence for the semantic context effect (BF = 3.64). There was only weak evidence against the interaction between name agreement and semantic context (BF = 0.86). In short, accuracy was somewhat affected by semantic context but not affected much by name agreement.

Utterance duration. As shown in Figure 3.2 and Table 3.2, a Bayesian mixed-effect model showed that log-transformed utterance duration was significantly longer for low name agreement pictures than for high name agreement pictures (β = -0.055, SE = 0.018, 95% Cr.I = [-0.091, -0.019]), but did not vary by semantic context. Name agreement and semantic context did not interact. Correspondingly, as shown in Table 3.3, Bayes factors showed moderate evidence in favor of the name agreement effect (BF = 7.60), but presented moderate evidence against the semantic context effect (BF = 0.22). There was moderate evidence against the interaction between name agreement and semantic context (BF = 5.49). In sum, utterance duration was affected by name agreement only.

Table 3.1. Means and standard deviations of the dependent variables calculated from trial onset (i.e., the start of the first picture) by name agreement and semantic context.

	High name agreement		Low name agreement	
	homogeneous	heterogeneous	homogeneous	heterogeneous
Accuracy	85	88	87	91
Utterance duration (ms)	10424 (2628)	10152 (2560)	10960 (2636)	10762 (2621)
Total pause time (ms)	2579 (2012)	2339 (1991)	3022 (2049)	2855 (2007)
Total chunk number	5.3 (3.3)	5.1 (3.4)	6.1 (3.5)	5.8 (3.5)
First chunk length	5.2 (4.0)	5.2 (4.1)	4.3 (3.3)	4.5 (3.7)
Onset latency (ms)	1355 (385)	1312 (364)	1441 (447)	1415 (437)

Note. Standard deviations are given in parentheses. All time and chunking measures reflect fully correct trials only.

Total pause time. As shown in Figure 3.2 and Table 3.2, a Bayesian mixed-effect model showed that log-transformed total pause time was longer for low name agreement pictures than for high name agreement pictures (β = -0.254 , SE = 0.057, 95% Cr.I = [-0.366, -0.143]). There was moderate evidence for a semantic context effect (β = 0.108, SE = 0.057, 95% Cr.I = [-0.005, 0.22]). Note that while the 95% Cr.I contains zero, the point estimate is high relative to the error around it, and 97% of the posterior distribution around the estimated effect is above zero. This demonstrates that log-transformed total pause time was longer in the homogeneous than in the heterogeneous conditions. Again, name agreement and semantic context did not interact.

Bayes factors showed a slightly different pattern: as shown in Table 3.3, Bayes factors showed extreme evidence in favor of the name agreement effect (BF = 343.85)[10], but only weak evidence against the semantic context effect (BF = 0.40). There was moderate evidence against the interaction between name agreement and semantic context (BF = 7.85). Thus, consistent with the results of utterance duration, total pause time was affected by name agreement only.

Total chunk number. As shown in Figure 3.3 and Table 3.2, a Bayesian mixed-effect model showed that participants grouped their responses in more chunks for low name agreement pictures than high name agreement pictures (β = -0.139, SE = 0.038, 95% Cr.I = [-0.214, -0.063]). Total chunk number was not impacted by semantic context, with no interaction between name agreement and semantic context. Bayes factors showed the same pattern, as shown in Table 3.3, with moderate evidence in favor of the name agreement effect (BF = 6.34), but moderate evidence against the semantic context effect (BF = 0.03). There was very strong evidence against the interaction between name agreement and semantic context (BF = 38.32). In sum, again, total chunk number was influenced by name agreement only.

First chunk length. As shown in Figure 3.3 and Table 3.2, a Bayesian mixed-effect model showed that participants planned fewer names in their first response chunk for low name agreement pictures than high name agreement pictures (β = 0.172, SE = 0.057, 95% Cr.I = [0.059, 0.282]), but first chunk length was not impacted by semantic context and there was no interaction between name agreement and semantic context. As shown in Table 3.3, Bayes factors showed a matching pattern: only weak evidence in favor of the name agreement effect (BF = 1.55), and moderate evidence against the semantic context effect (BF = 0.02). There was strong evidence against the interaction between name agreement and semantic context (BF = 24.34). Thus, first chunk length appeared to depend on name agreement, but not semantic context.

Onset latency. As shown in Table 3.2, a Bayesian mixed-effect model showed that log-transformed onset latency was longer for low than high name agreement pictures (β = -0.055, SE = 0.013, 95% Cr.I = [-0.079, -0.03]). There was moderate evidence for a semantic context effect (β = 0.025, SE = 0.013, 95% Cr.I = [-0.001, 0.05]). Note that while the 95% Cr.I contains zero, the point estimate is high relative to the error around it, and 97% of the posterior

---

[10] Changing this prior to something less informative reduces this Bayes factor, but still shows strong or moderate evidence in favor of the effect. See https://osf.io/6jg4p/ for details.

distribution around the estimated effect is above zero. This demonstrates that log-transformed onset latency was longer in the homogeneous context than in the heterogeneous context. Name agreement and semantic context did not interact. Bayes factors showed a slightly different pattern: as shown in Table 3.3, Bayes factors showed extreme evidence in favor of the name agreement effect (BF = 340.22), but presented only weak evidence against the semantic context effect (BF = 0.47). There was moderate evidence against the interaction between name agreement and semantic context (BF = 7.36). Thus, the results observed for onset latency matched those obtained for the remaining dependent variables: name agreement had an impact, but semantic context did not.

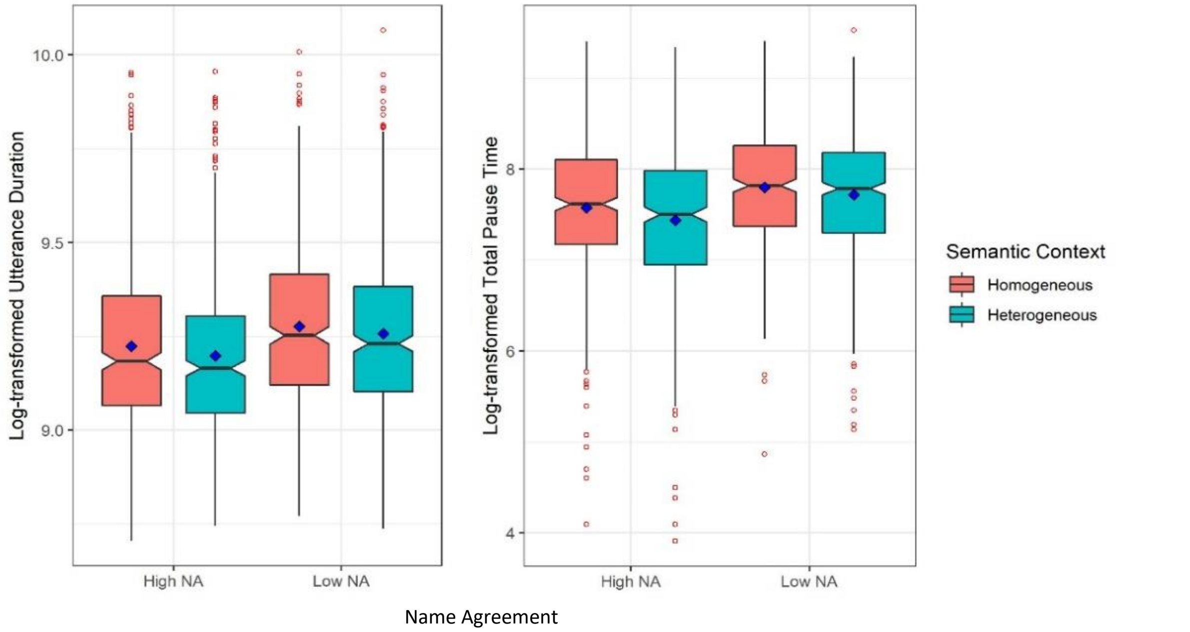

Figure 3.2. Log-transformed utterance duration (left) and log-transformed total pause time (right) calculated from trial onset split by name agreement (NA: high, low) and semantic context (homogeneous, heterogeneous). Blue squares represent condition means and red points reflect outliers.

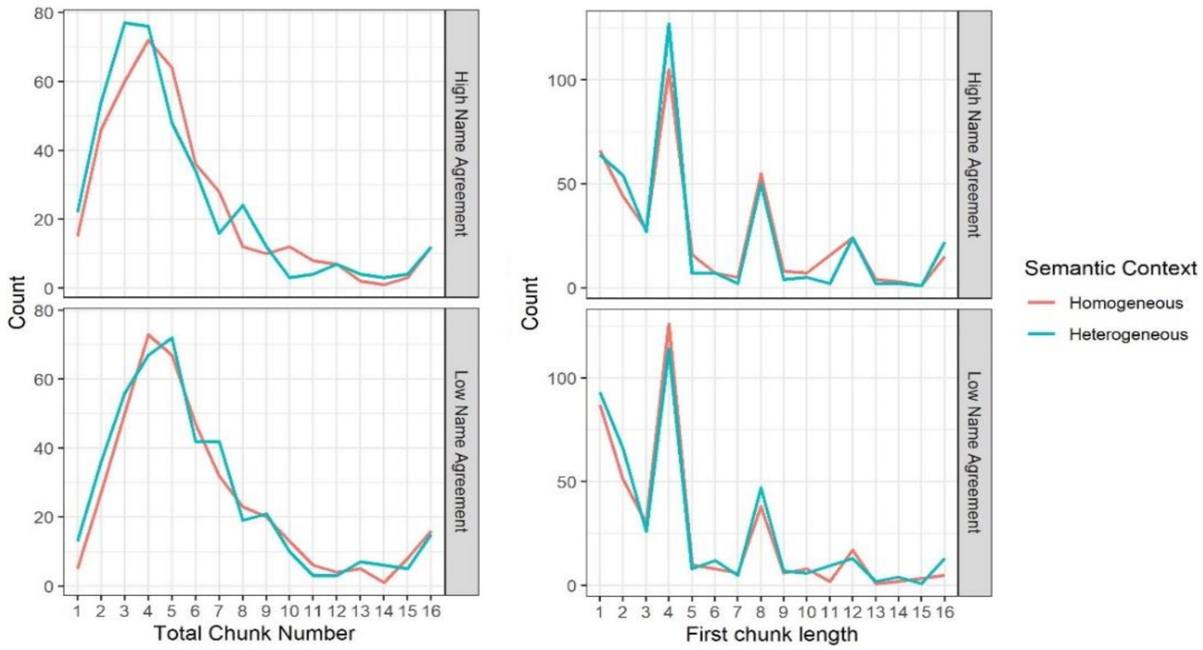

Figure 3.3. Total chunk number (left) and first chunk length (right) calculated from trial onset split by name agreement (high, low) and semantic context (homogeneous, heterogeneous).

Table 3.2. Results of Bayesian mixed-effect models for all dependent variables calculated from trial onset.

		Estimate	Est.error	95% Cr. I		Effective samples
				lower	upper	
Accuracy						
Population-level effects	Intercept	2.27	0.177	1.936	2.636	3803
	Name Agreement	-0.309	0.186	-0.677	0.052	10504
	Semantic Context	**-0.379**	**0.188**	**-0.753**	**-0.015**	**9697**
	NA × SC	0.238	0.375	-0.5	0.972	9925
Group-level effects	Participant_sd(Intercept)	0.853	0.147	0.604	1.173	4228
	Item_sd(Intercept)	0.34	0.144	0.042	0.619	2278
Log-transformed utterance duration						
Population-level effects	Intercept	9.242	0.033	9.176	9.305	1593
	Name Agreement	**-0.055**	**0.018**	**-0.091**	**-0.019**	**4057**
	Semantic Context	0.024	0.018	-0.012	0.059	3865
	NA × SC	0.008	0.036	-0.063	0.078	3891
Group-level effects	Participant_sd(Intercept)	0.189	0.023	0.151	0.242	2526
	Item_sd(Intercept)	0.06	0.007	0.047	0.075	4494
Log-transformed total pause time						
Population-level effects	Intercept	7.633	0.101	7.435	7.839	552
	Name Agreement	**-0.254**	**0.057**	**-0.366**	**-0.143**	**2703**
	Semantic Context	*0.108*	*0.057*	*-0.005*	*0.22*	*2581*
	NA × SC	0.06	0.112	-0.162	0.282	2970
Group-level effects	Participant_sd(Intercept)	0.592	0.072	0.466	0.749	1382
	Item_sd(Intercept)	0.176	0.024	0.135	0.227	3224
Total chunk number						
Population-level effects	Intercept	1.62	0.075	1.475	1.769	654
	Name Agreement	**-0.139**	**0.038**	**-0.214**	**-0.063**	**3889**

	Semantic Context	0.045	0.038	-0.031	0.12	3597
	NA × SC	0.016	0.078	-0.135	0.174	3461
Group-level effects	Participant_sd(Intercept)	0.439	0.054	0.347	0.558	1331
	Item_sd(Intercept)	0.109	0.018	0.077	0.147	3944
First chunk length						
Population-level effects	Intercept	1.436	0.092	1.251	1.617	690
	Name Agreement	**0.172**	**0.057**	**0.059**	**0.282**	**2749**
	Semantic Context	-0.009	0.058	-0.122	0.102	2601
	NA × SC	0.052	0.115	-0.174	0.284	2730
Group-level effects	Participant_sd(Intercept)	0.533	0.067	0.418	0.682	1285
	Item_sd(Intercept)	0.182	0.024	0.14	0.234	3412
Log-transformed onset latency						
Population-level effects	Intercept	7.198	0.028	7.141	7.253	1130
	Name Agreement	**-0.055**	**0.013**	**-0.079**	**-0.03**	**10977**
	Semantic Context	*0.025*	*0.013*	*-0.001*	*0.05*	*11029*
	NA × SC	0.011	0.025	-0.038	0.06	11221
Group-level effects	Participant_sd(Intercept)	0.167	0.021	0.131	0.213	2336
	Item_sd(Intercept)	0.021	0.01	0.002	0.039	2835

Note. Models for log-transformed total pause time and total chunk number were run for 5000 iterations, model for log-transformed utterance duration was run for 7000 iterations, and models for other dependent variables were run for 4000 iterations. Bolded values indicate effects where the 95% Cr.I does not contain zero; Italicized values indicate effects where the beta estimate is twice the estimate of the standard error. NA refers to name agreement, SC refers to semantic context.

Table 3.3. Bayes factors for all dependent variables calculated from trial onset.

	NA effect	SC effect	Null Interaction
Accuracy	1.75	**3.64**	0.86
Log-transformed utterance duration	**7.60**	*0.22*	**5.49**
Log-transformed total pause time	**343.85**	0.40	**7.85**
Total chunk number	**6.34**	*0.03*	**38.32**
First chunk length	1.55	*0.02*	**24.34**
Log-transformed onset latency	**340.22**	0.47	**7.36**

Note. NA refers to name agreement, SC refers to semantic context. Bolded values indicate moderate or above evidence in favor of the effects (BF > 3); Italicized values indicate moderate or above evidence against the effects (BF < 1/3); Regular values indicate only weak evidence in favor of or against the effects (1/3 < BF < 3).

Results from the onset of naming the fifth picture

Recall that earlier studies showed that semantic context effects are typically not seen when the pictures of a set are named for the first time (e.g., Belke, 2017; Belke et al., 2005; Damian & Als, 2005). As shown in Figure 3.4, our results are, at least descriptively, consistent with this pattern. Semantic context effects were not present when participants named the first row of objects, but appeared in the following rows. Analyses for the data set without the first row were conducted to assess the semantic context effect from the second row onwards. As the results were largely comparable to the full data set, we only report differences from the main analyses. See Appendix B for full details of each analysis.

Bayesian mixed-effect models showed that semantic context did not influence accuracy, but affected log-transformed utterance duration (β = 0.038, SE = 0.016, 95% Cr.I = [0.006, 0.071]), log-transformed total pause time (β = 0.17, SE = 0.075, 95% Cr.I = [0.023, 0.318]), and total chunk number (β = 0.070, SE = 0.034, 95% Cr.I = [0.003, 0.136]) (see Table B1). However, Bayes factors slightly contradicted these analyses (see Table B2): There was only weak evidence in favor of semantic context effects on the time measures (1/3 < BFs < 3). There was moderate evidence against the semantic context effect on total chunk number (BF = 0.10). Thus, even when the first row was excluded from the analyses, there was at best weak evidence for semantic context effects on any of the dependent measures.

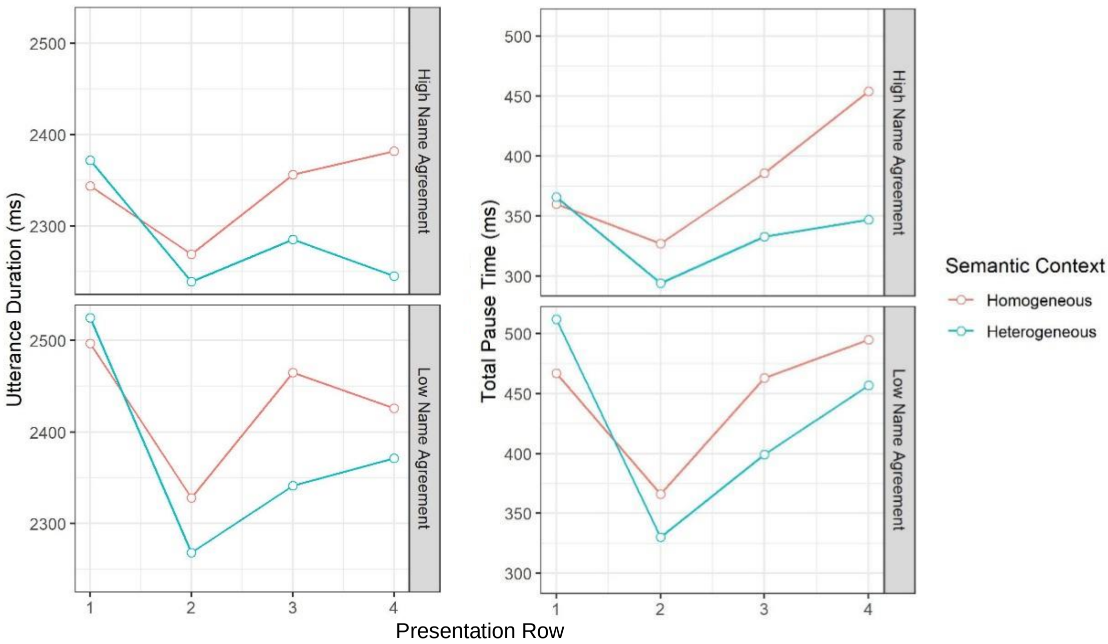

Figure 3.4. Utterance duration (left) and total pause time (right) in each row split by name agreement (high, low) and semantic context (homogeneous, heterogeneous).

Post-hoc power analyses

To test whether the weak semantic context effects and null interaction were due to relatively small sample size in our study, we conducted a post-hoc power analyses at different sample sizes by using lme4 package (Bates et al., 2015) in R version 4.0.3 (R Core Team, 2020). For time measures, separate linear mixed-effect models with the same structure as the Bayesian mixed-effect models were performed. In each estimation, 86% of items (i.e., 40 trials) were included, and actual values of means and standard deviations in each condition were used. The number of simulations was 1000. To obtain power values, we compared the model with each effect of interest and the one without the effect (see https://osf.io/6jg4p/ for details). As shown in Figure 3.5 (left), power values for the semantic context effects on time measures were relatively low for 38 participants (Powers < 0.5), while the values would be larger than 0.8 when testing at a minimum of 84 participants. This finding suggests that reliable semantic context effects can be detected for a large sample size. However, the power values for the interaction between name agreement and semantic context on time measures (see Figure 3.5, right) were extremely low even for a large enough sample size (e.g., Powers < 0.14 for 200 participants), which suggests that the null interaction cannot be attributed to the relatively small sample size in our study. Since the results for time measures calculated from the onset of naming the fifth picture are largely comparable to those from trial onset, we report them in Appendix C (see Figure C1).

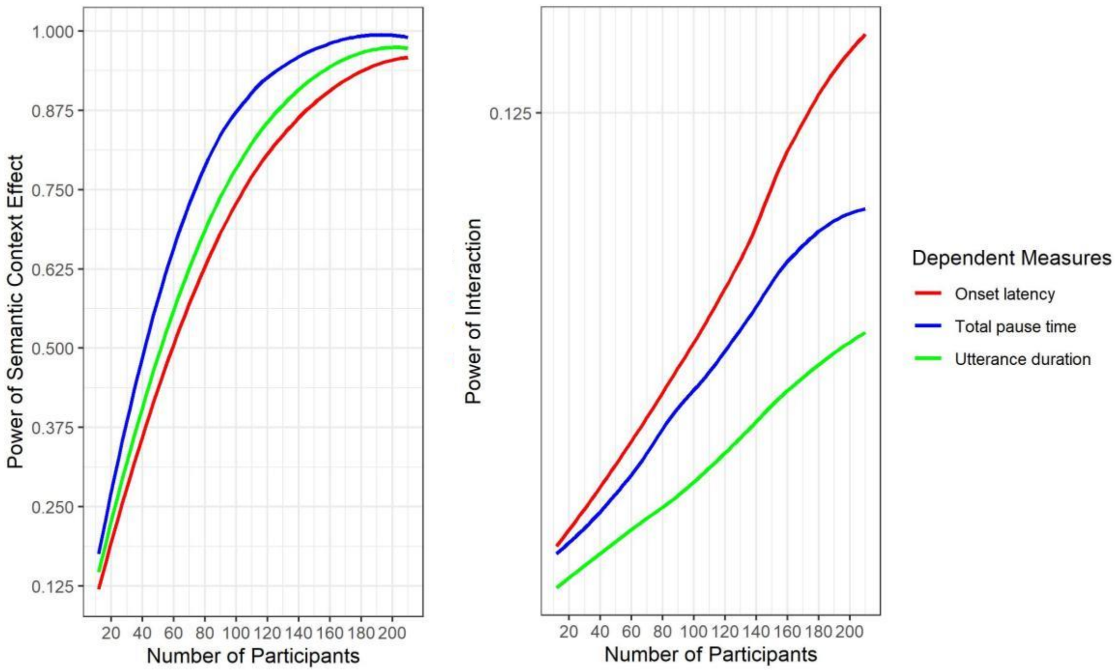

Figure 3.5. Results of post-hoc power analyses for the semantic context effects (left) and the interaction between name agreement and semantic context (right) on time measures calculated from trial onset.

3.4 Discussion

The present study investigated the feasibility of conducting spoken language production research in an online environment. We specifically explored the usefulness of measuring multiple dependent variables. We examined two previously studied effects related to lexical selection— semantic context and name agreement— in a modified blocked-cyclic naming paradigm. Six dependent variables were measured: naming accuracy, utterance duration, total pause time, total chunk number, first chunk length, and onset latency. We found strong evidence for name agreement effects, but little evidence for semantic context effects or interactions of the two variables. In this discussion, we comment on these findings, focusing primarily on their methodological implications.

As predicted, we found robust name agreement effects on all measures except accuracy, with longer speech onset latencies, utterance durations and pause times, more response chunks, and shorter first chunk length for the naming of low name agreement pictures than high name agreement pictures. These results suggest that participants achieved lexical selection for the object names incrementally, at several time points during the process of multiple-object naming, and that they tended to plan their speech more sequentially with audible pauses between their responses when speech planning demands was high. These findings are important, as they suggest that measures of utterance durations and speech fluency can be exploited to study lexical access of speech production, in addition to, or instead of speech onset latencies. Of course, the sensitivity of utterance durations and speech fluency to the duration of cognitive processes underlying speech planning is not a new insight. For instance, some of the earliest theories of speech planning relied on analyses of pauses and disfluencies (e.g., Goldman-Eisler, 1972; Levelt, 1989), and, as described earlier, the RAN paradigm (Denckla & Rudel, 1976) that is often used in reading research measures total utterance durations (e.g., Araújo et al., 2020). The present study therefore may be seen as a reminder of the usefulness of these dependent variables to complement measurement of speech onset latencies. In interpreting experimental findings, it is, of course, always important to keep in mind that every dependent measure, be it speech onset latency or utterance duration, is likely to be affected by multiple influences. Speech onset latencies may, for instance, reflect not only on the time required to retrieve the first object name, but also on the time required for any advance planning of the following object names a participant may engage in. Similarly, total utterance durations will not only depend on the retrieval times for all object names but also on the strategies participants use to coordinate speech planning and speaking. Because speech planning can happen during articulation, utterance duration may be less sensitive to the effects of planning difficulty than onset latencies.

In this web-based paradigm we did, somewhat unexpectedly, observe robust evidence for name agreement effects on speech onset latencies, which replicates the effects of lab-based studies (e.g., Alario et al., 2004; Shao et al., 2014). Thus, our initial concern that speech onset latencies would be unreliable turned out to be unwarranted. Other recent studies using internet-based paradigms have provided similar evidence for the reliability of onset latencies, as they replicated several key findings of the speech production literature, including the word frequency effect (Fairs & Strijkers, 2021), the cumulative semantic interference effect (Stark et al., 2022), and the semantic interference effect in the picture-word interference paradigm (Vogt et al., 2022). Fairs and Strijkers (2021) compared the results of their web-based study to those of an otherwise identical study run in the laboratory. They found overall longer latencies in the web-based study but no difference in the size of the word frequency effect. Similarly, Stark and colleagues (2022) reported cumulative semantic interference effect comparable to effects found in earlier lab-based studies. In short, there is now good evidence that speech onset latencies can be recorded with good accuracy in web-based language production studies.

To return to our study, when the dependent variables were calculated from trial onset, semantic context only affected accuracy and total pause time. By contrast, when the dependent variables were calculated from the onset of naming the fifth picture (the first one in the second row), semantic context effects were found for all dependent variables except accuracy. This pattern is consistent with earlier lab-based studies using the classic blocked-cyclic naming paradigm (with one picture being displayed and named per trial) and showing that semantic context effects are only obtained from the second naming cycle onwards (e.g., Belke, 2017; Belke et al., 2005; Damian & Als, 2005). However, in our experiment, Bayes factors showed only weak evidence in favor of these semantic context effects on any measure except accuracy (BFs < 3). This suggests that the semantic context effects in our web-based study were relatively weak.

There are a number of reasons why the semantic context effects may have been weak. First, it could be that the simultaneous presentation of objects, compared to the sequential presentation, increased facilitatory conceptual or repetition priming effects and counteracted the inhibitory semantic context effects (as would be consistent with Abdel Rahman & Melinger, 2009; Howard et al., 2006; Oppenheim et al., 2010). This implies that semantic context effects might always be weak when the pictures are shown simultaneously. The effects of simultaneous versus successive presentation of pictures on the occurrence of semantic context effects should be further investigated. More generally, the timing of picture presentation (simultaneous, successive at a rapid or fast pace) may affect speakers' memory for the pictures already named and their planning for upcoming pictures, which should be kept in mind when designing a study.

Second, compared with onset latencies, measures of utterance durations and speech fluency during multiple object naming may be less sensitive to semantic context, or to any other variable affecting the speed of lexical access. Consistent with this proposal, Belke and Meyer (2007) found a robust semantic context effect on onset latencies, a small semantic context effect on word durations, but no effect on pause rates for the young speakers in their study. Semantic context effects may be hard to detect in measures of utterance durations and speech fluency because these measures depend not only on lexical access times, but also on multiple other variables, including the time required for phonetic planning, prosodic planning, and articulation, which may vary from trial to trial. Thus, while speech durations and speech fluency can be exploited to assess the speed of word planning processes, subtle effects on word planning times may be obscured by other influences.

In addition, we found that semantic context did not interact with name agreement on any dependent variable, with Bayes factors showing moderate evidence or better (BFs > 3 for null interactions on all measures except accuracy). This might reflect that semantic context effects are not modulated by name agreement, suggesting that lexical selection can be achieved without competition, in line with the model proposed by Oppenheim and colleagues (2010). Alternatively, the interaction, just like the main effect of semantic context, may have been too subtle to be detected in analyses of utterance durations and speech fluencies.

A robust semantic context effect or an interaction between name agreement and semantic context may have been obtained with a larger sample size. We determined our sample size in terms of previous work: by collecting data from 41 participants, we doubled the number of participants tested in most lab-based semantic context experiments recording speech onset latencies (about 20 participants; e.g., Belke & Meyer, 2007; Damian & Als, 2005). A power simulation for determining sample size before the present study was not possible, as no comparable studies were available. However, we conducted post-hoc power calculations based on our results (see Figure 3.5), which suggest that robust semantic context effects indeed can be detected when testing at a minimum of 84 participants especially on total pause time. However, the interaction of semantic context and name agreement seems to be non-existent even for a large enough sample size (e.g., 200 participants). The results of post-hoc power analyses can now be used for a power simulation to estimate the sample size needed to observe effects of interest in future work.

In sum, we found strong evidence for name agreement effects, but weak evidence for semantic context effects. This pattern is consistent with the observation that name agreement effects on speech onset latencies tend to be descriptively larger than semantic context effects (e.g., Damian et al., 2001; Shao et al., 2014; Shao et al., 2015). Moreover, unlike semantic context effects, name agreement effects do not hinge on relationships between successive object names and consequently may be less likely to be affected by the timing of stimulus presentation.

Given the relative novelty of web-based studies of language production, we close by briefly commenting on the general quality of the data. It has been argued that the data quality of web-based experiments may be affected by poor compliance or distraction (e.g., Jun et al., 2017), and Fairs and Strijkers (2021) reported that 22% of their participants did not comply with the instructions. Other studies have shown no evidence for decreased attention and have demonstrated comparable data quality for web-based and lab-based studies (e.g., Casler et al., 2013; de Leeuw & Motz, 2016). Our results are consistent with the latter findings. There is little reason to assume that the participants in a web-based study will generally be less engaged or attentive than they would be in a laboratory setting. The speech recordings contained clearly articulated naming responses, no noise in the audio files, and little within-participant variation in the length of audio files per trial. Moreover, we had a much lower rate of participant dropout than earlier web-based studies, which reported dropout rates of over 30% (e.g., Sauter et al., 2020; Zhou & Fishbach, 2016). In our study, only 3 out of 41 participants (7.3%) were excluded from the analyses, one for technical reasons (the computer failed to record their speech responses) and two because they showed low overall accuracy (less than 40% correct responses). Unlike other web-based studies that used crowd-sourcing marketplaces such as Amazon Mechanical Turk (e.g., Anwyl-Irvine et al., 2021; Schnoebelen & Kuperman, 2010), we recruited participants from the pool of individuals that we also use for lab-based studies. They are generally highly motivated and often have experience in participating in psycholinguistic studies. This most likely helped to ensure high-quality data collection. More generally, the success of an experiment, be it laboratory or web-based, depends on the adequate selection, instruction and motivation of the participants. There is no reason to assume that web-based experiments necessarily yield data of poorer quality than lab-based experiments do.

To conclude, the present study, along with several others, supports the feasibility of conducting spoken language production research on web-based platforms. Speech onset latencies turned out to be more reliable than we had assumed. Moreover, the durational properties of multi-word utterances such as utterance duration and speech fluency can be measured to examine processing times for lexical access. These measurements, therefore, are promising dependent variables for future spoken language production research with a modified blocked-cyclic naming paradigm, at least for research questions concerning variations in the speed and success of lexical access. Overall, this study supports the validity of the modified blocked-cyclic naming paradigm as one that is more similar to real-world speaking relative to a classic single picture naming paradigm.

Combined, the present study suggests that web-based studies are a promising addition or alternative to lab-based research. They can be used not only when there are travel restrictions or mobility issues for experimenters, but also to reach groups of participants who may be reluctant or unable to visit a lab. In short, they may contribute to rendering psycholinguistics a more inclusive field.

Acknowledgments

We would like to thank Maarten van den Heuvel for programming; Carlijn van Herpt for double checking stimuli; Annelies van Wijngaarden, Inge Pasman, Dennis Joosen, Carlijn van Herpt, and Esther de Kerf for data coding.

Appendices

Appendix A: Stimuli in the present study

Table A1. Dutch names of pictures per condition. For low name agreement pictures, first and second modal names are given. English translations appear in parentheses.

	High Name Agreement	Low Name Agreement
Animal	dolfijn (dolphin)	varken / big (pig / piglet)
	vlinder (butterfly)	inktvis / octopus (squid / octopus)
	muis (mouse)	gier / aasgier (vulture / Egyptian vulture)
	leeuw (lion)	kat / poes (cat / cat)
Body Parts	oor (ear)	been / bovenbeen (leg / thigh)
	neus (nose)	kies / tand (molar / tooth)
	gezicht (face)	vinger / wijsvinger (finger / index finger)
	skelet (skeleton)	lippen / mond (lips / mouth)
Clothing	handschoen (glove)	trui / hoodie (sweater / hoodie)
	sok (sock)	blouse / overhemd (blouse / shirt)
	broek (pants)	schoen / wandelschoen (shoe / hiking boot)
	masker (mask)	luier / pamper (diaper / pamper)
Tool	kam (comb)	kwast / penseel (brush / paintbrush)
	schaar (scissors)	mes / zakmes (knife / pocket knife)
	hamer (hammer)	hengel / vishengel (rod / fishing rod)
	weegschaal (weighing scale)	tuinslang / slang (garden hose / hose)

Appendix B: Results for the analyses without the first row

Table B1. Results of Bayesian mixed-effect models for all dependent variables calculated from the onset of naming the fifth picture.

		Estimate	Est.error	95% CI lower	95% CI upper	Effective samples
Accuracy						
Population-level effects	Intercept	2.499	0.202	2.124	2.915	3733
	Name Agreement	-0.306	0.206	-0.718	0.099	10900
	Semantic Context	-0.243	0.207	-0.658	0.16	10714
	NA × SC	0.344	0.418	-0.503	1.168	11416
Group-level effects	Participant_sd(Intercept)	0.981	0.172	0.691	1.363	3744
	Item_sd(Intercept)	0.425	0.15	0.106	0.72	2874
Log-transformed utterance duration						
Population-level effects	Intercept	8.94	0.031	8.876	9	699
	Name Agreement	**-0.047**	**0.016**	**-0.079**	**-0.015**	**2655**
	Semantic Context	**0.038**	**0.016**	**0.006**	**0.071**	**2338**
	NA × SC	0.021	0.032	-0.042	0.083	2379
Group-level effects	Participant_sd(Intercept)	0.182	0.023	0.143	0.233	1271
	Item_sd(Intercept)	0.051	0.007	0.04	0.066	2995
Log-transformed total pause time						
Population-level effects	Intercept	7.264	0.116	7.038	7.494	1176
	Name Agreement	**-0.211**	**0.076**	**-0.359**	**-0.062**	**5363**
	Semantic Context	**0.17**	**0.075**	**0.023**	**0.318**	**5468**
	NA × SC	-0.068	0.147	-0.36	0.22	5188
Group-level effects	Participant_sd(Intercept)	0.678	0.086	0.532	0.868	1933
	Item_sd(Intercept)	0.187	0.041	0.109	0.271	4212
Total chunk number						
	Intercept	1.38	0.068	1.242	1.511	1211

Population-level effects	Name Agreement	**-0.124**	**0.033**	**-0.191**	**-0.059**	**11383**
	Semantic Context	**0.07**	**0.034**	**0.003**	**0.136**	**10980**
	NA × SC	0.052	0.066	-0.079	0.183	11505
Group-level effects	Participant_sd(Intercept)	0.407	0.05	0.321	0.518	2497
	Item_sd(Intercept)	0.078	0.019	0.043	0.117	4487

Note. The model for log-transformed utterance duration was run for 6000 iterations, and models for other dependent variables were run for 4000 iterations. Bolded values indicate effects where the 95% Cr.I does not contain zero. NA refers to name agreement, SC refers to semantic context.

Table B2. Bayes factors for all dependent variables calculated from the onset of naming the fifth picture.

	NA effect	SC effect	Null Interaction
Accuracy	1.48	0.97	0.66
Log-transformed utterance duration	**5.51**	1.29	**5.07**
Log-transformed total pause time	**3.53**	1.02	**6.03**
Total chunk number	**7.42**	*0.10*	**33.81**

Note. NA refers to name agreement, SC refers to semantic context. Bolded values indicate at least moderate evidence in favour of the effects (BF > 3); Italicized values indicate moderate evidence against the effects (BF < 1/3); Regular values indicate only weak evidence in favour of or against the effects (1/3 < BF < 3).

Appendix C: Results of post-hoc power analyses on time measures calculated from the onset of naming the fifth picture

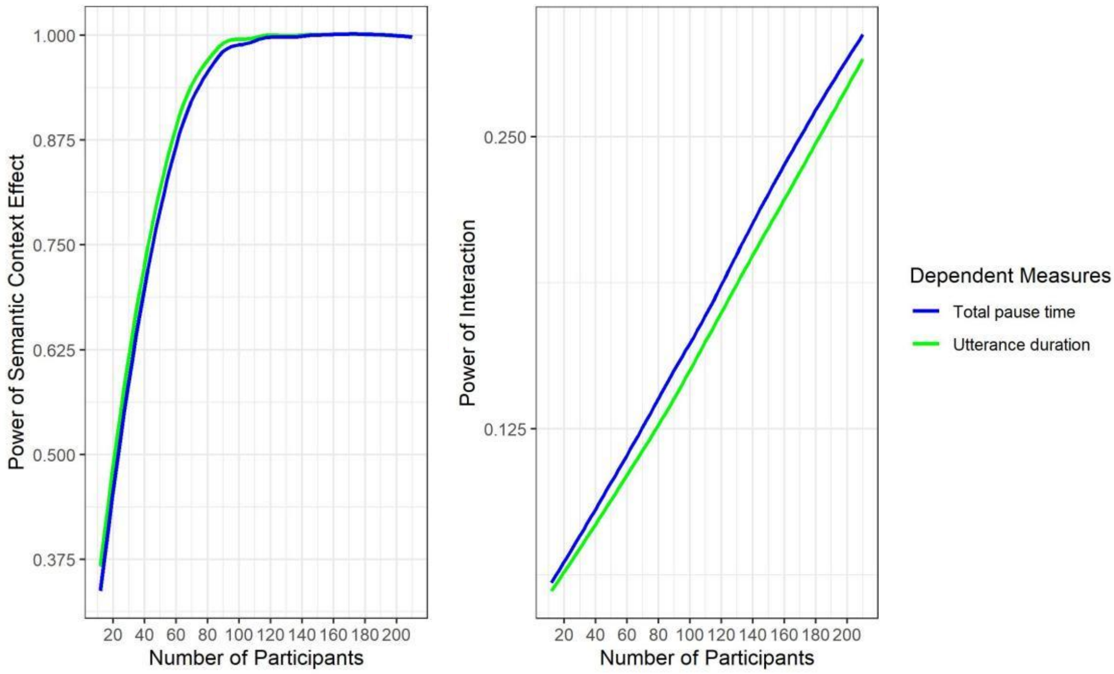

Figure C1. Results of post-hoc power analyses for the semantic context effects (left) and the interaction between name agreement and semantic context (right) on time measures calculated from the onset of naming the fifth picture.

4 | Effects of irrelevant unintelligible and intelligible background speech on spoken language production

Abstract

Speaking in noisy environments (e.g., in a restaurant) is very common. Earlier work has explored speech production during irrelevant background speech such as intelligible and unintelligible word lists (e.g., He et al., 2021, Chapter 2 of this dissertation). The present study compared how different types of background speech (word lists versus sentences) influenced speech production relative to a quiet control condition, and whether the influence depended on the intelligibility of the background speech. Experiment 1 presented native Dutch speakers with Chinese word lists and sentences. Experiment 2 presented native Dutch speakers with Dutch word lists and sentences. In both experiments, the demand of lexical selection in speech production was manipulated by varying name agreement (high versus low) of the to-be-named pictures. Results showed that background speech, regardless of its intelligibility, disrupted speech production relative to a quiet condition, but no effects of word lists versus sentences in either language were found. Moreover, the disruption by intelligible background speech compared to the quiet condition was eliminated when planning low name agreement pictures. These findings suggest that any speech, even unintelligible speech, is harmful for production relative to a quiet condition, which implies that the disruption of speech production is mainly phonological in nature. The disruption by intelligible background speech can be reduced or eliminated via a top-down attention engagement mechanism.

4.1 Introduction

Conversation, an everyday activity requiring the coordination of speech production and comprehension, often takes place in a variety of noisy environments (e.g., background conversations, radio and television broadcasting, or people speaking on the phone) and non-verbal noises (e.g., traffic and construction noises). Compared with extensive work on language comprehension in adverse listening conditions (e.g., Eckert et al., 2016; Vasilev et al., 2019), very little research has investigated how speakers plan their speech in the presence of irrelevant background speech. Understanding speech production in verbal and non-verbal sources of noise would advance our understanding of how speakers shield against auditory disruption when planning their speech. In the present study, we investigated how different types of irrelevant background speech (i.e., word lists versus sentences) influenced speech production with varying lexical selection demands, and whether the influence depended on the intelligibility of the background speech.

Over the past few decades, many studies have found that speech and non-speech sounds disrupt cognitive tasks such as serial recall (e.g., Parmentier & Beaman, 2015; Röer et al., 2014, 2015; Schlittmeier et al., 2012) and reading (e.g., Cauchard et al., 2012; Hyönä & Ekholm, 2016; Yan et al., 2018), even when these sounds are irrelevant for the task and can be ignored. This is referred to as the *irrelevant speech effect* (or *irrelevant sound effect*; Colle & Welsh, 1976; Jones et al., 1992). These studies have discussed two major accounts for the irrelevant speech effect (i.e., the *interference-by-similarity* account and the *attention capture* account), which can also be applied to explain how background speech interferes with speech production. We explain the predictions of both accounts for the interference elicited by background speech.

One reason for interference is the involvement of shared mechanisms or representations between both tasks; this is known as the interference-by-similarity account (e.g., Jones et al., 1993; Martin et al., 1988; Salamé & Baddeley, 1982, 1989). This account was initially proposed to explain the changing-state effect in serial recall. When participants need to recall a visually presented list of items (usually 6 - 8 digits or letters) in the correct order while ignoring task-irrelevant speech, they are impaired more by auditory changing-state sequences consisting of different distractor items (e.g., A B C D E F G H) than steady-state sequences comprising of a repeated distractor (e.g., A A A A A A A A) (Hughes, 2014; Hughes et al., 2007; Jones et al., 1993; Jones et al., 1992). The effect has been attributed to conflict between the intentional processing of the to-be-remembered items' order and the automatic processing of the irrelevant auditory distractors' order (i.e., interference-by-process account; e.g., Hughes, 2014; Jones et al.,1993).

Interference-by-similarity has been extended to reading (Martin et al., 1988; Salamé & Baddeley, 1982, 1989). Previous research has shown that irrelevant background speech impairs reading performance relative to a quiet condition (e.g., Cauchard et al., 2012; Hyönä & Ekholm, 2016; Yan et al., 2018). Two hypotheses attribute this impairment to different sources: similarity in the phonological representations (i.e., *phonological disruption view;* Salamé & Baddeley, 1982, 1989) or shared use of semantic processing (i.e., s*emantic disruption view*, Martin et al., 1988). The phonological disruption view (Salamé & Baddeley, 1982, 1989) hypothesizes that the irrelevant speech effect in reading results from the similarity in content of phonological codes of reading and background speech, which both are buffered in a phonological memory store (a component of the phonological loop; Baddeley, 2000, 2003). Since the phonological loop acts as a filter that lets in speech sounds but filters out non-speech sounds (Salamé & Baddeley, 1987), this view predicts that speech sounds (intelligible or not) would disrupt reading, while non-speech noise would not cause interference because it does not gain access to the phonological loop. This view has received support from studies showing that irrelevant vocal sounds (e.g., singing with/without instrumental accompaniment; Boyle & Coltheart, 1996) or a recording of a busy bar/restaurant (e.g., voices, music, and the clanking or dishes and silverware; Robison & Unsworth, 2015) do not significantly impair reading comprehension relative to a quiet condition.

Martin and colleagues (1988) propose instead that it is semantic processing that disrupts reading comprehension. In their experiments, intelligible speech (English) disrupted reading comprehension in English significantly more than unintelligible speech (Russian), and speech consisting of random words was more disruptive than speech consisting of random nonwords. These results are more consistent with a conflict of semantic processing (i.e., semantic disruption view): as reading for comprehension involves extracting the meaning of the text, the semantic content of the irrelevant speech can interfere with this process. This view thus predicts that disruption should be produced only by meaningful speech (i.e., when it is intelligible) because meaningless speech does not recruit semantic processing. Overall, the interference-by-similarity account (e.g., Jones et al., 1993; Martin et al., 1988; Salamé & Baddeley, 1982, 1989) attributes the irrelevant speech effect to a domain-specific linguistic, such as phonological or semantic, disruption.

A contrasting theoretical account of irrelevant speech effects is a *domain-general attention capture* view, which assumes that irrelevant speech or sound disrupts focal task performance because it diverts attention away from the task (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015). When the focus of attention is captured by task-irrelevant sounds, fewer attentional resources are available and hence task performance is impaired. Attentional capture can be divided into two classes (Eimer et al., 1996): *Aspecific attention capture* occurs when a sound captures attention because of the context in which it occurs, such as the sudden onset of speech following a period of silence (Eimer et al., 1996). This view thus predicts that irrelevant background speech with a varied context (stimuli-aspecific variation, e.g., the presence / absence of pauses) should interfere more with the focal task than irrelevant background speech with a constant context (e.g., continuous speech). Alternatively, *specific attention capture* occurs when the particular content of the sound or speech diverts attention (e.g., Eimer et al., 1996; Röer et al., 2013; Wood & Cowan, 1995). This implies that the attention-diverting power is attributable to the content of the stimulus itself (stimuli-specific variation). According to this specific attention capture view (e.g., Eimer et al., 1996; Röer et al., 2013; Wood & Cowan, 1995), the linguistic richness of background speech (e.g., whether it contains semantic/syntactic integration or not) should not elicit disruption when the background speech is unintelligible because the speech is meaningless to individuals.

The attention capture theory has some support in how irrelevant background speech interferes with serial recall performance. For example, the changing-state effect in serial recall has been suggested to occur because acoustic changes exogenously capture attention (cf. the "orienting response"; Sokolov, 1963), whereas with a steady sound, the capture response rapidly habituates (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015). For reading, Hyönä and Ekholm (2016) have also proposed that irrelevant speech can capture the readers' attention and thus impair performance on a reading task. However, these studies did not make a clear distinction between aspecific and specific attentional capture.

While the earlier work suggests that irrelevant speech clearly affects serial recall and reading, much less is known about whether and how it affects spoken language production. Prior literature has indicated that speech production and comprehension draw upon similar processes/representations for semantics and phonology (e.g., Glaser & Düngelhoff, 1984; Kittredge & Dell, 2016; Mitterer & Ernestus, 2008; Schriefers et al., 1990), and both require attention (Cleland et al., 2006; Lien et al., 2008; Roelofs & Piai, 2011). This implies that the interference-by-similarity (Martin et al., 1988; Salamé & Baddeley, 1982, 1989) and attention capture (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015) mechanisms may play roles in the disruption by irrelevant background speech on speech production, as the background speech varies in its specific linguistic content (e.g., semantics / phonology) and aspecific context (e.g., the presence / absence of pauses).

An earlier study by He and colleagues (2021, Chapter 2 of this thesis) has demonstrated that interference-by-similarity plays an important role in how irrelevant background speech affects speech production. In this study, Dutch speakers named sets of pictures with high versus low name agreement (saying, for instance, *snoepje, troon, kasteel, viool, brievenbus, engel ‘candy, throne, castle, violin, letterbox, angel’*) while ignoring Dutch word lists, Chinese word lists, or eight-talker babble (i.e., language-like noise). Their naming accuracy, onset latencies, utterance duration, total chunk number (groups of words spoken without intervening pause), and first chunk length were measured to index speech production performance. Background speech (Dutch and Chinese word lists) disrupted speech production more than eight-talker babble on all measures except accuracy, and Dutch word lists caused more disruption than Chinese word lists on all time measures. This suggests that more interference on speech production is obtained as the representational similarity between speech production and irrelevant background speech increases.

He et al. (2021) also manipulated the difficulty of speech production by varying name agreement (high, low) of to-be-named pictures. Name agreement is the extent to which participants agree on the name of a picture. Previous studies have found that naming a picture with high name agreement (e.g., the item called *banana*) is faster and more accurate than naming a picture with low name agreement (e.g., the item called *sofa* or *couch;* e.g., Alario et al., 2004; Cheng et al., 2011; Vitkovitch & Tyrell, 1995; Shao et al., 2014). The effect can arise at two levels of speech production: object recognition (due to confusion of what the object should be called) and lexical selection (due to the need to select among competing lexical candidates). He et al. (2021) focused on the latter effect, as the pictures were known to be easy to recognize. They found that the irrelevant speech effects on onset latencies and the first chunk length were obtained only for high name agreement pictures that had low lexical selection demands, which in in turn suggests that the interference can be eliminated when speech production becomes more demanding.

He et al. (2021) established that representational similarity disrupted speech production, but did not distinguish whether the disruption was phonological or semantic in nature. Furthermore, this study does not rule out that the disruption by background speech is due to an *attentional capture* mechanism (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015). This is because the background speech varied in both aspecific context properties (i.e., the presence of pauses in word lists but not in eight-talker babble) and specific linguistic content (i.e., word lists contain linguistic information but the eight-talker babble does not), which may divert attention from the speech production task and cause a drop in speech production performance. Therefore, the present study filled in these gaps by investigating how different types of background speech (word lists versus sentences; the manipulation of the presence/absence of pauses) influenced speech production relative to a quiet control condition.

Specifically, the present study was designed to distinguish between the variants of the interference-by-similarity and attention capture accounts. To distinguish between the semantic and phonological nature of the interference-by-similarity mechanisms, we examined disruption by unintelligible (i.e., Chinese, Experiment 1) and intelligible background speech (i.e., Dutch, Experiment 2) on speech production for native Dutch speakers. The phonological disruption view (Salamé & Baddeley, 1982, 1989) predicts that the presence of background speech, regardless of its intelligibility, should disrupt speech production relative to a quiet condition, predicting the same results across experiments, while the semantic disruption view (Martin et al., 1988) predicts that only intelligible background speech should interfere with speech production, predicting different results across experiments.

To assess the role of attention capture in the irrelevant speech effects on speech production, in both experiments we manipulated the presence/absence of pauses by comparing word lists containing silent pauses (e.g., *渔夫,合唱团,足球,苹果,尺子,鹿* '*fisherman, choir, football, apple, ruler, deer*') with sentences that form continuous speech without pauses (e.g., *鹿和尺子在苹果的左边, 并且足球和合唱团在渔夫的右边*. *'The deer and the ruler are to the left of the apple, and the football and the choir are to the right of the fisherman.'*). Because Dutch word lists and sentences differ on all information levels, it is hard to distinguish specific linguistic information (e.g., phonological/semantic representations) from aspecific context (e.g., the presence/absence of pauses) in intelligible background speech. Hence, the predictions following the attention capture view (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015) only hold for unintelligible background speech: if attention capture is only caused by *aspecific* context variation (i.e., the presence/absence of pauses), Chinese word lists should elicit more interference than Chinese sentences because they contain more pauses. By contrast, if attention capture only results from specific linguistic contents (e.g., semantics or syntax), Chinese word lists should have the same disruptive potency as the Chinese sentence because they are meaningless speech for speakers. The predictions for each account are shown in Table 4.1.

Table 4.1. A summary of predictions in the present study.

Account	Predictions
Interference-by-similarity account (e.g., Jones et al., 1993; Martin et al., 1988; Salamé & Baddeley, 1982, 1989)	
Phonological disruption view (Salamé & Baddeley, 1982, 1989)	Both Chinese speech (in Experiment 1) and Dutch speech (in Experiment 2) should disrupt speech production relative to a quiet condition.
Semantic disruption view (Martin et al., 1988)	Chinese speech (in Experiment 1) should not disrupt speech production relative to a quiet condition, but Dutch speech (in Experiment 2) should.
Attention capture account (e.g., Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015)	
Aspecific attention capture view (Eimer et al., 1996)	Experiment 1: Chinese word lists should be more disruptive than Chinese sentences.
Specific attention capture view (Eimer et al., 1996)	Experiment 1: Chinese word lists should have the same disruptive potency as the sentences.

In both experiments, we employed a continuous speaking-listening paradigm in which participants named four pictures while ignoring irrelevant background speech (also see He et al., 2021). To replicate the finding that difficult lexical selection of speech production reduces the disruption by background speech (He et al., 2021), we also manipulated the name agreement (high, low) of to-be-named pictures and focused on the effect occurring at lexical selection (due to the need to select among competing lexical candidates) in our study. Following earlier work (Alario et al., 2004; Cheng et al., 2011; Vitkovitch & Tyrell, 1995; Shao et al., 2014), we predicted that pictures with low name agreement would be named more slowly than those with high name agreement.

The interaction between the type of background speech and name agreement can provide some insight into whether irrelevant speech effects are attenuated when the required attentional demand of the focal task increases. He et al. (2021) demonstrated that high lexical selection demand (i.e., low name agreement) shielded against distraction from background speech, and attributed this to a top-down *attention engagement* mechanism (also referred to as *task engagement*; see Halin et al., 2014; Marsh et al., 2015). The attention engagement account assumes that when the focal task (e.g., naming pictures) is difficult, the meta-cognitive system triggers a compensatory shift in task-engagement (or concentration) such that a speaker can maintain a desired performance level by reducing the processing of background information (Ball et al., 2018; Eggemeier et al., 1983; Sörqvist & Marsh, 2015). Crucially, the attention engagement mechanism is sensitive to different types of auditory disruption: stimulus-aspecific disruption (e.g., the presence/absence of pauses) remains unaffected by changes in attention engagement in response to task difficulty because it is rooted in the automatic processing of the auditory input that escapes cognitive control (Hughes, 2014), while stimulus-specific distraction (e.g., linguistic richness) is reduced or eliminated by an increase in attention engagement because it requires central attention that taps into cognitive control (Hughes, 2014; Marsh et al., 2018). Thus, we predicted that the interference elicited by unintelligible background speech in Experiment 1 (variation in aspecific context properties) would not be affected by name agreement, while the disruption caused by intelligible background speech in Experiment 2 (variation in both aspecific context and specific linguistic properties) would be reduced for low name agreement pictures compared to high name agreement pictures.

4.2 Experiment 1

Experiment 1 investigated how different types of unintelligible background speech (i.e., Chinese) affected speech production by comparing Chinese word lists, Chinese sentences, and a quiet control condition. From the perspective of a Dutch speaker, the main difference between the Chinese word list and Chinese sentence conditions is the presence or absence of silent pauses. Comparing Chinese speech (word lists and sentences) with the quiet condition distinguishes between the two variants of the interference-by-similarity account. The phonological disruption view (Salamé & Baddeley, 1982, 1989) predicts that Chinese speech should be more disruptive than the quiet condition because it interferes with the phonological codes required for speech production, while the semantic disruption view (Martin et al., 1988) predicts that Chinese speech should not disrupt speech production relative to the quiet condition because it is meaningless to Dutch speakers.

The comparison of the Chinese word list and Chinese sentence conditions distinguishes between the two variants of attention capture (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015). The aspecific attention capture view (Eimer et al., 1996) predicts that the Chinese word lists should cause more interference than Chinese sentences, as they contain pauses that capture attention. By contrast, the specific attention capture view (Eimer et al., 1996) predicts that the Chinese word lists should have the same disruption potential as Chinese sentences, as both are unintelligible to native Dutch speakers.

The difficulty of lexical selection in speech production was also manipulated in the present experiment by varying the name agreement (high, low) of to-be-named pictures. We predicted a name agreement effect where naming performance would be worse for low name agreement pictures than high name agreement pictures. Because here the variation in unintelligible background speech is aspecific and rooted in automatic processing, we also predicted that the irrelevant speech effects would remain unaffected by name agreement.

4.2.1 Methods

Participants

We recruited 50 native speakers of Dutch (45 females, M_{age} = 25 years, range: 20 - 35 years) from the participant pool at the Max Planck Institute for Psycholinguistics. This sample size was selected because power simulations (see https://osf.io/wuafh/) showed that 50 participants and 144 items (i.e., with 80% of the items in the study named successfully) would allow 95% power to measure a plausibly-sized difference (20 ms, SD = 900 ms) between Chinese word list and sentence conditions on the measurement of utterance duration. All participants reported normal or corrected-to-normal vision, no speech or hearing problems, and little Chinese experience. They signed an online informed consent form and received a payment of €6 for their participation. The study was approved by the ethics board of the Faculty of Social Sciences of Radboud University.

Apparatus

The experiment was implemented in FRINEX (FRamework for INteractive EXperiments; for details, see Withers, 2017), a web-based platform developed by the technical group at the Max Planck Institute for Psycholinguistics. Participants used their own laptops with headphones/earphones. We restricted participants to use 14-inch or larger laptops (range: 14-24 inches) with Google Chrome, Firefox, Microsoft Edge, or Brave web browsers. Each participant's speech was recorded by a built-in voice recorder of the web browser. WebMAUS Basic was used for phonetic segmentation and transcription (https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic). Praat (Boersma & Weenink, 2009) was then used to extract the onsets and offsets of all segmented responses.

Materials

Visual stimuli. 240 pictures from He et al., (2021, Experiment 2; pictures selected from the MultiPic database, Duñabeitia et al., 2018; see Appendix A, Table A1) were used in the present study. Of these, 120 were high name agreement pictures, all with a name agreement percentage of 100%, and 120 were low name agreement pictures, with a name agreement percentage between 50% and 87% (M = 72%, SD = 11%). Independent *t*-tests revealed that the two sets of pictures differed significantly in name agreement, but not in any of the following psycholinguistic attributes: visual complexity, word frequency (WF), Age-of-Acquisition (AoA), number of phonemes, number of syllables, word prevalence, phonological neighborhood frequency (PNF), phonological neighborhood size (PNS), orthographic neighborhood frequency (ONF), and orthographic neighborhood size (ONS).

The 120 high name agreement and 120 low name agreement pictures were each divided into three subsets and paired with the two background speech conditions (Chinese word list, Chinese sentence) and a quiet control condition, meaning that each auditory condition was paired with 40 high name agreement and 40 low name agreement pictures. The three sets of pictures were matched on the above-mentioned 10 attributes, as were the high and low name agreement sets of pictures assigned to each auditory condition.

On each trial of the experiment, four pictures, all with high name agreement or all with low name agreement, were presented simultaneously in a 1 × 4 grid (size: 10 cm × 40 cm). The pictures per grid were neither semantically related (i.e. they were from different semantic categories) nor phonologically related (i.e. avoiding the overlap of their first phonemes), as judged by a native speaker of Dutch. There were 20 picture grids for each background speech condition, resulting in 60 grids in total. 24 additional pictures (6 picture grids) were selected as practice stimuli from the same database.

Irrelevant background speech. For the Chinese word list condition, 120 additional Dutch nouns (see Appendix A, Table A2) were selected from the MultiPic database (Duñabeitia et al., 2018) and translated into Chinese nouns by a native Mandarin Chinese speaker. To be paired with the 20 picture grids, these 120 Chinese nouns were divided into 20 word lists of 6 nouns. All 20 lists were matched on the number of phonemes and number of syllables. The number of syllables was also matched between the Chinese nouns and the sets of to-be-named pictures ($t_{(305.91)}$ = -1.58, $p > 0.05$). To avoid phonological overlap between picture naming and the background speech, we designed the word lists so that any six Chinese nouns per list did not share the 1st phoneme with each other, and any five consecutive Chinese nouns per list did not share the 1st phoneme with the to-be-named pictures in the same ordinal position. To create practice stimuli, 12 additional Dutch nouns were also selected from the same database (Duñabeitia et al., 2018) and then translated into Chinese, resulting in two lists. All of the word lists were recorded by a female native Mandarin Chinese speaker in neutral prosody using Audacity software (https://www.audacityteam.org/download/) at a sample rate of 44100 Hz. Each word list was then further processed using Adobe Audition (https://www.adobe.com/products/audition.html) and Praat to delete initial and final silences and compress by up to 0.74%, so that each word list lasted 8 seconds and so there were similar periods of silence (about 700 ms) between consecutive nouns.

For the Chinese sentence condition (see Appendix A, Table A3), the 20 Chinese word lists were transformed into 20 Chinese sentences by adding conjunctions (e.g., 和/并且, “*and*”) and prepositional phrases (e.g., 在左边/在右边; “*to the left/right of*”) to link the nouns. The order of nouns in each sentence was reversed from the corresponding word list, and was designed so that no six Chinese nouns per sentence were phonologically related to each other, and no five consecutive Chinese nouns per sentence were phonologically related to any to-be-named pictures in the same ordinal position. In addition, the two Chinese word lists were also transformed into two Chinese sentences as practice stimuli. The same speaker recorded these sentences in neutral prosody. They were further edited in the same fashion as each Chinese word list (by stretching by up to 9.59%) to last 8 seconds.

Given that this is a web-based experiment, we had to check whether participants were able to hear background speech as well as their concentration level. To this end, we designed attention check trials, in which 19 additional two-syllable Dutch nouns (4 for the practice stage, 15 for the test stage) were selected from the same database (Duñabeitia et al., 2018) to be used as attention check stimuli that needed to be repeated back during the experiment. These nouns were recorded by a native Dutch speaker in neutral prosody. All auditory files were matched on intensity (total RMS (root mean square) = -33.98dB) in Adobe Audition.

Design

The type of background speech (Chinese word list, Chinese sentences, quiet) and the difficulty of lexical selection in speech production (Name agreement: high, low) were treated as within-participant variables; both were randomized within experimental blocks and counterbalanced across participants. Items were repeated three times resulting in three blocks each containing 60 trials with one repetition of each background speech condition and each picture grid. Across blocks, the same set of four pictures was paired with all three background speech conditions, and the pictures were presented in a different arrangement within each repetition. A unique order of stimuli presentation was created for each participant with the Mix program (van Casteren & Davis, 2006), with the constraints that word lists and sentences sharing the same nouns were presented at least every three trials, and attention check trials were presented at least every five trials.

Procedure

Participants were tested on the web[11] and received instructions that they should perform this experiment in a quiet room with the door shut and with potentially distracting electronic equipment turned off. They were asked to image that they were in a laboratory during the experiment, to wear headphones properly, and to set the volume of their laptops to a level that they usually use (e.g., to watch a video) and not change it during the experiment. We asked for permission to record their speech responses and asked them to report their volume values before the test began.

During the experiment, a practice session of ten trials (six test trials and four attention check trials) was followed by the three blocks of experimental trials each containing sixty test trials and five attention check trials. Participants were allowed to take a short break after each block. After completing the main portion of the experiment, participants were asked to type the value of their volume again, which allowed us to check whether they changed the computer volume during the experiment. They also were asked to fill out a questionnaire asking about their Chinese experience (see Appendix A, Table A4), allowing us to check whether they

[11] Here is an example of the experiment for one participant:
https://frinexproduction.mpi.nl/image_naming_noise_cn/?stimulusList=List1

indeed had no Chinese experience. The experiment lasted about 30 minutes.

Practice and experimental trials began with a fixation cross presented for 500 ms, followed by a blank screen for 300 ms. Then, a 1 × 4 grid appeared on the screen in which four pictures were presented simultaneously while a sound file played for up to 8 seconds. Participants named the four pictures one by one from left to right as quickly and accurately as possible while ignoring the background speech. Once finished, they clicked the mouse to end the trial, at which point a blank screen was presented for 1500 ms. An example of a test trial is shown in Figure 4.1. Attention check trials were also included to test the concentration level of participants. The attention test trials shared the same structure as the test trials, but the stimulus screen was blank and an audio file of a single Dutch word was played. In these trials, participants were asked to repeat the Dutch word as quickly and accurately as possible.

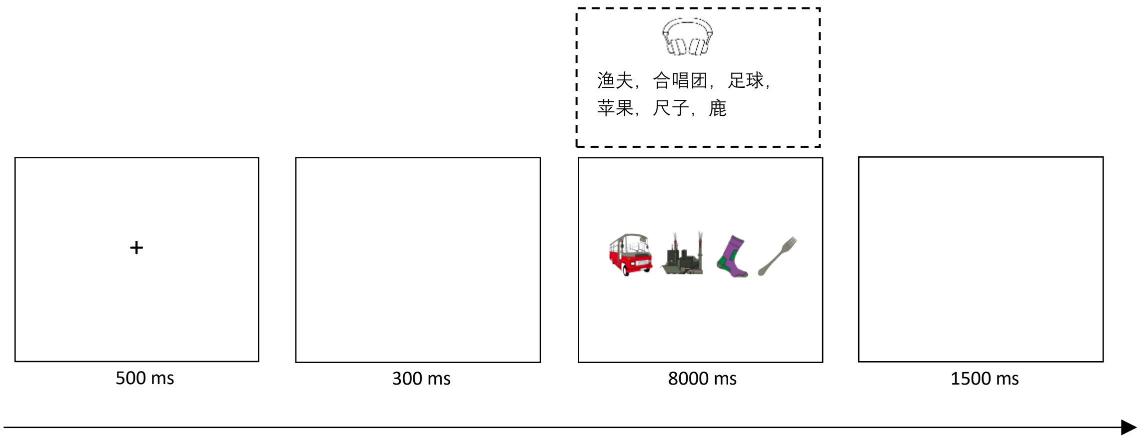

Figure 4.1. An example trial in which participants named pictures with high name agreement while ignoring a Chinese word list (translation: fisherman, choir, football, apple, ruler, deer).

Analysis

Seven dependent variables were coded to index naming performance. Production *accuracy* reflects the proportion of trials where all four pictures were named correctly. Picture names were coded as correct if they matched any of the multiple names given to the picture in the MultiPic database (Duñabeitia et al., 2018); if they were diminutive versions of the multiple names (e.g., *munt* ‘coin’ named as *muntje* ‘little coin’), or if they were judged reasonable by trained research assistants (e.g., *kruk* ‘stool’ named as *stoel* ‘chair’).

For trials where all pictures were named correctly and without hesitations or self-corrections (hereafter, “fully correct trials”), we calculated four main time-based measures. *Onset latency* was defined as the interval from the onset of stimulus presentation to the onset of the utterance, and indexes the beginning stages of speech planning. *Utterance duration* was defined as the interval between the utterance onset of the first picture name and the utterance offset of the fourth picture name, and reflects how long participants took to produce all four picture names. *Total pause time* was defined as the sum of all pauses between object names, and indexes the planning done between producing responses. *Articulation time* was defined as the sum of the articulation durations of all four picture names, and reflects the processing during articulations.

For fully correct trials, we also examined how participants grouped their four responses. Since earlier studies of spontaneous speech coded silent durations longer than 200 ms as silent pauses (e.g., Heldner & Edlund, 2010), we coded responses with 200 ms or less between them as a single response chunk. Two measures were derived: *Total chunk number* refers to how many response chunks participants made on one trial, with a larger number of response chunks meaning more separate planning units for production. *First chunk length* refers to how many names participants produced in their initial response, and provides a measure of how much information participants planned before starting to speak.

To examine the likely magnitude of all effects, Bayesian mixed-effect models (Nicenboim & Vasishth, 2016) were conducted in R version 4.0.3 (R Core Team, 2020) with the package *brms* (version 2.14.4, Bürkner, 2017). Predictors were name agreement (high/low) and the type of background speech (Chinese word list/Chinese sentence/quiet). Name agreement (high/low) was contrast coded with (0.5, -0.5). Two contrasts were made for the type of background speech: the first was coded with (0.25, 0.25, -0.5) to compare the two Chinese speech conditions (word list and sentence) with the quiet condition, and the second was coded with (0.5, -0.5, 0) to compare the Chinese word list and Chinese sentence conditions. The random effect structure for the models included random intercepts for participants and items, and random slopes for name agreement and the type of background speech by participants and items. Separate models were fitted for each dependent measure. All models had four chains and each chain had 24000 iterations depending on model convergence (listed in model output tables). We used a warm-up (or burn-in) period of 2000 iterations in each chain, which means we removed the data based on the first 2000 iterations in order to correct the initial sampling bias.

All models used weak, widely spread priors that would be consistent with a range of null to moderate effects. The model of accuracy used family *bernoulli* combined with a *logit* link, with a student-*t* prior with 1 degree of freedom and a scale parameter of 2.5. The models of log-transformed onset latency, log-transformed utterance duration, and log-transformed articulation time used a weak normal prior with an SD of 0.2, and the model of log-transformed total pause time used a weak normal prior with an SD of 1. These models were performed using the family *gaussian* combined with *identity* link. Total chunk number and first chunk length had weak normal priors centered at zero with an SD of 1, and used family *possion* combined with the *log* link. All models were run until R hat value for each estimated parameter was 1.00, indicating full convergence.

For these models, the size of reported betas reflects estimated effect sizes, with larger absolute values of betas reflecting larger effects. We reported the parameters for which 95% Credible Intervals (hereafter, Cr.I) do not contain zero, which is analogous to the frequentist null hypothesis significance test: the parameter has a non-zero effect with high certainty. We also reported any parameters for which the point estimate for the beta is about twice the size of its error, as this suggests that the estimated effect is large compared to the uncertainty around it. We also reported the posterior probability of all weak effects, indicating the proportion of samples with a value equal to or above the beta estimate.

4.2.2 Results

Six participants were removed from further analyses: three did not run the experiments successfully due to a bad internet connection, two gave no responses on attention check trials, and one had too much Chinese experience as indicated by their responses on the Chinese experience questionnaire. The data from the remaining 44 participants was checked for errors, removing from analysis any trials with implausible names (e.g., *koekje* ‘cookie’ named as *virus*), hesitations (e.g., *komkommer* ‘cucumber’ named as *kom...komkommer*), self-corrections (e.g., *komkommer* ‘cucumber’ misnamed as *courgette...komkommer* ‘courgette...cucumber’), and any trials where objects were omitted or named in the wrong order. The exclusion of these inaccurate trials resulted in a loss of 13.7% of the data (range by participants: 1.1% - 30% of removed trials). Then, any onset latencies below 200 ms were removed from this analysis, resulting in a loss of 0.47% of the data. Any total pause times below 20 ms were also removed from this analysis, resulting in a loss of 12.98% of the data. Finally, any data points more than 2.5 standard deviations below or above the mean values were removed for each time measure

(1.87% for log-transformed onset latency, 0.86% for log-transformed utterance duration, 0.97% for log-transformed total pause time, and 1.33% for log-transformed articulation time). Descriptive statistics of all dependent variables are shown in Table 4.2.

Table 4.2. Means and standard deviations of the dependent variables by name agreement and the type of background speech in Experiment 1.

	High NA			Low NA		
	Chinese Word List	Chinese Sentence	Quiet	Chinese Word List	Chinese Sentence	Quiet
Accuracy	91%	91%	92%	82%	82%	81%
Onset latency (ms)	1246 (462)	1279 (522)	1198 (408)	1434 (579)	1413 (539)	1345 (486)
Utterance duration (ms)	2868 (790)	2868 (771)	2791 (765)	3475 (1062)	3482 (1025)	3392 (970)
Total pause time (ms)	685 (621)	662 (590)	645 (582)	1078 (860)	1043 (790)	1040 (805)
Articulation time (ms)	2309 (431)	2332 (429)	2246 (392)	2518 (498)	2536 (522)	2450 (476)
Total chunk number	1.9 (1.0)	1.9 (1.0)	1.9 (1.0)	2.3 (1.1)	2.4 (1.1)	2.4 (1.1)
First chunk length	2.7 (1.3)	2.7 (1.3)	2.8 (1.3)	2.3 (1.3)	2.2 (1.2)	2.2 (1.2)

Note. Standard deviations are given in parentheses. All time and chunking measures reflect fully correct trials only.

Attention Checks. The mean accuracy for attention check responses was 97% (range by participants: 73% - 100%), showing that participants' attention levels were good and that they indeed heard the background speech.

Accuracy. Participants produced sensible responses on 86% of the naming trials. As shown in Table 4.3, a Bayesian mixed-effect model showed that accuracy was considerably lower for low name agreement pictures than high name agreement pictures (β= 0.099, SE = 0.025, 95% Cr.I = [0.051, 0.147]), but it was not influenced by the type of background speech. Name agreement and the type of background speech did not interact.

Onset latency. As shown in Table 4.3 and the left panel of Figure 4.2, a Bayesian mixed-effect model showed that log-transformed onset latency was affected by name agreement: it took participants reliably longer to plan names for low name agreement pictures than high name agreement pictures (β= -0.122, SE = 0.014, 95% Cr.I = [-0.149, -0.095]). There was moderate evidence for the first contrast (Chinese vs. Quiet) of background speech, showing that the log-transformed onset latencies in the two Chinese speech conditions (word list and sentence) were slower than in the quiet condition (β= 0.064, SE = 0.038, 95% Cr.I = [-0.011, 0.138]). Note that while the 95 % Cr.I contains zero, the point estimate is high relative to the error around it, and 96% of the posterior distribution around the estimated effect is above zero. Name agreement and the type of background speech did not interact.

Utterance duration. As shown in Table 4.3 and the right panel of Figure 4.2, a Bayesian mixed-effect model showed that the log-transformed utterance duration was significantly longer for low name agreement pictures than high name agreement pictures (β= -0.191, SE = 0.02, 95% Cr.I = [-0.231, -0.151]), but it was not influenced by the type of background speech. Again, name agreement and the type of background speech did not interact.

Total pause time. As shown in Table 4.3 and the left panel of Figure 4.2, a Bayesian mixed-effect model showed that the results for this measurement patterned in the same way as the log-transformed utterance duration. The log-transformed total pause time was considerably longer for low name agreement pictures than high name agreement pictures (β= -0.574, SE = 0.058, 95% Cr.I = [-0.687, -0.460]), but it did not vary with the type of background speech. Name agreement and the type of background speech did not interact.

Articulation time. As shown in Table 4.3 and the right panel of Figure 4.2, a Bayesian mixed-effect model showed that log-transformed articulation time was influenced by both name agreement and the type of background speech: It was significantly longer for low name agreement pictures than high name agreement pictures (β= -0.085, SE = 0.02, 95% Cr.I = [-0.125, -0.046]), and it was reliably longer in the two Chinese speech conditions (word list and sentence) than in the quiet condition (β= 0.038, SE = 0.014, 95% Cr.I = [0.01, 0.066]). Again, name agreement did not interact with the type of background speech.

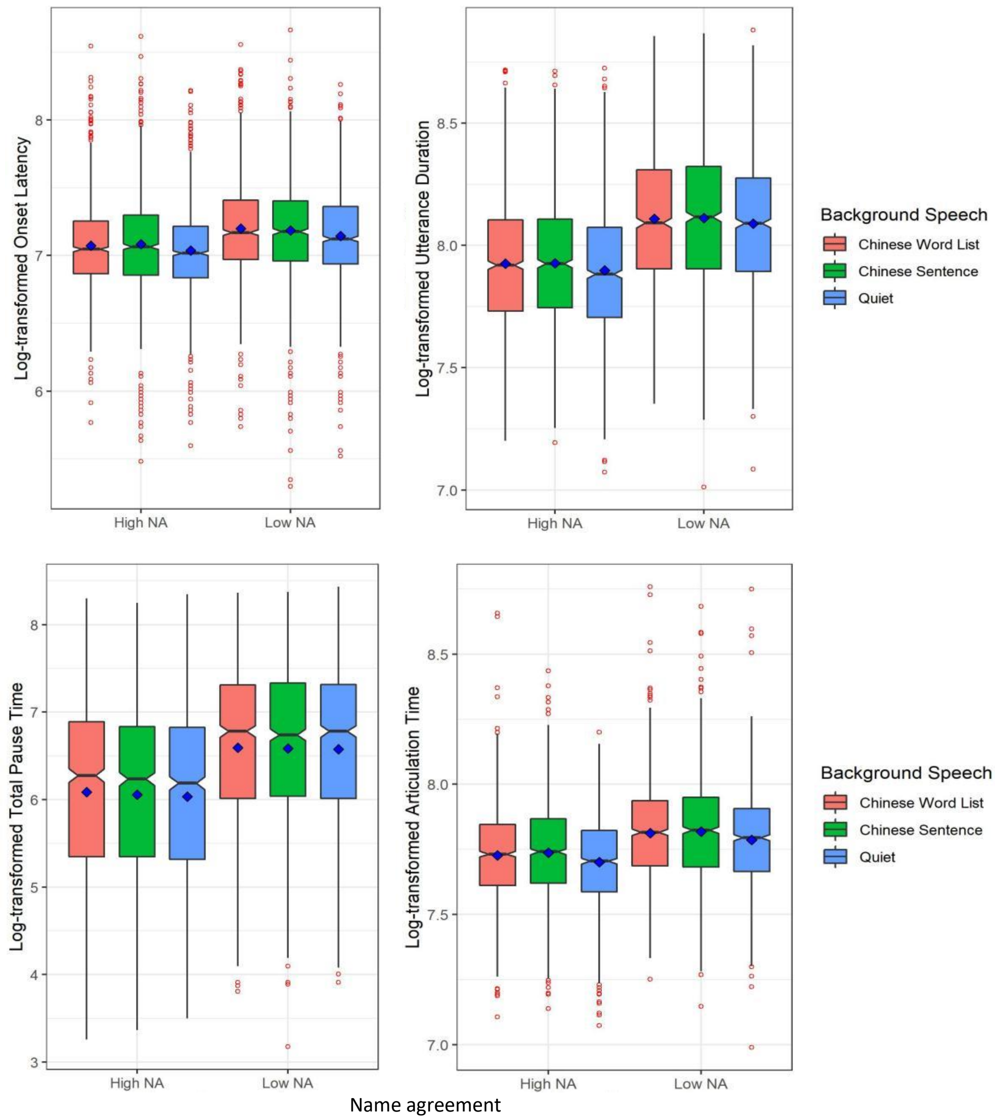

Figure 4.2. Log-transformed Onset latency (top left), log-transformed utterance duration (top right), log-transformed total pause time (bottom left), and log-transformed articulation time (bottom right) split by name agreement (NA: high, low) and the type of background speech (Chinese word list, Chinese sentence, Quiet) in Experiment 1. Blue squares represent condition means and red points reflect outliers.

Total chunk number. As shown in Table 4.3 and the left panel of Figure 4.3, a Bayesian mixed-effect model showed that participants grouped their responses in more chunks for low name agreement pictures than high name agreement pictures (β= -0.252, SE = 0.025, 95% Cr.I = [-

0.301, -0.203]). Total chunk number was not impacted by the type of background speech, with no interaction between name agreement and the type of background speech.

First chunk length. As shown in Table 4.3 and the right panel of Figure 4.3, a Bayesian mixed-effect model showed that participants planned fewer names in their first response chunk for low name agreement pictures than high name agreement pictures (β= 0.218, SE = 0.025, 95% Cr.I = [0.168, 0.268]), but first chunk length was not affected by the type of background speech and there was no interaction between name agreement and the type of background speech.

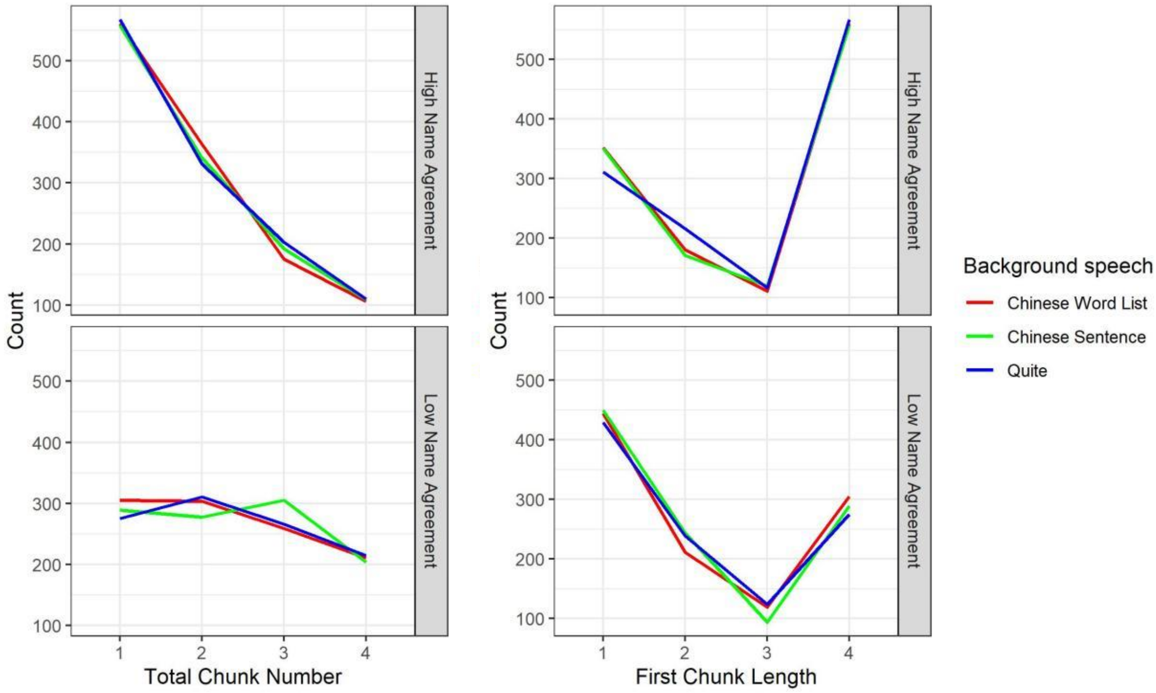

Figure 4.3. Total chunk number (left) and first chunk length (right) split by name agreement (NA: high, low) and the type of background speech (Chinese word list, Chinese sentence, Quiet) in Experiment 1.

Table 4.3. Results of Bayesian mixed-effect models for all dependent variables in Experiment 1.

		Estimate	Est.error	95% Cr. I lower	95% Cr. I upper	Effective samples
Accuracy						
Population-level effects	Intercept	0.863	0.017	0.83	0.895	32170
	Name Agreement	**0.099**	**0.025**	**0.051**	**0.147**	**59697**
	Speech vs. Quiet	0	0.014	-0.028	0.029	107958
	Word List vs. Sentence	0.003	0.011	-0.019	0.025	131954
	NA × (S vs. Q)	-0.02	0.028	-0.076	0.036	107878
	NA × (WL vs. S)	0.001	0.022	-0.042	0.045	134552
Group-level effects	Participants					
	sd(Intercept)	0.075	0.009	0.06	0.095	27257
	sd(NA)	0.043	0.01	0.024	0.064	54647
	sd(Svs.Q)	0.016	0.012	0.001	0.043	48050
	sd(WLvs.S)	0.012	0.009	0.001	0.033	56746
	sd(NA×(Svs.Q))	0.021	0.016	0.001	0.061	69866
	sd(NA×(WLvs.S))	0.023	0.017	0.001	0.065	55462
	Items					
	sd(Intercept)	0.058	0.02	0.016	0.092	6156
	sd(NA)	0.117	0.04	0.033	0.184	6086
	sd(Svs.Q)	0.05	0.018	0.011	0.085	20580
	sd(WLvs.S)	0.03	0.018	0.002	0.066	16829
	sd(NA×(Svs.Q))	0.099	0.037	0.023	0.17	22166
	sd(NA×(WLvs.S))	0.06	0.036	0.003	0.133	17133
Log-transformed onset latency						
Population-level effects	Intercept	7.133	0.028	7.078	7.188	5293
	Name Agreement	**-0.122**	**0.014**	**-0.149**	**-0.095**	**48510**
	Speech vs. Quiet	*0.064*	*0.038*	*-0.011*	*0.138*	*49911*
	Word List vs. Sentence	-0.002	0.037	-0.074	0.071	47960

	NA × (S vs. Q)	-0.006	0.07	-0.144	0.132	50854
	NA × (WL vs. S)	-0.014	0.069	-0.15	0.122	56068
Group-level effects	Participants					
	sd(Intercept)	0.177	0.02	0.143	0.223	10270
	sd(NA)	0.029	0.011	0.005	0.051	18616
	sd(Svs.Q)	0.077	0.015	0.049	0.109	31488
	sd(WLvs.S)	0.05	0.013	0.024	0.077	24869
	sd(NA×(Svs.Q))	0.035	0.025	0.001	0.091	27704
	sd(NA×(WLvs.S))	0.048	0.027	0.003	0.105	21254
	Items					
	sd(Intercept)	0.029	0.012	0.004	0.049	2331
	sd(NA)	0.058	0.024	0.008	0.098	2319
	sd(Svs.Q)	0.173	0.095	0.008	0.311	1284
	sd(WLvs.S)	0.177	0.1	0.006	0.316	1181
	sd(NA×(Svs.Q))	0.345	0.189	0.016	0.622	1222
	sd(NA×(WLvs.S))	0.325	0.202	0.011	0.626	1228
Log-transformed utterance duration						
Population-level effects	Intercept	8.021	0.023	7.974	8.066	6414
	Name Agreement	**-0.191**	**0.02**	**-0.231**	**-0.151**	**39748**
	Speech vs. Quiet	0.029	0.026	-0.022	0.08	54056
	Word List vs. Sentence	-0.003	0.022	-0.046	0.04	51599
	NA × (S vs. Q)	0.018	0.05	-0.081	0.117	56494
	NA × (WL vs. S)	0.005	0.044	-0.081	0.091	49868
Group-level effects	Participants					
	sd(Intercept)	0.142	0.016	0.115	0.178	12242
	sd(NA)	0.064	0.009	0.047	0.084	35908
	sd(Svs.Q)	0.014	0.01	0.001	0.036	35029
	sd(WLvs.S)	0.01	0.007	0	0.026	45776
	sd(NA×(Svs.Q))	0.019	0.014	0.001	0.054	49185

	sd(NA×(WLvs.S))	0.04	0.02	0.004	0.081	31111
	Items					
	sd(Intercept)	0.04	0.023	0.002	0.074	1565
	sd(NA)	0.081	0.045	0.004	0.148	1643
	sd(Svs.Q)	0.125	0.055	0.015	0.21	3193
	sd(WLvs.S)	0.111	0.036	0.037	0.173	5059
	sd(NA×(Svs.Q))	0.251	0.109	0.032	0.422	3182
	sd(NA×(WLvs.S))	0.222	0.073	0.072	0.346	4698
Log-transformed total pause time						
Population-level effects	Intercept	6.274	0.081	6.115	6.432	7041
	Name Agreement	**-0.574**	**0.058**	**-0.687**	**-0.46**	**43884**
	Speech vs. Quiet	0.009	0.07	-0.127	0.147	67063
	Word List vs. Sentence	0.017	0.064	-0.108	0.143	58586
	NA × (S vs. Q)	0.039	0.134	-0.224	0.304	69382
	NA × (WL vs. S)	0.033	0.126	-0.216	0.283	62853
Group-level effects	Participants					
	sd(Intercept)	0.508	0.058	0.41	0.635	13162
	sd(NA)	0.177	0.033	0.116	0.247	43499
	sd(Svs.Q)	0.122	0.052	0.017	0.222	26954
	sd(WLvs.S)	0.067	0.04	0.004	0.152	31799
	sd(NA×(Svs.Q))	0.078	0.06	0.003	0.223	53517
	sd(NA×(WLvs.S))	0.126	0.08	0.006	0.298	32126
	Items					
	sd(Intercept)	0.107	0.063	0.004	0.204	2282
	sd(NA)	0.222	0.124	0.01	0.409	2251
	sd(Svs.Q)	0.293	0.14	0.023	0.518	3763
	sd(WLvs.S)	0.292	0.102	0.078	0.469	6780
	sd(NA×(Svs.Q))	0.59	0.279	0.049	1.038	3738
	sd(NA×(WLvs.S))	0.579	0.205	0.151	0.935	6811

Log-transformed articulation time						
	Intercept	7.768	0.019	7.731	7.805	5872
	Name Agreement	**-0.085**	**0.02**	**-0.125**	**-0.046**	**46351**
Population-level effects	Speech vs. Quiet	**0.038**	**0.014**	**0.01**	**0.066**	**61569**
	Word List vs. Sentence	-0.007	0.012	-0.031	0.017	64224
	NA × (S vs. Q)	0.007	0.027	-0.046	0.06	66049
	NA × (WL vs. S)	-0.003	0.024	-0.05	0.044	62948
	Participants					
	sd(Intercept)	0.108	0.013	0.087	0.136	11302
	sd(NA)	0.053	0.007	0.041	0.069	28988
	sd(Svs.Q)	0.029	0.008	0.011	0.045	20619
	sd(WLvs.S)	0.008	0.005	0	0.02	35991
	sd(NA×(Svs.Q))	0.014	0.011	0.001	0.039	41441
Group-level effects	sd(NA×(WLvs.S))	0.021	0.014	0.001	0.051	21175
	Items					
	sd(Intercept)	0.042	0.026	0.001	0.078	1378
	sd(NA)	0.083	0.051	0.003	0.157	1380
	sd(Svs.Q)	0.06	0.036	0.002	0.113	1763
	sd(WLvs.S)	0.055	0.029	0.003	0.098	1923
	sd(NA×(Svs.Q))	0.121	0.071	0.005	0.225	1729
	sd(NA×(WLvs.S))	0.106	0.059	0.005	0.195	1932
Total chunk number						
	Intercept	0.715	0.041	0.635	0.795	9365
	Name Agreement	**-0.252**	**0.025**	**-0.301**	**-0.203**	**52559**
Population-level effects	Speech vs. Quiet	-0.016	0.035	-0.085	0.053	74601
	Word List vs. Sentence	-0.017	0.029	-0.074	0.040	79456
	NA × (S vs. Q)	0.014	0.070	-0.123	0.152	77761
	NA × (WL vs. S)	0.009	0.058	-0.105	0.123	78972

	Participants					
	sd(Intercept)	0.256	0.030	0.206	0.321	15391
	sd(NA)	0.062	0.021	0.020	0.104	46312
	sd(Svs.Q)	0.023	0.018	0.001	0.067	62627
	sd(WLvs.S)	0.020	0.016	0.001	0.058	63929
	sd(NA×(Svs.Q))	0.049	0.037	0.002	0.139	64075
Group-level effects	sd(NA×(WLvs.S))	0.043	0.033	0.002	0.122	61696
	Items					
	sd(Intercept)	0.035	0.020	0.002	0.073	8804
	sd(NA)	0.070	0.040	0.004	0.146	7966
	sd(Svs.Q)	0.124	0.058	0.012	0.229	9285
	sd(WLvs.S)	0.102	0.043	0.014	0.183	13656
	sd(NA×(Svs.Q))	0.246	0.116	0.020	0.458	9163
	sd(NA×(WLvs.S))	0.202	0.087	0.025	0.365	13743
First chunk length						
	Intercept	0.863	0.042	0.781	0.946	11967
	Name Agreement	**0.218**	**0.025**	**0.168**	**0.268**	**96798**
Population-level effects	Speech vs. Quiet	-0.012	0.034	-0.077	0.055	95932
	Word List vs. Sentence	0.013	0.030	-0.046	0.072	92168
	NA × (S vs. Q)	-0.030	0.067	-0.162	0.101	95948
	NA × (WL vs. S)	-0.027	0.060	-0.145	0.091	95897
	Participants					
	sd(Intercept)	0.262	0.031	0.210	0.330	19220
	sd(NA)	0.022	0.016	0.001	0.061	50297
Group-level effects	sd(Svs.Q)	0.025	0.019	0.001	0.069	64357
	sd(WLvs.S)	0.023	0.018	0.001	0.065	61516
	sd(NA×(Svs.Q))	0.047	0.036	0.002	0.135	64675
	sd(NA×(WLvs.S))	0.043	0.033	0.002	0.122	63963
	Items					

	sd(Intercept)	0.047	0.025	0.003	0.090	5967
	sd(NA)	0.094	0.050	0.005	0.179	5836
	sd(Svs.Q)	0.124	0.053	0.015	0.221	11407
	sd(WLvs.S)	0.116	0.042	0.028	0.195	19228
	sd(NA×(Svs.Q))	0.249	0.106	0.031	0.442	13355
	sd(NA×(WLvs.S))	0.230	0.085	0.051	0.389	18080

Note. Models for all dependent variables were run for 24000 iterations. Bolded values indicate effects where the 95% Cr.I does not contain zero. NA refers to name agreement, WL refers to word list, S refers to sentence, Q refers to quiet.

4.2.3 Discussion

This experiment was designed to explore how different types of unintelligible background speech influenced spoken language production, and to show whether this influence was modulated by the difficulty of lexical selection in speech production. There were four main findings. First, we found that name agreement affected all dependent measures, showing that speakers were less accurate, took longer to plan names, and produced fewer names at a time for pictures with low name agreement than high name agreement. This finding is consistent with earlier work using single- and multiple-picture naming paradigms (e.g., Alario et al., 2004; He et al., 2021; Shao et al., 2014).

Second, we found that the presence of Chinese background speech (word lists and sentences) increased articulation time significantly, but only had a weak impact on speech onset latencies, relative to a quiet condition. This finding suggests that unintelligible speech slows down the articulation and also initial planning before speaking, but does not affect the processing done between consecutive responses and the choice of planning units. The disruptive effects thus are consistent with the *phonological disruption view* (Salamé & Baddeley, 1982, 1989) that any speech, regardless of its intelligibility, should disrupt the focal task.

The third main finding was that there was no difference between the Chinese word list and Chinese sentence conditions on any dependent measures, which suggests that the aspecific context variation of unintelligible background speech (i.e., the presence / absence of pauses) does not elicit interference on speech production. This in turn implies that attention capture by background speech may be specific in nature.

The fourth main finding is that name agreement did not modulate the processing of unintelligible background speech on any measures, which is consistent with our prediction that auditory disruption by stimulus-aspecific variation is not affected by changes in attention engagement because it is automatic (Hughes, 2014; Marsh et al., 2018). This implies that speakers can block off the unintelligible background speech easily, which leaves enough attentional resources for them to perform lexical selection. This finding suggests that the Chinese background speech did not specifically affect lexical selection, which is unsurprising because we assume that the participants could not access the meanings of the Chinese words. We discuss the implications of this further in the General Discussion.

It is unclear whether the auditory distraction caused by unintelligible background speech generalizes to situations with intelligible background speech because direct evidence regarding how intelligible background speech affects speech production is rare (for an exception, see He et al., 2021). Understanding this is important from an applied perspective as it is of great practical interest to know whether and how speakers shield against the detrimental effects of auditory distraction in real-world settings, which often contain intelligible background speech (e.g., in a restaurant or on a train). Thus, we extended our investigation of the irrelevant speech effects to an intelligible-background-speech context by replacing Chinese speech with Dutch speech in Experiment 2.

4.3 Experiment 2

Experiment 2 examined whether the irrelevant speech effects observed in Experiment 1 would be replicated, or altered, when the background speech was intelligible. To this end, we replaced the Chinese speech with Dutch speech (word lists and sentences) and kept an identical quiet control condition. Given that the two Dutch speech conditions (word lists and sentences) contain both phonological and semantic information, the comparison of the Dutch speech with the quiet condition cannot distinguish the phonological- or semantic-nature of the interference-by-similarity account (Martin et al., 1988; Salamé & Baddeley, 1982, 1989). However, the contrast of the Dutch word list with Dutch sentence may provide some insights into the aspecific or specific nature of attention capture (Buchner et al., 2004; Cowan, 1995; Eimer et al., 1996; Elliott & Briganti, 2012; Röer et al., 2013, 2015). If Dutch word lists are more disruptive than Dutch sentences, the disruptive effect could be attributable to the aspecific context variation (i.e., the presence / absence of pauses) or to the possibility that the words in lists may capture attention more effectively when they are preceded by pauses (i.e., the disruption by specific linguistic contents is a byproduct of aspecific attention capture). By contrast, if Dutch sentences are more disruptive than Dutch word lists, then specific attention capture better accounts for the data because the sentences contain richer linguistic content (e.g., semantic / syntactic factors) that may capture more attention.

To assess whether the disruption by intelligible background speech was modulated by the difficulty of speech production, we again manipulated the name agreement (high versus low) of to-be-named pictures. Following the claim that the stimulus-specific auditory distraction should be reduced or eliminated by an increase in attention engagement because it requires central attention and cognitive control (Hughes, 2014; Marsh et al., 2018), we predicted an interaction between name agreement and the type of background speech, such that planning low name agreement pictures would reduce the processing of background speech due to top-down attention engagement (Halin et al., 2014; Marsh et al., 2015).

4.3.1 Method

Participants

We recruited 47 native Dutch speakers (33 females, M_{age} = 26 years, range: 18 - 39 years) from the participant pool at the Max Planck Institute for Psycholinguistics. This sample size was selected because power simulations (see https://osf.io/wuafh/ for scripts) showed that 46 participants and 144 items (an 80% accuracy rate) would allow 96% power to measure a plausibly-sized interaction between the type of background speech and name agreement on the measurement of utterance duration. The interaction effect size used in the simulation was an irrelevant speech effect (i.e., Dutch word lists > Dutch sentence) of 20 ms or smaller (SD = 900 ms) for low name agreement pictures, but 60 ms or larger (SD = 900 ms) for high name agreement pictures. All participants reported normal or corrected-to-normal vision and no speech or hearing problems. They signed an online informed consent form and received a payment of €6 for their participation. The study was approved by the ethics board of the Faculty of Social Sciences of Radboud University.

Apparatus

The same apparatus was used as in Experiment 1.

Materials

Visual stimuli. The same 240 pictures (i.e., 60 picture grids) from Experiment 1 were used as test stimuli. Trials were set up as in Experiment 1, with four pictures in a 1×4 grid (10 cm×40 cm) that were neither semantically nor phonologically related to each other. The same 24 pictures from Experiment 1 were also used as practice stimuli.

Irrelevant background speech. For the Dutch word lists (see Appendix C, Table C1), the 120 nouns from Experiment 1 were used in Dutch, and matched with picture names on word frequency, number of syllables, number of phonemes, age-of-acquisition, and word prevalence. To pair with the set of 20 picture grids, these 120 Dutch nouns were divided into 20 word lists of 6 nouns, each list matched on word frequency and number of syllables. To equate the amount of semantic and phonological overlap across trials between speech planning and auditory background speech, we made sure that the six Dutch nouns per word list were neither semantically nor phonologically related to each other, and any three consecutive Dutch nouns per word list were neither semantically nor phonologically related to the to-be-named pictures in the same ordinal position. In addition, 12 Dutch versions of nouns from Experiment 1 were used as practice stimuli, resulting in two Dutch word lists. All of the Dutch word lists were recorded by a female native Dutch speaker[12] in neutral prosody and further edited in the same fashion as the Chinese word lists to last 8 seconds each with similar silent periods (about 700 ms) between consecutive nouns, by stretching by up to 9.38%.

For the Dutch sentence condition (see Appendix C, Table C2), the 20 Dutch word lists were transformed into 20 Dutch sentences in the same manner as Experiment 1 by combining them with conjunctions (e.g., *en* 'and') and prepositional phrases (e.g., *bevinden zich links/rechts van* 'are to the left/right of'). The order of nouns in each sentence was again reversed from that in the corresponding word list. In addition, the two Dutch word lists were also transformed into two Dutch sentences as practice stimuli. The same female native Dutch speaker recorded these sentences in neutral prosody which were further edited to last 8 seconds each, by stretching by up to 14.29%. The same 19 attention catch trials (15 as test stimuli, 4 as practice stimuli) from Experiment 1 were also included. All auditory files were matched on intensity (total RMS = -33.98dB) in Adobe Audition.

[12] This was a different speaker from the one who recorded Dutch words for attention check trials.

Design

The design used in Experiment 2 was identical to the design used in Experiment 1, except that Chinese speech conditions (Chinese word list, Chinese sentence) were replaced with Dutch speech conditions (Dutch word list, Dutch sentence).

Procedure

The procedure was identical to Experiment 1 except that participants did not fill out the questionnaire of Chinese experience[13].

Analysis

The analysis was the same as Experiment 1.

4.3.2 Results

Six participants were removed from further analyses: one's speech responses were not recorded; three had no responses for attention check trials; one had also participated in Experiment 1, and one's speech responses contained too much noise to annotate. The data from the remaining 41 participants was checked for errors, removing from analysis any trials with implausible names, hesitations, self-corrections, and any trials where pictures were omitted or named in the wrong order. The exclusion of these inaccurate trials resulted in a loss of 12.7% of data (range by participants: 2.8% - 42% of removed trials). Then, any data points below 200 ms were removed for onset latency, resulting in a loss of 0.02% of the data. Any data time points below 20 ms were also removed for total pause time, resulting in a loss of 12.17% of the data. Finally, any data points more than 2.5 standard deviations below or above the mean values were removed for the time measures (1.61% for log-transformed onset latency, 0.85% for log-transformed utterance duration, 1.01% for log-transformed total pause time, and 1.18% for log-transformed articulation time). Descriptive statistics of all dependent variables are shown in Table 4.4.

[13] Here is an example of Experiment 2 for one participant:
https://frinexproduction.mpi.nl/image_naming_noise_nl/?stimulusList=List1

Table 4.4. Means and standard deviations of the dependent variables by name agreement and the type of background speech in Experiment 2.

	High NA			Low NA		
	Dutch Word List	Dutch Sentence	Quiet	Dutch Word List	Dutch Sentence	Quiet
Accuracy	92%	92%	93%	82%	82%	84%
Onset latency (ms)	1304 (496)	1300 (493)	1195 (362)	1451 (568)	1486 (611)	1392 (492)
Utterance duration (ms)	2864 (859)	2871 (872)	2690 (776)	3481 (1028)	3463 (1078)	3474 (1087)
Total pause time (ms)	771 (759)	726 (745)	632 (636)	1090 (877)	1072 (903)	1160 (909)
Articulation time (ms)	2260 (393)	2274 (415)	2172 (387)	2484 (467)	2482 (482)	2392 (458)
Total chunk number	1.9 (1.0)	1.9 (1.0)	1.9 (1.0)	2.4 (1.0)	2.4 (1.1)	2.5 (1.1)
First chunk length	2.7 (1.3)	2.8 (1.3)	2.8 (1.3)	2.2 (1.2)	2.3 (1.2)	2.2 (1.2)

Note. Standard deviations are given in parentheses. All time and chunking measures reflect fully correct trials only.

Attention Check. The mean accuracy for attention check responses was 98% (range by participants: 73% - 100%), showing that participants indeed heard the background speech during the experiment.

Accuracy. Participants produced the intended responses on 87% of the naming trials. As shown in Table 4.5, a Bayesian mixed-effect model showed that accuracy was significantly lower for low name agreement pictures than high name agreement pictures (β= 1.061, SE = 0.223, 95% Cr.I = [0.63, 1.506]), but it was not affected by the type of background speech. Name agreement and the type of background speech did not interact.

Onset latency. As shown in Table 4.5 and the left panel of Figure 4.4, a Bayesian mixed-effect model showed that log-transformed onset latency was reliably longer when planning names for low name agreement pictures than high name agreement pictures (β= -0.128, SE = 0.014, 95% Cr.I = [-0.155, -0.1]). There was moderate evidence for the first contrast of background speech (Dutch speech vs. Quiet), such that the log-transformed onset latencies in the two Dutch speech conditions (word list and sentence) were slower than in the quiet condition (β= 0.076, SE = 0.04, 95% Cr.I = [-0.003, 0.155]). Note that while the 95 % Cr.I contains zero, the point estimate is high relative to the error around it, and 93% of the posterior distribution around the estimated effect is above zero. Again, name agreement did not interact with the type of background speech.

Utterance duration. As shown in Table 4.5 and the right panel of Figure 4.4, a Bayesian mixed-effect model showed that the log-transformed utterance duration was significantly longer for low name agreement pictures than high name agreement pictures (β= -0.215, SE = 0.022, 95% Cr.I = [-0.257, -0.172]). There was moderate evidence for the first contrast of background speech (Dutch speech vs. Quiet), such that the log-transformed utterance durations in the two Dutch speech conditions (word list and sentence) were slower than in the quiet condition (β= 0.05, SE = 0.031, 95% Cr.I = [-0.012, 0.111]). Note that while the 95 % Cr.I contains zero, the point estimate is high relative to the error around it, and 93% of the posterior distribution around the estimated effect is above zero. Again, name agreement did not interact with the type of background speech.

Total pause time. As shown in Table 4.5 and the left panel of Figure 4.4, a Bayesian mixed-effect model showed that log-transformed total pause time was longer for low name agreement pictures than high name agreement pictures (β= -0.599, SE = 0.072, 95% Cr.I = [-0.741, -0.458]), but it did not vary with the type of background speech. However, there was moderate evidence for the interaction of name agreement and the first contrast (Dutch speech vs. Quiet) of background speech (β= 0.28, SE = 0.173, 95% Cr.I = [-0.06, 0.621]). Note that while the 95 % Cr.I contains zero, the point estimate is high relative to the error around it, and 93% of the posterior distribution around the estimated effect is above zero. This demonstrates that the log-transformed total pause time in the Dutch speech condition was longer than that in the quiet condition for high name agreement pictures (β= 0.394, SE = 0.171, 95% Cr.I = [0.058, 0.727]), but not for low name agreement pictures.

Articulation time. As shown in Table 4.5 and the right panel of Figure 4.4, a Bayesian mixed-effect model showed that the log-transformed articulation time was affected by both name agreement and the type of background speech: It took significantly longer to articulate names of low name agreement than high name agreement pictures (β= -0.093, SE = 0.02, 95% Cr.I = [-0.133, -0.054]), and it was reliably longer in the two Dutch speech conditions (word list and sentence) than in the quiet condition (β= 0.054, SE = 0.016, 95% Cr.I = [0.023, 0.085]). There was no interaction between name agreement and the type of background speech.

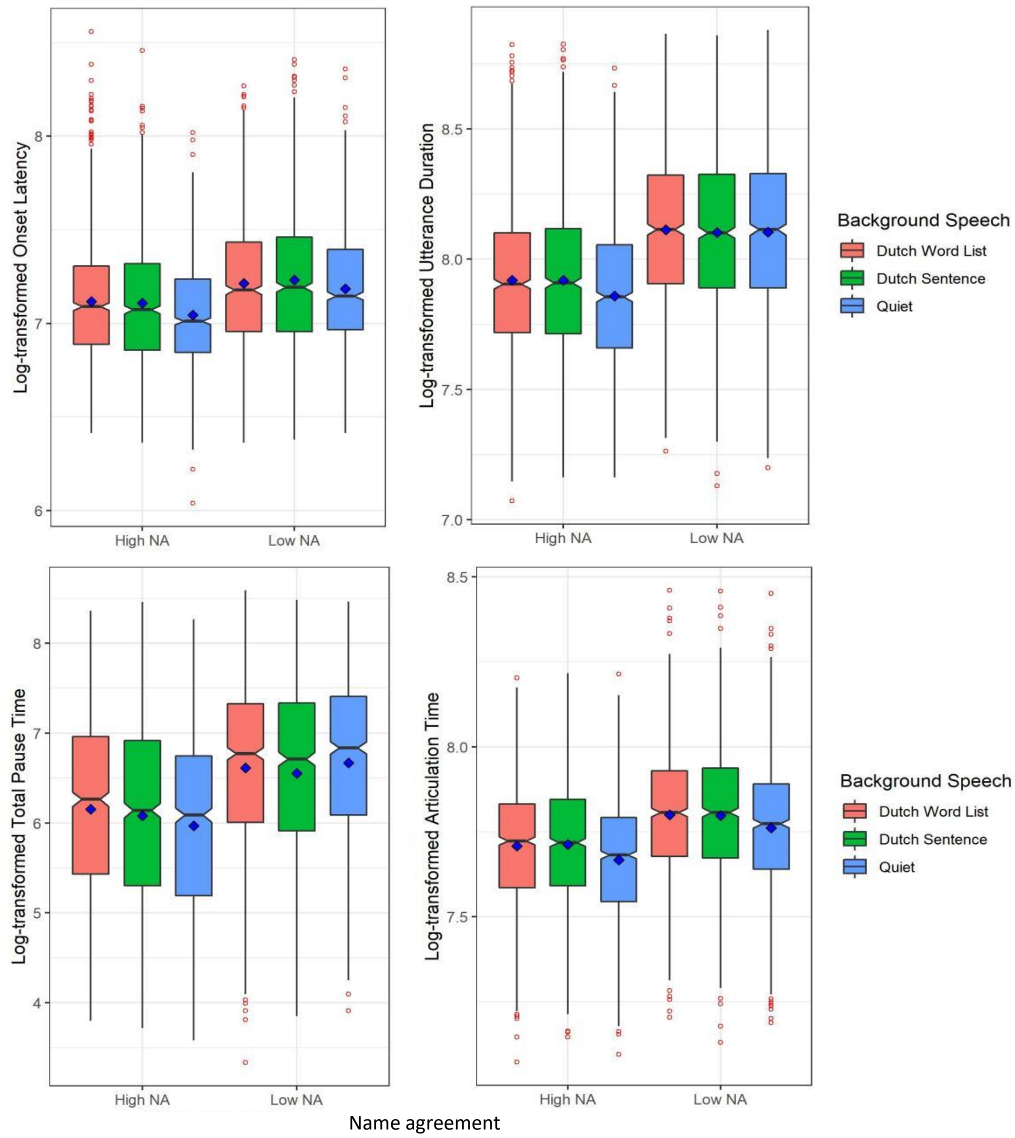

Figure 4.4. Log-transformed onset latency (top left), log-transformed utterance duration (top right), log-transformed total pause time (bottom left), and log-transformed articulation time (bottom right) split by name agreement (NA: high, low) and the type of background speech (Dutch word list, Dutch sentence, Quiet) in Experiment 2. Blue squares represent condition means and red points reflect outliers.

Total chunk number. As shown in Table 4.5 and Figure 4.5 (left), a Bayesian mixed-effect model showed that participants grouped their responses in more chunks for low name agreement pictures than high name agreement pictures (β= -0.266, SE = 0.03, 95% Cr.I = [-0.325, -0.208]). Total chunk number was not impacted by the type of background speech. Again, name agreement did not interact with the type of background speech.

First chunk length. As shown in Table 4.5 and the right panel of Figure 4.5, a Bayesian mixed-effect model showed that participants planned fewer names in their first response chunk for low name agreement pictures than high name agreement pictures (β= 0.237, SE = 0.027, 95% Cr.I = [0.183, 0.291]). First chunk length was not impacted by the type of background speech. Again, name agreement did not interact with the type of background speech.

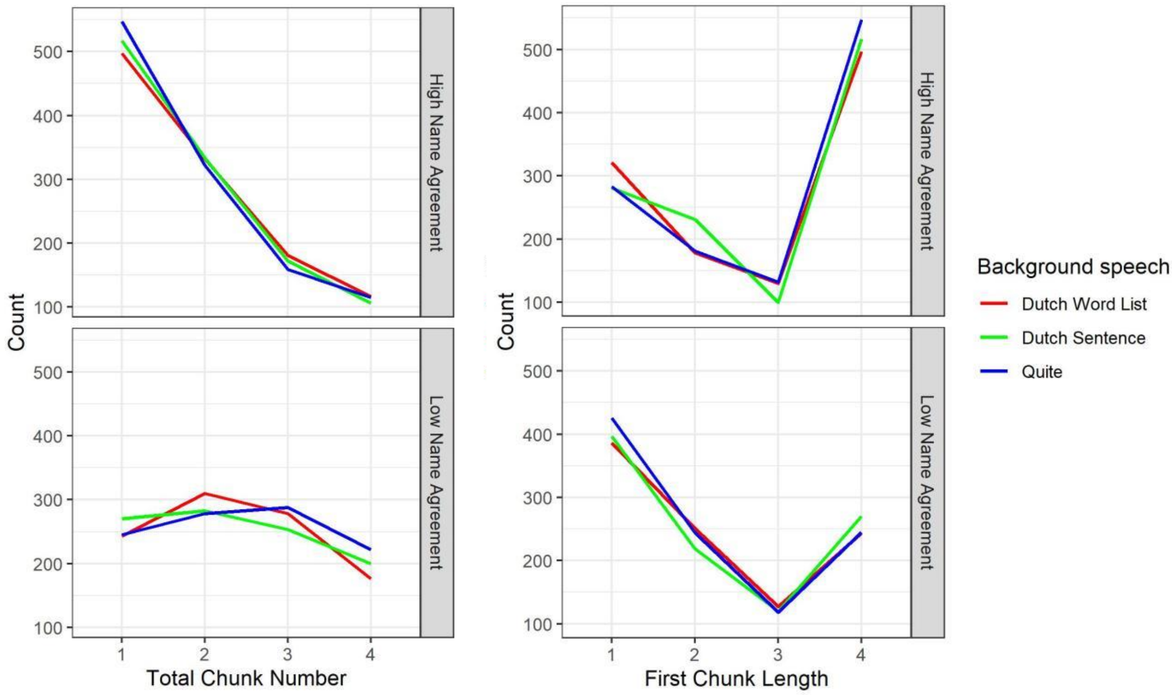

Figure 4.5. Total chunk number (left) and first chunk length (right) split by name agreement (NA: high, low) and the type of background speech (Dutch word list, Dutch sentence, Quiet) in Experiment 2.

Table 4.5. Results of Bayesian mixed-effect models for all dependent variables in Experiment 2.

		Estimate	Est.error	95% Cr. I lower	95% Cr. I upper	Effective samples
Accuracy						
Population-level effects	Intercept	2.295	0.165	1.974	2.628	29013
	Name Agreement	**1.061**	**0.223**	**0.630**	**1.506**	**79513**
	Speech vs. Quiet	-0.043	0.142	-0.328	0.230	118039
	Word List vs. Sentence	0.016	0.123	-0.231	0.256	109284
	NA × (S vs. Q)	-0.134	0.275	-0.669	0.412	118838
	NA × (WL vs. S)	0.063	0.246	-0.416	0.553	112914
Group-level effects	Participants					
	sd(Intercept)	0.812	0.103	0.634	1.038	28016
	sd(NA)	0.317	0.135	0.043	0.582	25107
	sd(Svs.Q)	0.171	0.123	0.007	0.455	45424
	sd(WLvs.S)	0.125	0.093	0.005	0.345	54483
	sd(NA×(Svs.Q))	0.220	0.169	0.008	0.630	64394
	sd(NA×(WLvs.S))	0.236	0.178	0.009	0.663	53301
	Items					
	sd(Intercept)	0.478	0.265	0.020	0.868	2980
	sd(NA)	0.901	0.531	0.034	1.714	3066
	sd(Svs.Q)	0.340	0.189	0.021	0.715	19407
	sd(WLvs.S)	0.315	0.187	0.017	0.692	18572
	sd(NA×(Svs.Q))	0.652	0.371	0.039	1.394	21918
	sd(NA×(WLvs.S))	0.601	0.366	0.030	1.338	18389
Log-transformed onset latency						
Population-level effects	Intercept	7.161	0.028	7.105	7.216	5610
	Name Agreement	**-0.128**	**0.014**	**-0.155**	**-0.1**	**60813**
	Speech vs. Quiet	*0.076*	*0.04*	*-0.003*	*0.155*	*61479*
	Word List vs. Sentence	-0.004	0.046	-0.096	0.086	65617

	NA × (S vs. Q)	0.04	0.074	-0.104	0.187	64085
	NA × (WL vs. S)	0.022	0.086	-0.147	0.19	66181
Group-level effects	Participants					
	sd(Intercept)	0.171	0.02	0.136	0.217	12128
	sd(NA)	0.024	0.011	0.003	0.044	22175
	sd(Svs.Q)	0.05	0.014	0.021	0.078	26754
	sd(WLvs.S)	0.028	0.014	0.002	0.054	20076
	sd(NA×(Svs.Q))	0.027	0.02	0.001	0.074	39897
	sd(NA×(WLvs.S))	0.026	0.018	0.001	0.067	39453
	Items					
	sd(Intercept)	0.029	0.016	0.001	0.053	1183
	sd(NA)	0.059	0.031	0.003	0.107	1196
	sd(Svs.Q)	0.184	0.106	0.008	0.339	1012
	sd(WLvs.S)	0.233	0.117	0.016	0.405	2193
	sd(NA×(Svs.Q))	0.376	0.213	0.015	0.68	1029
	sd(NA×(WLvs.S))	0.454	0.237	0.029	0.807	2111
Log-transformed utterance duration						
Population-level effects	Intercept	8.012	0.028	7.957	8.067	4298
	Name Agreement	**-0.215**	**0.022**	**-0.257**	**-0.172**	**34356**
	Speech vs. Quiet	*0.050*	*0.031*	*-0.012*	*0.111*	*48720*
	Word List vs. Sentence	0.005	0.024	-0.042	0.052	54738
	NA × (S vs. Q)	0.070	0.060	-0.047	0.187	50417
	NA × (WL vs. S)	-0.007	0.047	-0.100	0.085	58527
Group-level effects	Participans					
	sd(Intercept)	0.171	0.021	0.136	0.216	11188
	sd(NA)	0.073	0.011	0.054	0.097	31638
	sd(Svs.Q)	0.045	0.014	0.014	0.072	16224
	sd(WLvs.S)	0.008	0.006	0.000	0.023	55147
	sd(NA×(Svs.Q))	0.039	0.027	0.002	0.097	21573

	sd(NA×(WLvs.S))	0.019	0.014	0.001	0.054	45545
	Items					
	sd(Intercept)	0.044	0.023	0.002	0.078	1561
	sd(NA)	0.085	0.046	0.004	0.155	1554
	sd(Svs.Q)	0.151	0.065	0.021	0.253	2658
	sd(WLvs.S)	0.112	0.059	0.006	0.200	1808
	sd(NA×(Svs.Q))	0.301	0.130	0.040	0.504	2617
	sd(NA×(WLvs.S))	0.225	0.119	0.012	0.401	1766
Log-transformed total pause time						
Population-level effects	Intercept	6.298	0.09	6.12	6.476	8463
	Name Agreement	**-0.599**	**0.072**	**-0.741**	**-0.458**	**50058**
	Speech vs. Quiet	0.055	0.086	-0.114	0.224	74556
	Word List vs. Sentence	0.059	0.068	-0.075	0.194	87601
	NA × (S vs. Q)	*0.28*	*0.173*	*-0.06*	*0.621*	*74891*
	NA × (WL vs. S)	-0.006	0.137	-0.275	0.263	88114
Group-level effects	Participants					
	sd(Intercept)	0.542	0.065	0.432	0.687	16813
	sd(NA)	0.28	0.042	0.207	0.373	38849
	sd(Svs.Q)	0.078	0.051	0.004	0.188	27262
	sd(WLvs.S)	0.035	0.027	0.001	0.099	55607
	sd(NA×(Svs.Q))	0.28	0.12	0.035	0.51	25088
	sd(NA×(WLvs.S))	0.117	0.078	0.005	0.29	35367
	Items					
	sd(Intercept)	0.125	0.067	0.007	0.227	2808
	sd(NA)	0.249	0.134	0.014	0.455	2789
	sd(Svs.Q)	0.401	0.163	0.067	0.665	4686
	sd(WLvs.S)	0.297	0.168	0.012	0.549	2653
	sd(NA×(Svs.Q))	0.786	0.326	0.123	1.322	4524
	sd(NA×(WLvs.S))	0.589	0.337	0.024	1.099	2693

Log-transformed articulation time						
Population-level effects	Intercept	7.744	0.021	7.704	7.785	8367
	Name Agreement	**-0.093**	**0.020**	**-0.133**	**-0.054**	**63460**
	Speech vs. Quiet	**0.054**	**0.016**	**0.023**	**0.085**	**97570**
	Word List vs. Sentence	-0.003	0.013	-0.029	0.022	100970
	NA × (S vs. Q)	0.010	0.030	-0.048	0.069	103634
	NA × (WL vs. S)	0.000	0.026	-0.050	0.051	101332
Group-level effects	Participants					
	sd(Intercept)	0.120	0.014	0.096	0.152	16082
	sd(NA)	0.055	0.008	0.042	0.071	33143
	sd(Svs.Q)	0.031	0.007	0.018	0.046	24300
	sd(WLvs.S)	0.007	0.005	0.000	0.018	43960
	sd(NA×(Svs.Q))	0.033	0.017	0.002	0.067	20736
	sd(NA×(WLvs.S))	0.017	0.011	0.001	0.041	37705
	Items					
	sd(Intercept)	0.042	0.025	0.001	0.078	1772
	sd(NA)	0.083	0.051	0.003	0.156	1798
	sd(Svs.Q)	0.066	0.040	0.002	0.124	1927
	sd(WLvs.S)	0.058	0.035	0.002	0.108	2217
	sd(NA×(Svs.Q))	0.130	0.080	0.004	0.247	1977
	sd(NA×(WLvs.S))	0.116	0.069	0.004	0.217	2209
Total chunk number						
Population-level effects	Intercept	0.728	0.041	0.647	0.808	8660
	Name Agreement	**-0.266**	**0.030**	**-0.325**	**-0.208**	**41811**
	Speech vs. Quiet	-0.003	0.037	-0.077	0.071	73370
	Word List vs. Sentence	0.015	0.030	-0.045	0.074	77365
	NA × (S vs. Q)	0.070	0.075	-0.078	0.217	74377
	NA × (WL vs. S)	0.014	0.061	-0.105	0.133	79264

	Participants					
	sd(Intercept)	0.246	0.030	0.196	0.312	15554
	sd(NA)	0.086	0.022	0.045	0.132	47199
	sd(Svs.Q)	0.024	0.019	0.001	0.070	62041
	sd(WLvs.S)	0.020	0.015	0.001	0.057	68947
	sd(NA×(Svs.Q))	0.051	0.040	0.002	0.148	61109
Group-level effects	sd(NA×(WLvs.S))	0.040	0.031	0.002	0.114	70155
	Items					
	sd(Intercept)	0.047	0.026	0.002	0.092	4816
	sd(NA)	0.094	0.052	0.005	0.184	4829
	sd(Svs.Q)	0.140	0.066	0.012	0.257	7236
	sd(WLvs.S)	0.102	0.057	0.005	0.204	6819
	sd(NA×(Svs.Q))	0.278	0.132	0.023	0.512	7343
	sd(NA×(WLvs.S))	0.201	0.114	0.010	0.407	6661
First chunk length						
	Intercept	0.858	0.045	0.767	0.948	8363
	Name Agreement	**0.237**	**0.027**	**0.183**	**0.291**	**74876**
Population-level effects	Speech vs. Quiet	-0.008	0.043	-0.092	0.076	64681
	Word List vs. Sentence	-0.022	0.036	-0.093	0.048	70214
	NA × (S vs. Q)	-0.090	0.085	-0.257	0.078	65380
	NA × (WL vs. S)	-0.005	0.072	-0.146	0.137	70142
	Participants					
	sd(Intercept)	0.272	0.034	0.214	0.346	17057
	sd(NA)	0.030	0.021	0.001	0.079	35240
Group-level effects	sd(Svs.Q)	0.026	0.019	0.001	0.073	58663
	sd(WLvs.S)	0.021	0.016	0.001	0.060	67790
	sd(NA×(Svs.Q))	0.059	0.044	0.002	0.164	54199
	sd(NA×(WLvs.S))	0.040	0.031	0.002	0.115	72032
	Items					

sd(Intercept)	0.050	0.027	0.003	0.095	4599
sd(NA)	0.100	0.053	0.006	0.190	4610
sd(Svs.Q)	0.185	0.064	0.049	0.300	8825
sd(WLvs.S)	0.150	0.063	0.020	0.258	6981
sd(NA×(Svs.Q))	0.367	0.128	0.093	0.595	9005
sd(NA×(WLvs.S))	0.301	0.125	0.040	0.519	7420

Note. Models for all dependent variables were run for 24000 iterations. Bolded values indicate effects where the 95% Cr.I does not contain zero; Italicized values indicate effects where the beta estimate is twice the estimate of the standard error. NA refers to name agreement, WL refers to word list, S refers to sentence, Q refers to quiet.

4.3.3 Discussion

The results of Experiment 2 were remarkably similar to those of Experiment 1. First, we replicated the robust name agreement effects on all dependent measures, demonstrating again that high demand of lexical selection decreases planning speed and reduces planned utterance units in each response for multiple-object naming. Second, the presence or absence of background speech, now in the participants' native language, increased onset latencies and articulation time, and also had weak impact on utterance durations. The results suggest that intelligible background speech disrupts speech production relative to a quiet condition.

Third, contrary to our predictions, there were no systematic differences between the Dutch word list and sentence conditions on any of the dependent measures. As explained above, one could reasonably expect that word lists would interfere more with speech production than sentences (because the presence / absence of pauses may capture more attention, or the words in lists may capture attention more effectively when they are preceded by pauses), or that word lists would interfere less than sentences (because processing is less demanding when no syntactic / semantic integration processes occur). However, neither of these results were obtained. One possibility is that the absence of the word lists versus sentence effect may be because the disruption resulting from aspecific and specific attention capture may have canceled each other out. For now, we note that we obtained the same pattern as for Chinese background speech: an effect of background speech, but no effect of word lists versus sentences. We discuss this further in the General Discussion.

Finally, we found, different from Experiment 1, a weak modulation of name agreement on the processing of background speech, such that Dutch background speech (relative to the quiet control condition) increased the total pause time during the planning of high, but not low, name agreement pictures. This is consistent with earlier work by He et al. (2021) and suggests that stronger attentional engagement in the more difficult low name agreement condition leads to less interference from background speech. This implies that the disruption by intelligible background speech can be eliminated when the speech production task is sufficiently difficult.

4.4 General Discussion

In two experiments, we explored how different types of unintelligible (Experiment 1) and intelligible (Experiment 2) background speech affected spoken language production, with a focus on their impact on lexical selection in speech planning. There were four major findings. First, we obtained consistent name agreement effects on all measures in both experiments, showing that pictures with low name agreement decreased naming accuracies, slowed down planning speed, and reduced planned utterance units in each response relative to those with high name agreement. Second, background speech in both experiments disrupted speech production relative to a quiet condition, showing significantly increased articulation time and a weak impact on onset latencies in Experiment 1 (i.e., Chinese background speech), and significantly increased onset latencies and articulation time but a weak impact on utterance duration in Experiment 2 (i.e., Dutch background speech). Third, no systematic difference between word lists and sentences was found in either experiment. Finally, the disruption by Chinese background speech in Experiment 1 was not affected by name agreement, but the disruptive effects by Dutch background speech in Experiment 2 were mildly modulated by the demand of lexical selection in speech production, such that Dutch background speech resulted in increased total pause time for high, but not for low, name agreement pictures. Combined, these findings suggest that the presence of background speech, regardless of its intelligibility, interferes with speech production, and disruption caused by intelligible background speech can be eliminated by the attention engagement associated with difficult speech production.

4.4.1 Lexical selection demand affects speech production

The largest effect across both experiments was the effect of name agreement (indexing lexical selection demands in production), which was obtained on all measures. The pattern of results replicated the finding of earlier work using both single and multiple-picture naming paradigms (e.g., Alario et al., 2004; He et al., 2021; Shao et al., 2014). The name agreement effects on time measures, including onset latencies, utterance durations, total pause time, and articulation time, suggest that the demand of lexical selection affects processing before and after speech onset. This also indicates that speakers retrieve picture names during the whole process of planning sequence of picture names, which is consistent with the claim that speakers plan speech incrementally because they cannot retrieve all picture names before articulation but have to coordinate the planning and articulation of successive words with each other (e.g., Levelt et al., 1999; Roelofs, 1998; Wheeldon & Lahiri, 1997). The finding that name agreement affected response chunking measures indicates that increased lexical selection demand reduced planned utterance units in each response, which may reflect that speakers tend to plan names with less temporal overlap, resulting in more and shorter response chunks, for pictures with low, compared to high name agreement.

The name agreement effects obtained in both experiments suggest the effects are stable and consistent. This is also demonstrated by the finding that the magnitude of name agreements was similar across the two experiments, showing that name agreement (high versus low) did not interact with experiment (Experiment 1, Experiment 2) on any of the measures (see Appendix E, Table E1). Moreover, the present study differed in several aspects from the study by He et al. (2021). Here, we employed different error-coding criteria such that any reasonable name (rather than only the first and second most common names) to a picture was coded as correct, presented four (rather than six) pictures per trial, and we used Bayesian mixed-effect models rather than linear mixed-effect models. We replicated the finding in He et al., (2021) that name agreement affected accuracy, time measures, and response chunking measures. This suggests that name agreement effects in multi-word production are stable across different error-coding criteria, picture presentation ways, statistical methods, and groups of participants.

4.4.2 Irrelevant background speech disrupts speech production relative to a quiet condition

The present study showed clear irrelevant speech effects on articulation time in both experiments, with increased articulation time in the background speech conditions (word lists and sentences) relative to a quiet control condition. This result is in line with the prediction by the phonological disruption view (Salamé & Baddeley, 1982, 1989) that any background speech (no matter whether it is intelligible or not) should disrupt speech production due to the similarity of phonological codes between the focal task and background speech. One might predict though that the disruption by Dutch background speech may be due to similarity in both phonology and semantics. If this is the case, the disruption by intelligible background speech should be larger in Experiment 2 than in Experiment 1. However, in an analysis including Experiment as a factor, we found no interaction between the first contrast of background speech (i.e., Speech vs. Quiet) and experiment (Experiment 1, Experiment 2) on the measure of articulation time (see Appendix E, Table E1), which implies that the magnitude of disruption was similar across experiments. This argues against the importance of semantic similarity in disrupting speech planning.

We also found that the presence of background speech increased onset latencies relative to a quiet condition in both experiments, although the effects were weak. This finding suggests that background speech, regardless of its intelligibility, interferes with initial speech planning in multi-picture naming. As with articulation time, no interaction between the first contrast of background speech (speech vs. quiet) and experiment was obtained (see Appendix E, Table E1). This again indicates that the disruptive effect is largely phonological and not semantic in nature. Earlier results from He et al., (2021) showed that word lists (regardless of intelligibility) interfered with onset latencies relative to a speech-like noise condition (i.e., eight-talker babble), which excludes a possible contribution of low-level acoustic properties shared between speech production and speech-like noise. Thus, these results are largely in agreement with the phonological disruption view (Salamé & Baddeley, 1982, 1989).

One major difference between Experiment 1 and Experiment 2 is that auditory disruption by Dutch background speech also appeared as a weak effect on utterance durations, which was supported by an interaction between the first contrast of background speech (speech vs. quiet) and experiment (See Appendix E, Table E1). This is consistent with He et al. (2021), where Dutch word lists increased utterance durations relative to Chinese word lists, indicating that intelligible background speech elicits more disruption than unintelligible background speech. One possibility is that this effect arises because intelligible background speech interferes with planning done between producing chunks of words. This is partially supported by the result that Dutch speech elicited more disruption on total pause time relative to a quiet condition for high name agreement pictures. Alternatively, the extra disruption on utterance duration may result from similarity in semantics, in addition to phonology. That is, intelligible background speech activates semantic representations that lead to code conflict with the processing, but only after starting to speak.

Note that the extra disruption on utterance duration by intelligible background speech could also be caused by an attention capture mechanism. That is, it may be argued that compared with unintelligible background speech, intelligible background speech activates linguistic information at multiple levels, captures more attention, and interferes more with speech production. Yet, the contrast of background speech with quiet in the present study does not provide direct evidence for or against this argument.

4.4.3 The presence of pauses in background speech does not disrupt speech production

In contrast to robust differences between background speech and quiet conditions, we did not observe any difference between the two speech conditions, word lists and sentences, in either experiment. The absence of the word lists versus sentences effect in Experiment 1 may reflect that the manipulation of aspecific context variation (the presence / absence of pauses) in unintelligible background speech does not elicit disruption on speech production, which goes against the aspecific attention capture view (Eimer et al., 1996).

However, the absence of evidence in favor of an aspecific attention capture view implies that stimulus-specific variation in background speech may interfere with speech production. In other words, one would expect that the background speech with more linguistic information (i.e., the semantic / syntactic information in Dutch sentences) should disrupt speech production more than that with less linguistic representations (i.e., Dutch word lists). However, we did not find any difference between Dutch word lists and sentences on any measures in Experiment 2. The absence of word lists versus sentences effects in Dutch can be accounted for in at least three ways. First, the manipulation of stimulus-specific variation does not actually matter in this study because the background speech stimuli were too uniform and boring (i.e., word lists had a regular acoustic pattern, sentences had uniform syntactic structure), which may not have been very engaging to our participants. Second, our manipulation of stimulus-specific variation matters, but aspecific and specific effects have canceled each other out. In other words, the disruption by the presence of pauses (i.e., aspecific context variation) in Dutch word lists canceled the interference by richer linguistic information of semantic / syntactic integration (i.e., specific linguistic variation) in Dutch sentences.

The last possibility for the lack of the word lists versus sentences effect is that the stimulus-specific effect indeed exists, but it was too small and attenuated over repetition of stimuli. This is because all stimuli were presented three times across three blocks in the present study. To test this possibility, we conducted all analyses including the repetition (i.e., block) as a within-participant factor. However, we did not find any interaction between the contrast of word list versus sentence and the block in either experiment (see Appendices, Table B1 for

Experiment 1; Table D1 for Experiment 2), which suggests that there is no evidence that the disruptive effect by stimulus-specific variation decreases with repetition.

Combined, the absence of the word lists versus sentences effect in unintelligible background speech fails to support the aspecific attention capture view (Eimer et al., 1996), which in turn implies that the stimuli-specific variation may capture attention. However, we did not find any stimuli-specific effect in intelligible background speech either, which could be because the manipulation of stimulus-specific variation may not work, or because the stimulus-specific effect may have canceled the stimulus-aspecific effect in the present study. The two possibilities cannot be distinguished here and need to be tested by future research.

4.4.4 The modulation of name agreement on the irrelevant speech effects

The interaction between the type of background speech and name agreement was absent in Experiment 1, but present in Experiment 2, which is consistent with the predictions of the attention engagement account (Halin et al., 2014; Marsh et al., 2015). That is, disruption by Chinese background speech remains unaffected by changes in attention engagement (i.e., the name agreement manipulation) because the processing of unintelligible auditory input is automatic and escapes cognitive control (Hughes, 2014). In contrast, the interference by Dutch background speech is reduced by an increase in attention engagement (i.e., on the low name agreement trials), because the processing of intelligible background speech requires central attention that taps into cognitive control (Marsh et al., 2018). Since the evidence for this is a weak interaction on total pause time in Experiment 2, the implication is that intelligible background speech can be tuned out by performing lexical selection during pauses between articulations. This differs from He et al., (2021) who found a similar interaction, but only on the measure of onset latencies, suggesting that auditory disruption can be blocked out during planning before beginning speaking. The inconsistency may be due to variations in baseline tasks (quiet in the present study but the eight-talker babble in He et al., 2021), in statistical approaches (Bayesian mixed-effect models in the present study but linear mixed-effect mode in He et al., 2021), or in speech production task (naming four pictures in the present study but naming six pictures in He et al., 2021). Future work is needed to determine the cause of the difference.

4.4.5 Outlook

While the present study provides some insights into how different types of background speech interfere with speech production, more work is needed to reveal how speakers plan their speech in the presence of background speech. For example, the disruption may have been relatively weak because the stimuli and task were relatively simple. As mentioned earlier, the background speech was uniform and boring, which may not have been very engaging to our participants. It may be more difficult to plan speech when background speech is more engaging, as engaging speech may make it harder to selectively attend to the speech production and filter out the irrelevant information. The present research used a multi-object naming task that was relatively easy, which may be less susceptible to the disruption by background speech. Therefore, future studies should utilize more difficult speech production tasks such as phrase or sentence production.

Moreover, we saw clear evidence for phonological disruption but we did not see any evidence for semantic interference, which may be because the background speech stimuli was not sufficiently related to the speech produced by the participants. This hypothesis could be tested by using background speech that is semantically related or unrelated to the target words for production, thereby assessing the role of conceptual or lexical competition in speaking with background speech.

Finally, we note that disruption from background speech may be larger in certain participant populations. For example, children or older people may show larger effects than young adults due to their poorer control of attention and their weaker ability to filter out task-irrelevant stimuli. This is an avenue that should be explored in future research, informing theories of irrelevant speech effects and contribute to a better understanding of how people plan speech in noisy environments.

4.5 Conclusion

Two online experiments using a speaking-while-listening paradigm showed that irrelevant background speech (regardless of its intelligibility) disrupts speech production relative to a quiet condition, although intelligible background speech elicits extra disruption. The finding stresses the importance of similarity in phonological representations between the speech production and background speech in eliciting interference. Moreover, the absence of differences between the word lists and sentences conditions in unintelligible background speech suggests that the aspecific properties of background speech may not capture attention and cause a drop on naming performance. Finally, while intelligible background speech has more detrimental impact on speech production, the impact can be reduced through greater engagement with the task, e.g., increasing the difficulty of speech production. The implication is that when the disruption by background speech occurs in speech production, speakers may be able to manage this disruption by changing when and how they plan their speech.

Acknowledgments

We would like to thank Maarten van den Heuvel and Thijs Rinsma for programming; Annelies van Wijngaarden and Sophie Slaats for translating and recording materials; Dennis Joosen, Esther de Kerf, Elizardo Laclé, Elsa Opheij, Marije Veeneman, and Sanne van Eck for data coding.

Appendices

Appendix A: Stimuli used in Experiment 1

Table A1. 240 pictures used in both Experiments.

Picture Grid	*Picture 1*	*Picture 2*	*Picture 3*	*Picture 4*	*Picture Grid*	*Picture 1*	*Picture 2*	*Picture 3*	*Picture 4*
Pictures with high name agreement									
1	koelkast	pijl	dolfijn	gevangenis	16	spiegel	ananas	robot	zaklamp
2	leeuw	kruiwagen	driehoek	tomaat	17	schilderij	tunnel	kangoeroe	broek
3	harp	radio	knie	paprika	18	sleutel	dobbelsteen	ketting	rechter
4	vlinder	trap	cactus	batterij	19	stopcontact	arm	ezel	diamant
5	zaag	kiwi	vliegtuig	bezem	20	kapper	zebra	aardbei	wolk
6	waaier	schaap	glas	baard	21	schaduw	kompas	geit	horloge
7	ster	konijn	doedelzak	handschoen	22	pompoen	vlieger	kaars	skelet
8	pijp	hamer	berg	duim	23	heks	aardappel	vleermuis	boog
9	eekhoorn	keuken	banaan	orkest	24	masker	bijbel	zwembad	kanon
10	kwal	slager	anker	vuist	25	schaar	rups	kraan	puzzel
11	microfoon	bloem	koning	stier	26	eiland	schildpad	clown	bril
12	kokosnoot	steen	gitaar	egel	27	fruit	vlag	aansteker	lepel

13	roos	trechter	kroon	ballon	28	kikker	wasmachine	bokser	trompet
14	slak	rug	weegschaal	honing	29	bus	fabriek	sok	vork
15	muis	drumstel	parachute	tandarts	30	papegaai	helikopter	toetsenbord	riem
Pictures with low name agreement									
1	jager	klauw	baksteen	trui	16	antenne	olie	piano	knuffel
2	lade	schedel	melk	foto	17	planeet	motor	litteken	gang
3	speer	nagel	kerkhof	duif	18	komkommer	badkamer	domino	wortels
4	engel	parel	troon	viool	19	schatkist	elf	koffie	put
5	kasteel	snoepje	brievenbus	vogelkooi	20	schelp	prullenbak	ridder	meloen
6	kerk	schoolbord	bank	walrus	21	hengel	gevangene	brug	driewieler
7	soldaat	vis	gorilla	kruk	22	vinger	magneet	zanger	plas
8	armband	rimpels	kogel	hagedis	23	blad	raam	jurk	hoorn
9	ijsje	spuit	paus	badkuip	24	rivier	monster	pion	goochelaar
10	broekzak	naald	varken	wasbak	25	rugzak	chocolade	balkon	schep
11	staart	inktvis	herder	perzik	26	koekje	garage	cirkel	mossel
12	sigaret	ijsberg	hersenen	kwast	27	camping	pruik	sneeuw	ballerina
13	gymzaal	leraar	handdoek	worst	28	munt	strand	kameel	lamp
14	museum	tuinslang	druif	kegel	29	kleed	tram	doodskist	garnaal
15	koningin	buik	trein	soep	30	haven	bliksem	schrift	kaarten

Table A2. 20 Chinese word lists used in Experiment 1.

	Noun 1	*Noun 2*	*Noun 3*	*Noun 4*	*Noun 5*	*Noun 6*
List1	剑	苍蝇	梨	画家	暖气	幸运草
List2	肉	火箭	羽毛	鞋带	正方形	树枝
List3	美洲豹	邮票	胸	电视	剃刀	发梳
List4	奶酪	枭	植物	救护车	眼睛	手鼓
List5	老鹰	火	风扇	纽扣	鼓	摄影师
List6	巢	早餐	樵夫	屁股	立方体	铁刷
List7	鸟	船舵	刽子手	嘴唇	温室	步枪
List8	手风琴	肩膀	秃鹫	鞋	衣柜	骨头
List9	肺	盆子	栅栏	计算器	迷宫	蛇
List10	仙女	奖章	船	秃头	桌子	面包机
List11	树	火山	袋子	磨坊	鳄鱼	洋娃娃
List12	波浪	橄榄	钉子	相机	音乐会	鹅
List13	机场	杯子	肥皂	狼	盒子	向日葵
List14	血管	帽子	文件夹	河马	烟	豆子
List15	橡子	游泳者	盘子	钱包	鸡	眉毛
List16	独木舟	戒指	西瓜	马	公主	椅子
List17	渔夫	合唱团	足球	苹果	超市	鹿
List18	瓶塞	灭火器	柠檬	香水	铅笔	锁
List19	盐	坦克	奶牛	服务员	黄金	床垫
List20	裙子	电缆	脚	摇篮	护士	水族馆

Table A3. 20 Chinese sentences used in Experiment 1.

No.	Chinese sentences
1	幸运草和暖气在画家的左边，并且梨和苍蝇在剑的右边。
2	树枝和正方形在鞋带的左边，并且羽毛和火箭在肉的左边。
3	发梳和剃刀在电视的右边，并且胸和邮票在美洲豹的左边。
4	手鼓和眼睛在救护车的右边，并且植物和枭在奶酪的右边。
5	摄影师和鼓在纽扣的左边，并且风扇和火在老鹰的右边。
6	铁刷和立方体在屁股的左边，并且樵夫和早餐在巢的左边。
7	步枪和温室在嘴唇的右边，并且刽子手和船舵在鸟的左边。
8	骨头和衣柜在鞋的右边，并且秃鹫和肩膀在手风琴的右边。
9	蛇和迷宫在计算器的左边，并且栅栏和盆子在肺的右边。
10	面包机和桌子在秃头的左边，并且船和奖章在仙女的左边。
11	洋娃娃和鳄鱼在磨坊的右边，并且袋子和火山在树的左边。
12	鹅和音乐会在相机的右边，并且钉子和橄榄在波浪的右边。
13	向日葵和盒子在狼的左边，并且肥皂和杯子在机场的右边。
14	豆子和烟在河马的左边，并且文件夹和帽子在血管的左边。
15	眉毛和鸡在钱包的右边，并且盘子和游泳者在橡子左边。
16	椅子和公主在马的右边，并且西瓜和戒指在独木舟的右边。
17	鹿和超市在苹果的左边，并且足球和合唱团在渔夫的右边。
18	锁和铅笔在香水的左边，并且柠檬和灭火器在瓶塞的左边。
19	床垫和黄金在服务员右边，并且奶牛和坦克在盐的左边。
20	水族馆和护士在摇篮的右边，并且脚和电缆在裙子的右边。

Table A4. A questionnaire of Chinese experience in Experiment 1.

Tot slot willen we je vragen om een aantal vragen te beantwoorden over jouw ervaring met Mandarijn Chinees. Nadat je een vraag hebt aangevinkt, dien je op ‘Volgende’ te klikken om naar de volgende vraag te gaan.

1) Ben je in een land geweest waar Mandarijn Chinees wordt gesproken? Zo ja, hoeveel maanden?

A. Nooit B. <3 maanden C. 3-6 maanden D. 6-12 maanden E. >12 maanden

2) Ben je bij een gezin geweest waar Mandarijn Chinees wordt gesproken? Zo ja, hoeveel maanden?

A. Nooit B. <3 maanden C. 3-6 maanden D. 6-12 maanden E. >12 maanden

3) Ben je in een school/werkomgeving geweest waar Mandarijn Chinees wordt gesproken? Zo ja, hoeveel maanden?

A. Nooit B. <3 maanden C. 3-6 maanden D. 6-12 maanden E. >12 maanden

4) Gebruik onderstaande schaal, waar 0 “helemaal geen kennis” is, en 10 “vloeiend, alsof het je moedertaal is”. Geef aan wat jouw vaardigheidsniveau is op het gebied van het spreken, verstaan en lezen van Mandarijn Chinees.

A. Spreken van Mandarijn Chinees: 0 1 2 3 4 5 6 7 8 9 10

B. Verstaan van gesproken Mandarijn Chinees: 0 1 2 3 4 5 6 7 8 9 10

C. Lezen van Mandarijn Chinees: 0 1 2 3 4 5 6 7 8 9 10

5) Gebruik onderstaande schaal, waar 0 “helemaal geen kennis” is, en 10 “vloeiend, alsof het je moedertaal is”. Geef aan in hoeverre je op dit moment blootgesteld wordt aan Mandarijn Chinees in de volgende situaties.

A. Contact hebben met Chinese vrienden: 0 1 2 3 4 5 6 7 8 9 10

B. Kijken van Chinese TV: 0 1 2 3 4 5 6 7 8 9 10

C. Luisteren naar Chinese radio/muziek: 0 1 2 3 4 5 6 7 8 9 10

D. Lezen van Chinese boeken/tijdschriften: 0 1 2 3 4 5 6 7 8 9 10

Appendix B: Results of block analysis in Experiment 1

Table B1. Results of block analysis in Experiment 1.

		Estimate	Est.error	95% Cr. I lower	95% Cr. I upper	Effective samples
Log-transformed onset latency						
Population-level effects	Intercept	7.134	0.028	7.079	7.19	5611
	Name Agreement	**-0.121**	**0.015**	**-0.15**	**-0.092**	**60182**
	Speech vs. Quiet	**0.062**	**0.024**	**0.015**	**0.11**	**59671**
	Word List vs. Sentence	0	0.021	-0.041	0.041	62160
	Block 12 vs. Block 3	**0.194**	**0.029**	**0.136**	**0.25**	**51032**
	Block 1 vs. Block 2	**0.245**	**0.028**	**0.19**	**0.299**	**42710**
	NA × (S vs. Q)	-0.004	0.042	-0.086	0.079	66494
	NA × (WL vs. S)	-0.019	0.039	-0.096	0.059	68857
	NA × (Block 12 vs. 3)	-0.035	0.046	-0.125	0.055	69736
	NA × (Block 1 vs. 2)	-0.01	0.037	-0.083	0.062	66348
	(S vs. Q) × (Block 12 vs. 3)	0.026	0.051	-0.074	0.126	74295
	(WL vs. S) × (Block 12 vs. 3)	-0.023	0.049	-0.12	0.075	67668
	(S vs. Q) × (Block 1 vs. 2)	0.093	0.047	0	0.185	65723
	(WL vs. S) × (Block 1 vs. 2)	-0.029	0.055	-0.136	0.078	70992
	NA × (S vs. Q) × (Block 12vs.3)	0.047	0.095	-0.138	0.233	77572
	NA × (WL vs. S) × (Block 12vs.3)	0.025	0.087	-0.146	0.194	82091
	NA × (S vs. Q) × (Block 1vs.2)	-0.017	0.082	-0.179	0.146	79468
	NA × (WL vs. S) × (Block 1vs.2)	-0.013	0.098	-0.205	0.18	76734
Log-transformed utterance duration						
	Intercept	8.021	0.023	7.975	8.067	6748
	Name Agreement	**-0.191**	**0.02**	**-0.23**	**-0.151**	**52806**
	Speech vs. Quiet	**0.03**	**0.012**	**0.006**	**0.054**	**85083**
	Word List vs. Sentence	-0.003	0.011	-0.025	0.019	87020
	Block 12 vs. Block 3	**0.168**	**0.019**	**0.132**	**0.205**	**49646**

	Block 1 vs. Block 2	**0.134**	**0.016**	**0.103**	**0.166**	**46638**
	NA × (S vs. Q)	0.015	0.024	-0.031	0.062	90001
	NA × (WL vs. S)	0.005	0.023	-0.041	0.051	80784
	NA × (Block 12 vs. 3)	**-0.101**	**0.025**	**-0.149**	**-0.052**	**87321**
Population-level effects	NA × (Block 1 vs. 2)	**-0.073**	**0.024**	**-0.12**	**-0.026**	**82973**
	(S vs. Q) × (Block 12 vs. 3)	-0.022	0.053	-0.125	0.083	63183
	(WL vs. S) × (Block 12 vs. 3)	-0.066	0.046	-0.156	0.025	65632
	(S vs. Q) × (Block 1 vs. 2)	0.031	0.049	-0.066	0.127	64491
	(WL vs. S) × (Block 1 vs. 2)	-0.029	0.04	-0.107	0.049	61714
	NA × (S vs. Q) × (Block 12vs.3)	0.033	0.096	-0.156	0.22	69797
	NA × (WL vs. S) × (Block 12vs.3)	-0.005	0.085	-0.171	0.163	73221
	NA × (S vs. Q) × (Block 1vs.2)	-0.048	0.09	-0.224	0.129	74468
	NA × (WL vs. S) × (Block 1vs.2)	0.03	0.073	-0.113	0.173	69539
Log-transformed total pause time						
	Intercept	5.019	0.291	4.447	5.59	4615
	Name Agreement	**-1.429**	**0.241**	**-1.904**	**-0.952**	**14775**
	Speech vs. Quiet	-0.428	0.238	-0.896	0.037	37330
	Word List vs. Sentence	-0.115	0.2	-0.505	0.278	45039
	Block 12 vs. Block 3	**1.131**	**0.22**	**0.699**	**1.562**	**29293**
	Block 1 vs. Block 2	**0.912**	**0.18**	**0.558**	**1.263**	**28534**
	NA × (S vs. Q)	-0.023	0.365	-0.74	0.7	68037
Population-level effects	NA × (WL vs. S)	0.137	0.348	-0.546	0.819	55847
	NA × (Block 12 vs. 3)	-0.05	0.419	-0.871	0.779	54403
	NA × (Block 1 vs. 2)	-0.214	0.302	-0.808	0.378	70396
	(S vs. Q) × (Block 12 vs. 3)	0.569	0.564	-0.544	1.676	57132
	(WL vs. S) × (Block 12 vs. 3)	-0.252	0.566	-1.361	0.864	55234
	(S vs. Q) × (Block 1 vs. 2)	0.118	0.475	-0.813	1.048	59261
	(WL vs. S) × (Block 1 vs. 2)	0.578	0.449	-0.309	1.458	55047
	NA × (S vs. Q) × (Block 12vs.3)	1.233	1.129	-0.994	3.441	48396

	NA × (WL vs. S) × (Block 12vs.3)	-0.12	1.101	-2.281	2.031	56935
	NA × (S vs. Q) × (Block 1vs.2)	-0.75	0.935	-2.586	1.093	63045
	NA × (WL vs. S) × (Block 1vs.2)	0.981	0.818	-0.619	2.586	59252

Note. NA refers to name agreement, WL refers to word lists, S refers to sentences. These results are for 36 participants who wore their headphones/earphones correctly.

Appendix C: Stimuli used in Experiment 2

Table C1. 20 Dutch word lists used in Experiment 2.

	Noun 1	*Noun 2*	*Noun 3*	*Noun 4*	*Noun 5*	*Noun 6*
List1	fee	medaille	boot	luipaard	zonnebloem	kers
List2	tak	beker	prinses	schild	veer	raket
List3	postzegel	vlees	jas	tamboerijn	map	kam
List4	plant	Kaas	accordeon	oog	scheermes	uil
List5	rekenmachine	mand	vulkaan	zeep	paard	kano
List6	gier	vierkant	schoen	ambulance	kast	boom
List7	krokodil	veter	tas	molen	pop	bot
List8	ring	slang	dienblad	hek	watermeloen	kubus
List9	nest	ontbijt	borstel	trommel	stoel	kruik
List10	potlood	Kurk	brandblusser	citroen	spons	vuur
List11	nijlpaard	koffer	spijker	camera	fakkel	boon
List12	vliegveld	Wolf	kopje	houthakker	doos	boter
List13	televisie	zwaard	voet	peer	schilder	klavertje
List14	vlieg	Rok	zuster	kabel	aquarium	wieg
List15	zwemmer	Lijst	bord	portemonnee	hert	koor
List16	ventilator	Zout	adelaar	tank	liniaal	brief
List17	koe	voetbal	goud	wortel	parfum	serveerster
List18	kas	Gans	tafel	verwarming	fotograaf	roer
List19	appel	theepot	knoop	vogel	wandelstok	slot
List20	pet	cadeau	haak	olijf	kip	visser

Table C2. 20 Dutch sentences used in Experiment 2.

No.	*Dutch sentences*
1	De kers en de zonnebloem bevinden zich links van het luipaard, en de boot en de medaille bevinden zich rechts van de fee.
2	De raket en de veer bevinden zich links van het schild, en de prinses en de beker bevinden zich links van de tak.
3	De kam en de map bevinden zich rechts van de tamboerijn, en de jas en het vlees bevinden zich links van de postzegel.
4	De uil en het scheermes bevinden zich rechts van het oog, en de accordeon en de kaas bevinden zich rechts van de plant.
5	De kano en het paard bevinden zich links van de zeep, en de vulkaan en de mand bevinden zich rechts van de rekenmachine.
6	De boom en de kast bevinden zich links van de ambulance, en de schoen en het vierkant bevinden zich links van de gier.
7	Het bot en de pop bevinden zich rechts van de molen, en de tas en de veter bevinden zich links van de krokodil.
8	De kubus en de watermeloen bevinden zich rechts van het hek, en het dienblad en de slang bevinden zich rechts van de ring.
9	De kruik en de stoel bevinden zich links van de trommel, en de borstel en het ontbijt bevinden zich rechts van het nest.
10	Het vuur en de spons bevinden zich links van de citroen, en de brandblusser en de kurk bevinden zich links van het potlood.
11	De boon en de fakkel bevinden zich rechts van de camera, en de spijker en de koffer bevinden zich links van het nijlpaard.
12	De boter en de doos bevinden zich rechts van de houthakker, en het kopje en de wolf bevinden zich rechts van het vliegveld.
13	Het klavertje en de schilder bevinden zich links van de peer, en de voet en het zwaard bevinden zich rechts van de televisie.
14	De wieg en het aquarium bevinden zich links van de kabel, en de zuster en de rok bevinden zich links van de vlieg.
15	Het koor en het hert bevinden zich rechts van de portemonnee, en het bord en de lijst bevinden zich links van de zwemmer.
16	De brief en de liniaal bevinden zich rechts van de tank, en de adelaar en het zout bevinden zich rechts van de ventilator.
17	De serveerster en het parfum bevinden zich links van de wortel, en het goud en de voetbal bevinden zich rechts van de koe.
18	Het roer en de fotograaf bevinden zich links van de verwarming, en de tafel en de gans bevinden zich links van de kas.
19	Het slot en de wandelstok bevinden zich rechts van de vogel, en de knoop en de theepot bevinden zich links van de appel.
20	De visser en de kip bevinden zich rechts van de olijf, en de haak en het cadeau bevinden zich rechts van de pet.

Appendix D: Results of block analysis in Experiment 2

Table D1. Results of block analysis in Experiment 2.

		Estimate	Est.error	95% Cr. I lower	95% Cr. I upper	Effective samples
Log-transformed onset latency						
Population-level effects	Intercept	7.161	0.028	7.106	7.217	4693
	Name Agreement	**-0.127**	**0.013**	**-0.153**	**-0.101**	**56007**
	Speech vs. Quiet	**0.076**	**0.022**	**0.033**	**0.119**	**55853**
	Word List vs. Sentence	-0.005	0.019	-0.043	0.033	59827
	Block 12 vs. Block 3	**0.236**	**0.027**	**0.183**	**0.288**	**36045**
	Block 1 vs. Block 2	**0.301**	**0.028**	**0.246**	**0.356**	**35931**
	NA × (S vs. Q)	0.043	0.039	-0.034	0.121	60049
	NA × (WL vs. S)	0.029	0.036	-0.043	0.1	61253
	NA × (Block 12 vs. 3)	-0.06	0.038	-0.136	0.014	61001
	NA × (Block 1 vs. 2)	-0.064	0.037	-0.137	0.009	62117
	(S vs. Q) × (Block 12 vs. 3)	0.074	0.051	-0.026	0.175	63417
	(WL vs. S) × (Block 12 vs. 3)	-0.01	0.043	-0.095	0.075	62381
	(S vs. Q) × (Block 1 vs. 2)	**0.221**	**0.048**	**0.126**	**0.315**	**56880**
	(WL vs. S) × (Block 1 vs. 2)	-0.045	0.046	-0.137	0.047	61468
	NA × (S vs. Q) × (Block 12vs.3)	-0.014	0.091	-0.19	0.165	68028
	NA × (WL vs. S) × (Block 12vs.3)	-0.046	0.081	-0.205	0.115	67893
	NA × (S vs. Q) × (Block 1vs.2)	-0.11	0.084	-0.274	0.056	70312
	NA × (WL vs. S) × (Block 1vs.2)	-0.024	0.086	-0.193	0.145	66811
Log-transformed utterance duration						
	Intercept	8.012	0.028	7.957	8.067	4964
	Name Agreement	**-0.214**	**0.022**	**-0.256**	**-0.171**	**36308**
	Speech vs. Quiet	**0.05**	**0.015**	**0.02**	**0.081**	**56830**
	Word List vs. Sentence	0.004	0.011	-0.018	0.027	72507
	Block 12 vs. Block 3	**0.189**	**0.018**	**0.153**	**0.225**	**34819**

	Block 1 vs. Block 2	**0.16**	**0.015**	**0.131**	**0.19**	**52287**
	NA × (S vs. Q)	**0.073**	**0.028**	**0.018**	**0.128**	**65023**
	NA × (WL vs. S)	-0.007	0.023	-0.051	0.038	69775
	NA × (Block 12 vs. 3)	**-0.095**	**0.026**	**-0.146**	**-0.045**	**70942**
Population-level effects	NA × (Block 1 vs. 2)	**-0.063**	**0.025**	**-0.112**	**-0.014**	**65090**
	(S vs. Q) × (Block 12 vs. 3)	-0.061	0.056	-0.17	0.049	50549
	(WL vs. S) × (Block 12 vs. 3)	-0.05	0.051	-0.15	0.051	48181
	(S vs. Q) × (Block 1 vs. 2)	0.014	0.049	-0.082	0.109	47859
	(WL vs. S) × (Block 1 vs. 2)	-0.021	0.044	-0.108	0.066	50218
	NA × (S vs. Q) × (Block 12vs.3)	0.097	0.096	-0.093	0.285	58207
	NA × (WL vs. S) × (Block 12vs.3)	0.096	0.09	-0.082	0.272	57433
	NA × (S vs. Q) × (Block 1vs.2)	0.052	0.089	-0.123	0.226	56100
	NA × (WL vs. S) × (Block 1vs.2)	0.066	0.08	-0.092	0.224	57018
Log-transformed total pause time						
	Intercept	6.294	0.088	6.121	6.468	6219
	Name Agreement	**-0.598**	**0.073**	**-0.741**	**-0.454**	**37565**
	Speech vs. Quiet	0.052	0.053	-0.052	0.156	74627
	Word List vs. Sentence	0.055	0.046	-0.036	0.146	77117
	Block 12 vs. Block 3	**0.475**	**0.07**	**0.338**	**0.612**	**40543**
	Block 1 vs. Block 2	**0.413**	**0.06**	**0.295**	**0.531**	**50115**
	NA × (S vs. Q)	**0.292**	**0.111**	**0.075**	**0.512**	**72640**
Population-level effects	NA × (WL vs. S)	-0.017	0.094	-0.202	0.167	78343
	NA × (Block 12 vs. 3)	**-0.27**	**0.101**	**-0.469**	**-0.07**	**77865**
	NA × (Block 1 vs. 2)	-0.138	0.097	-0.331	0.053	72022
	(S vs. Q) × (Block 12 vs. 3)	-0.041	0.185	-0.405	0.322	61523
	(WL vs. S) × (Block 12 vs. 3)	-0.03	0.173	-0.369	0.312	60175
	(S vs. Q) × (Block 1 vs. 2)	-0.046	0.175	-0.389	0.296	56617
	(WL vs. S) × (Block 1 vs. 2)	0.106	0.15	-0.189	0.402	57255
	NA × (S vs. Q) × (Block 12vs.3)	0.324	0.35	-0.364	1.013	67276
	NA × (WL vs. S) × (Block 12vs.3)	0.482	0.335	-0.179	1.136	64208

	NA × (S vs. Q) × (Block 1vs.2)	0.215	0.308	-0.388	0.821	63082
	NA × (WL vs. S) × (Block 1vs.2)	0.256	0.285	-0.306	0.816	64384

Note. NA refers to name agreement, WL refers to word lists, S refers to sentences. These results are for 36 participants who wore their headphones/earphones correctly.

Appendix E: Comparison of two experiments

Table E1. Results of Bayesian mixed-effect models across experiments.

		Estimate	Est.error	95% Cr. I		Effective samples
				lower	upper	
Log-transformed onset latency						
Population-level effects	Intercept	7.147	0.019	7.11	7.186	5824
	Name Agreement	**-0.125**	**0.012**	**-0.149**	**-0.101**	**63985**
	Speech vs. Quiet	**0.07**	**0.036**	**0**	**0.141**	**71154**
	Word List vs. Sentence	-0.003	0.04	-0.081	0.075	68553
	Experiment	-0.026	0.037	-0.098	0.046	6025
	NA × (S vs. Q)	0.017	0.068	-0.117	0.15	71792
	NA × (WL vs. S)	0.005	0.074	-0.142	0.15	70402
	NA × Experiment	0.005	0.013	-0.021	0.031	70888
	(S vs. Q) × Experiment	-0.013	0.032	-0.076	0.05	74191
	(WL vs. S) × Experiment	0.003	0.029	-0.054	0.06	72758
	NA × (S vs. Q) × Experiment	-0.049	0.056	-0.158	0.059	75539
	NA × (WL vs. S) × Experiment	-0.039	0.054	-0.145	0.067	75976
Group-level effects	*Participant_sd (Intercept)*	0.17	0.014	0.146	0.199	10874
	sd(Name Agreement)	0.027	0.008	0.01	0.041	22835
	sd(Speech vs. Quiet)	0.065	0.01	0.047	0.084	36544
	sd(Word List vs. Sentence)	0.04	0.009	0.021	0.058	20658
	sd(NA × (S vs. Q))	0.025	0.017	0.001	0.064	28855
	sd(NA × (WL vs. S))	0.021	0.016	0.001	0.059	26258
	Item_sd (Intercept)	0.027	0.013	0.002	0.048	1450
	sd(Name Agreement)	0.055	0.026	0.004	0.096	1385
	sd(Speech vs. Quiet)	0.167	0.095	0.007	0.307	1211
	sd(Word List vs. Sentence)	0.192	0.105	0.009	0.344	1842
	sd(Experiment)	0.018	0.011	0.001	0.038	2045
	sd(NA × (S vs. Q))	0.347	0.189	0.016	0.616	1209
	sd(NA × (WL vs. S))	0.381	0.211	0.016	0.687	1817

	sd(NA × Experiment)	0.037	0.021	0.002	0.075	1954
	sd((S vs. Q) × Experiment)	0.124	0.07	0.006	0.23	1526
	sd((WL vs. S) × Experiment)	0.131	0.058	0.012	0.225	3159
	sd(NA × (S vs. Q) × Experiment)	0.25	0.14	0.011	0.461	1548
	sd(NA × (WL vs. S) × Experiment)	0.258	0.117	0.023	0.446	3247
Log-transformed utterance duration						
Population-level effects	Intercept	8.016	0.019	7.979	8.053	4034
	Name Agreement	**-0.204**	**0.019**	**-0.24**	**-0.166**	**31359**
	Speech vs. Quiet	0.039	0.027	-0.014	0.093	38700
	Word List vs. Sentence	0.001	0.022	-0.043	0.045	37920
	Experiment	0.01	0.033	-0.054	0.075	3561
	NA × (S vs. Q)	0.045	0.053	-0.06	0.149	39293
	NA × (WL vs. S)	-0.001	0.044	-0.087	0.085	38949
	NA × Experiment	0.024	0.018	-0.011	0.059	21478
	(S vs. Q) × Experiment	*-0.02*	*0.015*	*-0.05*	*0.009*	*62382*
	(WL vs. S) × Experiment	-0.007	0.013	-0.032	0.017	69948
	NA × (S vs. Q) × Experiment	**-0.055**	**0.027**	**-0.109**	**-0.001**	**69610**
	NA × (WL vs. S) × Experiment	0.012	0.026	-0.038	0.062	65325
Group-level effects	*Participant_sd (Intercept)*	0.153	0.012	0.131	0.179	7187
	sd(Name Agreement)	0.067	0.007	0.054	0.081	28946
	sd(Speech vs. Quiet)	0.026	0.011	0.003	0.046	11714
	sd(Word List vs. Sentence)	0.008	0.005	0	0.019	33445
	sd(NA × (S vs. Q))	0.02	0.014	0.001	0.054	24533
	sd(NA × (WL vs. S))	0.023	0.014	0.001	0.053	22589
	Item_sd (Intercept)	0.041	0.022	0.002	0.074	1562
	sd(Name Agreement)	0.083	0.044	0.004	0.147	1599
	sd(Speech vs. Quiet)	0.139	0.054	0.023	0.225	2527
	sd(Word List vs. Sentence)	0.112	0.044	0.018	0.182	2874
	sd(Experiment)	0.018	0.009	0.001	0.035	7237

	sd(NA × (S vs. Q))	0.273	0.108	0.041	0.447	2380
	sd(NA × (WL vs. S))	0.226	0.087	0.039	0.365	2790
	sd(NA × Experiment)	0.035	0.019	0.002	0.07	7414
	sd((S vs. Q) × Experiment)	0.041	0.023	0.002	0.084	6087
	sd((WL vs. S) × Experiment)	0.04	0.021	0.002	0.08	5466
	sd(NA × (S vs. Q) × Experiment)	0.08	0.046	0.004	0.169	5992
	sd(NA × (WL vs. S) × Experiment)	0.081	0.043	0.005	0.16	5395
Log-transformed total pause time						
Population-level effects	Intercept	6.284	0.062	6.163	6.405	4174
	Name Agreement	**-0.589**	**0.055**	**-0.697**	**-0.481**	**26776**
	Speech vs. Quiet	0.031	0.072	-0.111	0.174	37500
	Word List vs. Sentence	0.037	0.06	-0.083	0.155	37909
	Experiment	-0.03	0.113	-0.252	0.19	3829
	NA × (S vs. Q)	0.163	0.142	-0.119	0.443	35595
	NA × (WL vs. S)	0.017	0.121	-0.219	0.255	37295
	NA × Experiment	0.026	0.064	-0.099	0.152	18480
	(S vs. Q) × Experiment	-0.045	0.059	-0.162	0.071	51571
	(WL vs. S) × Experiment	-0.05	0.052	-0.152	0.052	62542
	NA × (S vs. Q) × Experiment	**-0.234**	**0.112**	**-0.455**	**-0.012**	**63364**
	NA × (WL vs. S) × Experiment	0.037	0.106	-0.17	0.246	59726
Group-level effects	*Participant_sd (Intercept)*	0.514	0.041	0.441	0.603	7707
	sd(Name Agreement)	0.227	0.026	0.18	0.281	29906
	sd(Speech vs. Quiet)	0.101	0.041	0.016	0.177	13912
	sd(Word List vs. Sentence)	0.031	0.023	0.001	0.085	28697
	sd(NA × (S vs. Q))	0.112	0.073	0.005	0.27	16436
	sd(NA × (WL vs. S))	0.11	0.062	0.007	0.239	18382
	Item_sd (Intercept)	0.118	0.06	0.006	0.205	1575
	sd(Name Agreement)	0.217	0.123	0.01	0.406	1524
	sd(Speech vs. Quiet)	0.348	0.141	0.052	0.576	2218

	sd(Word List vs. Sentence)	0.289	0.124	0.031	0.487	2346
	sd(Experiment)	0.058	0.034	0.003	0.125	8725
	sd(NA × (S vs. Q))	0.678	0.283	0.09	1.14	2238
	sd(NA × (WL vs. S))	0.575	0.248	0.067	0.97	2335
	sd(NA × Experiment)	0.117	0.069	0.006	0.253	8968
	sd((S vs. Q) × Experiment)	0.153	0.085	0.009	0.318	6683
	sd((WL vs. S) × Experiment)	0.16	0.089	0.009	0.328	6183
	sd(NA × (S vs. Q) × Experiment)	0.292	0.17	0.015	0.628	6590
	sd(NA × (WL vs. S) × Experiment)	0.322	0.178	0.018	0.656	6527
Log-transformed articulation time						
Population-level effects	Intercept	7.757	0.015	7.727	7.786	4999
	Name Agreement	**-0.089**	**0.019**	**-0.127**	**-0.052**	**37001**
	Speech vs. Quiet	**0.046**	**0.014**	**0.018**	**0.074**	**49698**
	Word List vs. Sentence	-0.005	0.012	-0.029	0.019	45323
	Experiment	0.025	0.024	-0.021	0.073	3748
	NA × (S vs. Q)	0.01	0.028	-0.045	0.064	48524
	NA × (WL vs. S)	-0.002	0.024	-0.049	0.046	47017
	NA × Experiment	0.008	0.013	-0.017	0.033	18403
	(S vs. Q) × Experiment	-0.016	0.009	-0.034	0.003	52214
	(WL vs. S) × Experiment	-0.004	0.006	-0.016	0.008	72990
	NA × (S vs. Q) × Experiment	-0.002	0.014	-0.03	0.026	89838
	NA × (WL vs. S) × Experiment	-0.004	0.013	-0.028	0.021	88482
Group-level effects	*Participant_sd (Intercept)*	0.11	0.009	0.095	0.13	8141
	sd(Name Agreement)	0.053	0.005	0.044	0.063	21399
	sd(Speech vs. Quiet)	0.03	0.005	0.021	0.041	29762
	sd(Word List vs. Sentence)	0.007	0.004	0	0.015	26055
	sd(NA × (S vs. Q))	0.018	0.011	0.001	0.042	16427
	sd(NA × (WL vs. S))	0.014	0.01	0.001	0.036	16253
	Item_sd (Intercept)	0.043	0.024	0.002	0.077	1422

sd(Name Agreement)	0.086	0.048	0.004	0.154	1456
sd(Speech vs. Quiet)	0.064	0.036	0.003	0.117	1607
sd(Word List vs. Sentence)	0.056	0.03	0.003	0.102	1895
sd(Experiment)	0.008	0.005	0	0.017	12710
sd(NA × (S vs. Q))	0.13	0.073	0.006	0.235	1537
sd(NA × (WL vs. S))	0.116	0.061	0.006	0.205	1857
sd(NA × Experiment)	0.016	0.009	0.001	0.034	14920
sd((S vs. Q) × Experiment)	0.01	0.007	0	0.028	33328
sd((WL vs. S) × Experiment)	0.01	0.007	0	0.027	25544
sd(NA × (S vs. Q) × Experiment)	0.02	0.015	0.001	0.056	30810
sd(NA × (WL vs. S) × Experiment)	0.02	0.014	0.001	0.054	26730

Note. NA refers to name agreement, WL refers to word lists, S refers to sentences, and Exp refers to Experiment.

5 | Auditory disruption by irrelevant background sentences on spoken language production

Abstract

Everyday conversations often occur in noisy settings such as on a train or in a busy cafeteria. There speakers are exposed to external auditory stimulation from nonverbal and verbal noise that may distract them, impairing speech production. The present study explored how the interestingness (funny versus boring) and contextual variation (varied versus constant) of background sentences influenced speech production, and whether the influence was modulated by the difficulty of speech production (indexed by name agreement: high versus low). Native Dutch speakers named sets of pictures with high or low name agreement while ignoring background sentences in a constant context (only boring sentences) or a varied context (boring and funny sentences intermixed). In the varied context, funny sentences caused less disruption on picture naming performance than boring sentences. Boring sentences elicited less interference in the varied context than in the constant context. Moreover, the effects of interestingness and context were larger for low than high name agreement pictures. These findings suggest that both the interestingness and contextual variation of background sentences influence speech production, and that this influence is modulated by the lexical selection demand of speech production. This implies that speakers may increase top-down cognitive control to shield against auditory disruption when background speech is funny and varied enough.

5.1 Introduction

Although it is perhaps best to have conversations when no background distracting stimuli are present, such ideal settings are not typical for daily life. Rather, much of daily conversation occurs in the presence of external auditory stimulation, such as noise from nearby traffic or construction, television broadcasting in the background, or a colleague talking on the phone. It has been shown previously that the two aspects of conversation -- spoken language comprehension (e.g., Eckert et al., 2016; Vasilev et al., 2018) and production (e.g., He et al., 2021; Chapter 4 of this dissertation) -- receive interference from irrelevant background noise. However, compared with extensive work on speech comprehension (e.g., Eckert et al., 2016; Vasilev et al., 2018), few studies have explored how speakers plan their speech in the presence of irrelevant background speech (e.g., word lists or sentences, He et al., 2021; Chapter 4 of this dissertation). Further, a direct comparison between auditory disruption by different types of background sentences (e.g., boring versus funny sentences) in different contexts (e.g., varied versus constant context) that are more naturalistic and resemble typical real-world settings on speech production remains necessary. The comparison is important because different types of background sentences vary in linguistic richness and acoustic pattern, which may have different disruptive potential on speech production and which may cause speakers to take different strategies to shield against the disruptions. The present study thus investigated how the interestingness (funny versus boring) and contextual variation (varied versus constant) of irrelevant background sentences affect speech production in the face of varied lexical selection demands (indexed by name agreement; high versus low). This provides insights into how speakers plan their speech and shield against auditory disruption in real-world conversation.

Before we examine the effect of auditory disruption on speech production, we review the literature about the *irrelevant speech effect* (or *irrelevant sound effect*; e.g., Colle & Welsh, 1976; Jones et al., 1992) which is foundational to the present study. The irrelevant speech effect refers to the impairment of performance when a task (e.g., a short-term memory task or a reading task) is performed in the presence of irrelevant background stimuli (e.g., pure tones, syllables, words; Hughes et al., 2007; Hyönä & Ekholm, 2016; Vachon et al., 2017) relative to a quiet condition. Two major theoretical views have been proposed to account for the irrelevant speech effect. The first is the *interference-by-similarity* account (also referred as *interference-by-process* account), which assumes that the auditory disruption is caused by a conflict between similar and competing representations or processes that are used for background stimuli and focal tasks (e.g., Jones et al., 1993; Macken et al., 2009). For instance, in a typical serial recall task, a sequence of items (usually 6 – 8 digits or letters) is to be maintained and recalled in the correct order in the presence of irrelevant background speech (Jones & Morris, 1992). The interference-by-similarity account suggests that changing-state speech consisting of different distractor items (e.g., A C B E D H G F) impairs serial recall performance because it interferes with the processing of serial order in the main task, and steady-state speech comprising of a repeated distractor (e.g., A A A A A A A A) does not interfere because serial order processing is unnecessary (i.e., *changing-state effect*; Jones et al., 1993; Jones et al., 1992). Further, the interference-by-similarity account has been extended to explain auditory disruption in reading performance. It has been argued that auditory distraction in reading could be attributed to similarity in shared use of semantic processing (i.e., *semantic disruption view*, Martin et al., 1988) or similarity in the phonological representations (i.e., *phonological disruption view;* Salamé & Baddeley, 1982, 1989) between background speech and reading, and both views have received some support (for a review, see Vasilev et al., 2018).

The other major theoretical account of the irrelevant speech effect is the *attention capture* account, which assumes that irrelevant background speech disrupts focal task performance by diverting attention away from the task (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015). For instance, in serial recall, changing-state speech produces more disruption because its acoustic changes cause attentional capture, whereas steady-state speech is more predictable and does not capture attention (Chein & Fiez, 2010; Cowan, 1995; Elliott & Briganti, 2012;). The deviant effect in serial recall can also be explained by the attention capture account (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015): more distraction occurs from unexpected speech that deviates from the auditory environment (e.g., A A A A B A A A) relative to a steady-state sequence (e.g., A A A A A A A A) because the deviant item captures attention and diverts attention away from the serial recall task (Cowan 1995). The attention capture account can be divided into two versions (Eimer et al., 1996): *aspecific* attentional capture occurs when background speech captures attention because of the context in which it occurs (due to perceptual properties such as changes in frequency), while *specific* attentional capture occurs when the particular content of background speech diverts attention (due to e.g., semantic information; Röer et al., 2013; Wood & Cowan, 1995).

These accounts of the irrelevant speech effect have been applied to explain how background speech disrupts speech production. For instance, a recent study by He et al. (2021)

explored the role of representational similarity in auditory disruption by background speech on speech production. In this study Dutch speakers named sets of pictures in Dutch while ignoring background speech in Dutch, or in Chinese which they could not understand, or language-like noise (eight-talker babble). Background speech disrupted speech production more than babble, and Dutch caused more interference than the Chinese. The results suggest that more disruption of speech production occurs as the representational similarity between background speech and the speech production task increases, which is consistent with the interference-by-similarity account (e.g., Jones et al., 1993; Martin et al., 1988; Salamé & Baddeley, 1982, 1989). However, this study did not distinguish whether the disruption by background speech was phonological or semantic in nature, and also did not rule out the possibility that the disruption was caused via attentional capture.

To overcome these limitations, a further investigation by He and colleagues (Chapter 4) was conducted in which Dutch speakers named sets of pictures while ignoring word lists or sentences in Chinese (Experiment 1) and Dutch (Experiment 2), or in a quiet control condition. The comparison between speech and quiet conditions reveals whether the disruption is phonological or semantic in nature: the phonological disruption view (Salamé & Baddeley, 1982, 1989) predicts that the presence of background speech (regardless of its intelligibility) should disrupt speech production relative to a quiet condition, while the semantic disruption view (Martin et al., 1988) predicts that only intelligible background speech (i.e., Dutch) should interfere with speech production. He et al. (Chapter 4) found that both Chinese and Dutch background speech disrupted speech production relative to a quiet condition, which suggests that at least part of the disruption is phonological in nature (Salamé & Baddeley, 1982, 1989). He et al. (Chapter 4) also compared word lists and sentences (especially in Chinese) to distinguish between the two variants of attention capture (Eimer et al., 1996): the aspecific attention capture view predicts that Chinese word lists should elicit more interference than Chinese sentences because of the presence of more stimulus-aspecific pauses, while specific attention capture predicts that Chinese word lists should have the same disruption potency as the Chinese sentences because they are equally meaningless to Dutch speakers. No difference between Chinese word lists and sentences was obtained, suggesting that stimulus-aspecific variation may not capture attention, which in turn supports the specific attention capture view (Eimer et al., 1996). In addition, there was no difference between word lists and sentences in Dutch experiment, which implies that the two types of Dutch background speech interfered to similar degrees with speech production, regardless of aspecific contextual variation and specific syntactic complexity. These results suggest that speakers were, in general, quite good at blocking off auditory information when they plan their own speech and received relatively little interference across the board.

However, the background speech used in He et al. (Chapter 4 of this dissertation) was notably uniform and fairly boring: word lists had a regular acoustic pattern (with silent pauses of 700 ms between consecutive words, e.g., "*fisherman, choir, football, apple, ruler, deer*"), and sentences had uniform syntactic structure (e.g., *"The deer and the ruler are to the left of the apple, and the football and the choir are to the right of the fisherman.").* This suggests that the background speech had relatively impoverished attentional capture properties, which may underestimate the potential of different types of background speech for disruption. To put this possibility to an empirical test, the present study explored whether a larger and more ecologically relevant manipulation, i.e., the relative interestingness (funny versus boring), of background sentences would moderate the disruptive effect on speech production performance. The boring sentences were the same as those used in Chapter 4. The funny sentences described events and featured light semantic anomalies (e.g., *"The chair danced along with the lamp in the living room while the rest of the house was asleep.").*

The comparison between funny and boring sentences also allowed us to test whether the attentional capture account for the irrelevant speech effect on speech production is specific or aspecific in nature. As mentioned earlier, the specific attention capture view (Eimer et al., 1996) assumes that the linguistic content of background speech can divert attention and cause more disruption on focal language tasks. This view thus predicts that relative to boring sentences, funny sentences should alert the attentional system to allocate further processing resources to the auditory modality, resulting in a decrease in speech production performance. In contrast, the aspecific attention capture view assumes that the perceptual properties of background speech should capture attention more or less effectively, which predicts that attention capture may be independent of the interestingness of background sentences: funny sentences should not elicit more interference than boring sentences on speech production performance.

The fairly uniform background speech used in He et al. (Chapter 4) could also make speakers adapt quite easily, leading to little auditory disruption on speech production and the lack of word list versus sentence effects in either experiment. The ease of adaptation might be akin to the finding that steady-state sequences interfere less with serial recall performance than changing-state sequences, because repeated exposure to the same auditory distractor causes habituation of attentional orienting (Cowan, 1995). More specifically, the *habituation* view postulates that the irrelevant speech effect is caused by orienting responses to the auditory distractors, which are assumed to be subject to habituation (Cowan, 1995, 1999). When a stimulus remains constant or is repeated, the cognitive architecture is thought to adapt, as a consequence of which attention orienting decreases. Habituation serves to ensure that background stimuli that are irrelevant to the focal task consume only limited attentional resources.

To test whether habitation to background stimuli occurs in this paradigm, the present study made a larger manipulation of the contextual variation by presenting background sentences in two blocks: the constant context block contained a set of boring sentences, while the varied context block contained both boring sentences and funny sentences. The habituation view predicts an attenuation of interference by boring sentences in the constant context compared to the varied context, as the constant context leads to habituation (Cowan, 1995, 1999). By contrast, the interference-by-similarity account (Jones et al., 1993; Martin et al., 1988; Salamé & Baddeley, 1982, 1989) predicts that the interference effect will not be attenuated with repeated exposure to the boring sentences in the constant context because representational similarity does not change with exposure.

The previous paragraphs laid out the role of “bottom-up” properties of background stimuli in irrelevant speech effects, but “top-down” cognitive control that is internal to speakers could also matter. Earlier studies have shown that an increase in focal task difficulty can shield against distraction from background stimuli via a top-down cognitive control (attention engagement) mechanism (Halin et al., 2014; Marsh et al., 2015). The attention engagement account assumes that when the focal task is difficult, the metacognitive system triggers a compensatory shift in attention engagement (or concentration) such that individuals can maintain a desired performance level by reducing the processing of background information (Ball et al., 2018; Marsh et al., 2015).

Earlier work has shown evidence for top-down attention engagement when picture naming in the presence of background speech was made especially difficult with low name agreement of to-be named pictures (e.g., He et al., 2021, Chapter 4). Name agreement (hereafter NA) is the extent to which participants agree on the name of a picture. Previous studies have found that naming a picture with high name agreement (e.g., the item called *banana*) is faster and more accurate than naming a picture with low name agreement (e.g., the item called *sofa* or *couch*), which is referred to as the name agreement effect (Alario et al., 2004; Vitkovitch &

Tyrell, 1995; Shao et al., 2014). The effect can arise at two levels of speech production: object recognition (due to confusion of what the object should be called) and lexical selection (due to the need to select among competing lexical candidates). He and colleagues (2021, Chapter 4) focused on the latter effect and found that disruption by background speech occurred for only easy picture naming (high name agreement pictures), and not for difficult picture naming (low name agreement pictures). This finding suggests that auditory disruption by background speech can be attenuated or eliminated when the lexical selection demand of speech production increases. To explore whether the disruption elicited by the interestingness and contextual variation of background sentences is further modulated by speech production difficulty, the present study also manipulated the name agreement (high versus low) of to-be-named pictures. This provides insight into whether and how speakers take top-down strategies to shield against auditory disruption when planning their speech.

In sum, the current study examined the irrelevant speech effect in light of several proposed mechanisms including the attention capture account (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015), the habituation account (Cowan, 1995, 1999), and the attention engagement account (Halin et al., 2014; Marsh et al., 2015). The aim was to shed light on how the interestingness (funny versus boring) and contextual variation (varied versus constant) of background sentences affected speech production with varied lexical selection demands (indexed by name agreement: high, low). Following earlier work (e.g., Alario et al., 2004; Vitkovitch & Tyrell, 1995; Shao et al., 2014), we expected that pictures with low name agreement would be named more slowly and elicit more errors than those with high name agreement. Under a specific attention capture view (Eimer et al., 1996), we predicted an interestingness effect, such that in the varied context, funny sentences should cause more disruption than boring sentences because funny sentences contain rich semantic and phonological information and complex syntactic and thematic structure. According to the habituation view (Cowan, 1995), we predicted that there should be a context effect, such that boring sentences in the constant context should elicit less interference than those in the varied context because of habituation to the boring sentences in the constant context. Finally, following the attention engagement account (Halin et al., 2014; Marsh et al., 2015), we predicted that both the interestingness effect and the context effect should be reduced for low name agreement pictures compared to high name agreement pictures.

5.2 Methods

Participants

We recruited 59 native speakers of Dutch (43 females, M_{age} = 24 years, range: 18-33 years) from the Max Planck Institute for Psycholinguistics' participant database. A-priori power simulations showed that testing 50 participants and 90 items (i.e., with 80% of the items in the study named successfully) would allow 99% power to measure a plausibly-sized interaction between name agreement and the interestingness effect on the measurement of utterance duration. The interaction effect size used in the simulation was an interestingness effect (i.e., funny sentence > boring sentence in the varied context) of 30 ms or smaller (SD = 900 ms) for low name agreement pictures, but 60 ms or larger (SD = 900 ms) for high name agreement pictures. All participants reported normal or corrected-to-normal vision and no speech or hearing problems. They signed an online informed consent form and received a payment of €6 for their participation. The study was approved by the ethics board of the Faculty of Social Sciences of Radboud University.

Apparatus

The experiment was implemented in FRINEX (FRamework for INteractive EXperiments; for details, see Withers, 2017), a web-based platform developed by the technical group at the Max Planck Institute for Psycholinguistics. Participants used their own laptops with headphones / earphones. We restricted participants to using 14-inch or larger laptops (range: 14-24 inches) with Google Chrome, Firefox, Microsoft Edge, Internet Explorer, Brave, or Opera web browsers. Each participant's speech was recorded by a built-in voice recorder of the web browser. WebMAUS Basic was used for phonetic segmentation and transcription (https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic).
Praat (Boersma & Weenink, 2009) was then used to extract the onsets and offsets of all segmented responses.

Materials

Visual stimuli. A subset of the pictures (224 of the original 240 pictures) from He et al. (2021, Experiment 2; pictures selected from MultiPic database, Duñabeitia et al., 2018; see Appendix A, Table A1) was used in the present study. Of these, 112 were high name agreement pictures, all with a name agreement 100%, and 112 were low name agreement pictures, with a name agreement between 50% and 87% (M = 72%, SD = 11%). Independent *t*-tests revealed that the two sets of pictures differed significantly in name agreement, but not in any of the following psycholinguistic attributes: visual complexity, word frequency, age-of-acquisition, number of phonemes, number of syllables, word prevalence, phonological neighborhood frequency, phonological neighborhood size, orthographic neighborhood frequency, and orthographic neighborhood size.

The high and low name agreement pictures were each divided into two subsets and paired with the boring sentence conditions in two blocks, meaning that the boring sentence condition in each block was paired with 56 high name agreement and 56 low name agreement pictures. The two subsets of pictures were matched on the 10 above-mentioned attributes, as were the high and low name agreement sets of pictures assigned to each block of boring sentences. All of the pictures (224 pictures) were also paired with the funny sentence condition in the varied block. The pictures in the boring (112 pictures) and funny sentence conditions (224 pictures) in the varied block were also matched on the 10 above-mentioned attributes.

On each trial of the experiment, four pictures matched in name agreement were presented simultaneously in a 1 × 4 grid (size: 10 cm × 40 cm). The pictures in each grid were neither semantically related (i.e., they were from different semantic categories) nor phonological related (i.e., avoiding the overlap of their first phonemes), as judged by a native speaker of Dutch. There were 28 picture grids in the constant block and 84 picture grids (28 for boring sentence condition, 56 for funny sentence condition) in the varied block. Thirty-two additional pictures (8 picture grids) were selected from the same database as practice stimuli.

Irrelevant background sentences. We created 56 boring sentences to pair with the 56 picture grids, each containing six nouns and a simple syntactic structure (e.g., De dierenarts en het terras bevinden zich links van het hotel, en het vuilnis en de rem bevinden zich rechts van de fluit. *'The vet and the terrace are to the left of the hotel, and the garbage and the brake are to the right of the whistle.'*). These were comprised of the 20 sentences from our previous study (Experiment 2 in He et al., Chapter 4), and 36 additional sentences made from 216 Dutch nouns (see Appendix A, Table A2) selected from Experiment 2 in He et al., (2021) by adding a conjunction (and) and prepositional phrases (to the left/right of) to link the nouns. All 56 boring sentences were matched on word frequency and number of syllables. To pair with the 28 picture grids in each block, these 56 boring sentences were divided into two subsets, matched on word frequency, number of syllables, number of phonemes, word prevalence, and age-of-acquisition. To avoid phonological and semantic overlap between picture naming and boring sentences, we designed the boring sentences so that no nouns in a sentence were semantically or phonologically related, and no three consecutive nouns per sentence were semantically or phonologically related to the to-be-named pictures in the same ordinal position. To create practice stimuli, 24 additional nouns were selected from He et al. (2021) and then transformed into four additional boring sentences. All of the boring sentences were recorded by a female native Dutch speaker in neutral prosody using Audacity software (https://www.audacityteam.org/) at a sample rate of 44100 Hz. Each boring sentence was then further processed using Adobe Audition (https://www.adobe.com/products/audition.html/) and Praat to delete initial and final silences and compressed by up to 0.19%, so that each boring sentence lasted 8 seconds.

The 56 funny sentences (see Appendix A, Table A3; e.g., De stoel danste samen met de lamp in de woonkamer terwijl de rest van het huis aan het slapen was. *'The chair danced along with the lamp in the living room while the rest of the house was asleep.'*), were written by trained Dutch-speaking student assistants with the following instructions: the critical word that marks the implausibility should appear near the beginning of the sentence (in the 2nd, 3rd, or 4th position); the subject should be a concrete noun such as dog, grape etc., but not a person's name; the length of the sentence should be between 15 and 20 words; the sentence should not contain negative words such as death, violence etc.; the nouns within the sentence should not be semantically related; and different nouns should appear across sentences. All sentences were double-checked by a different native Dutch speaker. To avoid semantic and phonological overlap, the critical noun in each sentence was not semantically or phonologically related to the to-be-named pictures. In addition, four funny sentences were written by student assistants as practice stimuli. The same speaker recorded all funny sentences in neutral prosody. They were further edited in the same fashion as the boring sentences (by stretching by up to 0.33% and compressed by up to 0.1%) to last 8 seconds.

To check participants' concentration level and whether they were able to hear the background sentences, 15 additional two-syllable Dutch nouns (3 for the practice block, 3 for the constant block, and 9 for the varied block) were selected from the same database (Duñabeitia et al., 2018) as attention check stimuli that needed to be repeated back during the experiment. These nouns were recorded by a native Dutch speaker in neutral prosody. All auditory files were matched on intensity (total RMS [root mean square] = -33.98dB) in Adobe Audition (https://www.adobe.com/products/audition.html/).

Design

The sentence block (constant block with boring sentences, varied block including boring and funny sentences) and the difficulty of lexical selection in speech production (name agreement: high versus low) were treated as within-participant variables. Name agreement was randomized within experimental blocks and counterbalanced across participants. Picture stimuli were repeated twice resulting in two blocks. The constant block contained 28 trials with one repetition of half of the boring sentences (e.g., boring sentences 1-28) and half of the picture grids (e.g., picture grid 1-28). The varied block contained 84 trials with one repetition of the other half of the boring sentences (e.g., boring sentences 29-56) and the other half of picture grids (e.g., picture grid 29-56), and one repetition of all funny sentences (i.e., funny sentences 1-56) and all picture grids (i.e., picture grid 1-56). The constant block always preceded the varied block in order to prevent a response strategy where participants may expect and attend for amusing sentences even in the constant block. A unique order of stimulus presentation was created for each participant using the Mix program (van Casteren & Davis, 2006), with the constraints that attention check trials were presented at least every three trials.

Procedure

Participants were tested online (e.g., for one participant, https://frinexstaging.mpi.nl/image_n-aming_noise_b/?stimulusList=List1) and received instructions that they should perform this experiment in a quiet room with the door shut and with potentially distracting electronic equipment turned off. They were asked to image that they were in a laboratory during the experiment, to wear headphones properly, and to set the volume of their laptop to the level that they usually use (e.g., to watch a video) and not change it during the experiment. We asked for permission to record their vocal responses and asked them to report their volume values before the test began.

During the experiment, a practice session of 11 trials (8 test trials and 3 attentional check trials) was followed by the two blocks of experimental trials (one block of 28 test trials and 3 attention check trials, and one block of 84 test trials and 9 attention check trials). After completing the main portion of the experiment, participants were asked to type the value of their volume again, which allowed us to check whether they changed the computer volume during the experiment. The experiment lasted about 30 minutes.

Practice and experimental trials began with a fixation cross presented for 500 ms, followed by a blank screen for 300 ms. Then, a 1 × 4 grid appeared on the screen in which four pictures were presented simultaneously while a sound file played for up to 8 seconds. Participants named the four pictures one by one from left to right as quickly and accurately as possible while ignoring the background sentences. Once finished, they clicked the mouse to end the trial, at which point a blank screen was presented for 1500 ms. An example of a test trial is shown in Figure 5.1. The attentional check trials shared the same structure with the test trials, but the stimulus screen was blank and an audio file of a single Dutch word was played. On these trials, participants were asked to repeat the Dutch word as quickly and accurately as possible.

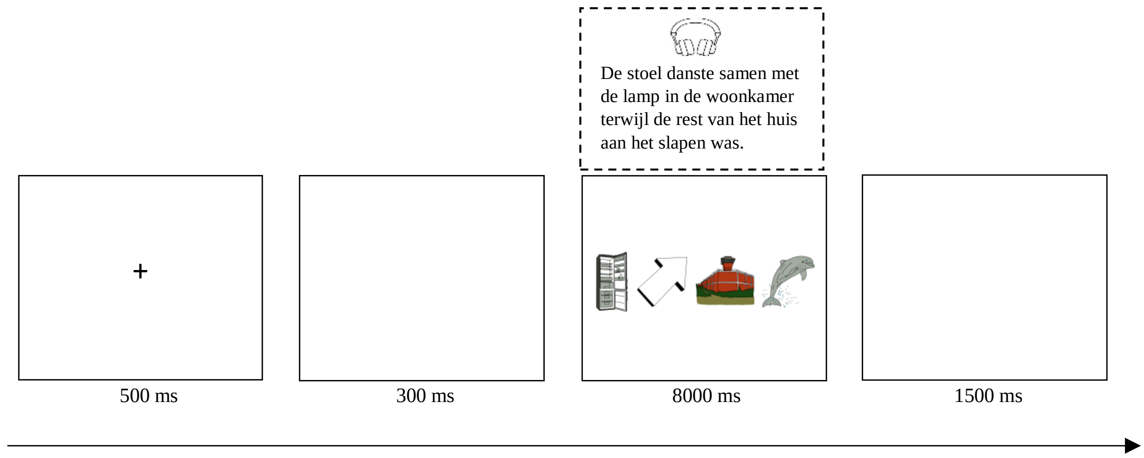

Figure 5.1. An example trial in which participants named pictures with high name agreement while ignoring a funny sentence (translation: The chair danced along with the lamp in the living room while the rest of the house was asleep.).

Analysis

Seven dependent variables were coded to index naming performance. Production *accuracy* reflects the proportion of trials where all four pictures were named correctly. Picture names were coded as correct if they matched any of the multiple names given to the picture in the MultiPic database (Duñabeitia et al., 2018), were diminutive versions of the multiple names (e.g., *munt* ‘coin’ named as *muntje* ‘little coin’), or were judged reasonable by trained research assistants (e.g., *kruk* ‘stool’ named as *stoel* ‘chair’).

For trials where all pictures were named correctly and without hesitations or self-corrections (hereafter, “fully correct trials”), we calculated four main time-based measures: onset latency, utterance duration, total pause time, and articulation time. *Onset latency* was defined as the interval from the onset of stimulus presentation to the onset of the utterance and indexes the beginning stages of speech planning. *Utterance duration* was defined as the interval between the utterance onset of the first picture name and the utterance offset of the fourth picture name and reflects how long participants took to produce all four picture names. *Total pause time* was defined as the sum of all pauses between object names and indexes the planning done between producing responses. *Articulation time* was defined as the sum of the articulation durations of all four picture names and reflects planning during articulations.

For fully correct trials, we also examined how participants grouped their four responses. Since earlier studies of spontaneous speech coded silent durations longer than 200 ms as silent pauses (e.g., Heldner & Edlund, 2010), we coded responses with 200 ms or less between them as a single response chunk. Two measures were derived: total chunk number and first chunk length. *Total chunk number* refers to how many response chunks participants made on one trial with a larger number meaning more separate planning units for production. *First chunk length* refers to how many names participants produced in their initial response and provides a measure of how much information participants planned before starting to speak.

To examine the likely magnitude of all effects, Bayesian mixed-effect models (Nicenboim & Vasishth, 2016) were conducted in R version 4.0.3 (R Core Team, 2020) with the package *brms* (version 2.14.4, Bürkner, 2017). Predictors were name agreement (high / low) and sentence block (constant block with boring sentences / varied block with funny sentences / varied block with boring sentences). Name agreement (high / low) was contrast coded with (0.5, -0.5). Two contrasts were made for the sentence blocks: the first was coded with (0, 0.33, -0.66) to compare the funny sentences and boring sentences in the varied block (indexing the interestingness effect), and the second was coded with (0.5, 0, -0.5) to compare the boring sentences in the constant block and the varied block (indexing the context effect). Note that this contrast scheme uses a weighted mean, which is appropriate for conditions with different numbers of observations (Sweeney & Ulveling, 1972). The random effect structure for the models included random intercepts for participants and items and random slopes for name agreement and the sentence block by participants and items. Separate models were fitted for each dependent measure. All models had four chains and each chain had 24000 iterations depending on model convergence (listed in model output tables). We used a warm-up (or burn-in) period of 2000 iterations in each chain, which means we removed the data based on the first 2000 iterations in order to correct the initial sampling bias.

All models used weak, widely spread priors that would be consistent with a range of null to moderate effects. The model of accuracy used family *bernoulli* combined with a *logit* link, with a student-t prior with 1 degree of freedom and a scale parameter of 2.5. The models of log-transformed onset latency, log-transformed utterance duration, and log-transformed articulation time used a weak normal prior with an SD of 0.2, and the model of log-transformed total pause time used a weak normal prior with an SD of 1. These models were performed using the family *gaussian* combined with *identity* link. Total chunk number and first chunk length had weak normal priors centered at zero with an SD of 1, and used family *possion* combined with the *log* link. All models were run until the R hat value for each estimated parameter was 1.00, indicating full convergence. For these models, the size of reported betas reflects estimated effect sizes, with larger absolute values of betas reflecting larger effects. We reported the parameters for which 95% Credible Intervals (hereafter, Cr.I) do not contain zero, which is analogous to the frequentist null hypothesis significance test: the parameter has a non-zero effect with high certainty.

5.3 Results

Nine participants were removed from further analyses: four did not run the experiments successfully due to a bad internet connection, four failed to record speech responses, and one gave no responses on attention check trials. The data from the remaining 50 participants was checked for errors, removing from analysis any trials with implausible names (e.g., *koekje* 'cookie' named as *virus*), hesitations (e.g., *komkommer* 'cucumber' named as *kom... komkommer*), self-corrections (e.g., *komkommer* 'cucumber' misnamed as *courgette...komkommer* 'courgette...cucumber'), and any trials where objects were omitted or named in the wrong order. The exclusion of these inaccurate trials resulted in a loss of 9.64% of the data (range by participants: 0 – 41.07% of removed trials). Then, any onset latencies below 200 ms were removed from this analysis, resulting in a loss of 1.30% of the data. Any total pause times below 20 ms were also removed from this analysis, resulting in a loss of 9.70% of the data. Finally, any data points more than 2.5 standard deviations below or above the mean values were removed for each time measure (1.44% for log-transformed onset latency, 0.81% for log-transformed utterance duration, 0.85% for log-transformed total pause time, and 0.77% for log-transformed articulation time). Descriptive statistics of all dependent variables are shown in Table 5.1.

Table 5.1. Means and standard deviations of the dependent variables by name agreement and sentence block.

	High name agreement			Low name agreement		
	Boring Constant	Funny Varied	Boring Varied	Boring Constant	Funny Varied	Boring Varied
Accuracy	96%	94%	94%	83%	87%	87%
Onset latency (ms)	1500 (597)	1330 (533)	1348 (522)	1806 (772)	1532 (656)	1613 (671)
Utterance duration (ms)	3291 (840)	2937 (896)	3038 (941)	4162 (1045)	3629 (1123)	3731 (1155)
Total pause time (ms)	821 (663)	750 (683)	776 (713)	1474 (894)	1174 (896)	1239 (921)
Articulation time (ms)	2567 (607)	2334 (494)	2370 (509)	2739 (696)	2559 (592)	2584 (629)
Total chunk number	2.3 (1.1)	2.0 (1.1)	2.1 (1.0)	3.0 (1.0)	2.5 (1.1)	2.6 (1.1)
First chunk length	2.3 (1.3)	2.6 (1.3)	2.5 (1.3)	1.6 (0.9)	2.1 (1.2)	2.0 (1.2)

Note. Standard deviations are given in parentheses. All time and chunking measures reflect fully correct trials only.

Attention Checks. The mean accuracy for attention check responses was 98% (range by participants: 67% - 100%), showing that participants' attention levels were good and that they indeed heard the background speech.

Accuracy. Participants produced sensible responses on 90% of the naming trials. As shown in Table 5.2, a Bayesian mixed-effect model showed that accuracy was considerably lower for low name agreement pictures than high name agreement pictures ($\beta = 1.194$, SE = 0.28, 95% Cr.I = [0.651, 1.757]), but it was not influenced by the type of background sentence. Name agreement did not interact with the type of background sentences.

Onset latency. As shown in Table 5.2 and the top left panel of Figure 5.2, a Bayesian mixed-effect model showed that log-transformed onset latency was affected by name agreement and the type of background sentences: participants were reliably faster to plan names for high name agreement pictures than low name agreement pictures ($\beta = -0.158$, SE = 0.022, 95% Cr.I = [-0.201, -0.115]). The log-transformed onset latency was faster in funny background sentences than boring background sentences in the varied context block (β = -0.133, SE = 0.016, 95% Cr.I = [-0.164, -0.102]), and it was also faster for the boring background sentences in the varied context block than in the constant context block (β = 0.193, SE = 0.022, 95% Cr.I = [0.148, 0.237]).

There was an interaction between name agreement and interestingness (β = 0.07, SE = 0.024, 95% Cr.I = [0.023, 0.117]), where the log-transformed onset latencies in funny background sentences were only slightly faster than in boring background sentences in the varied context block for high name agreement pictures (β = -0.096, SE = 0.02, 95% Cr.I = [-0.136, -0.056]), but were much faster for low name agreement pictures (β = -0.167, SE = 0.019, 95% Cr.I = [-0.205, -0.128]). An interaction was also found between name agreement and context (β = -0.070, SE = 0.030, 95% Cr.I = [-0.128, -0.012]), where the log-transformed onset latencies for the boring background sentences in the varied context block were somewhat faster than in the constant context block for high name agreement pictures (β = 0.154, SE = 0.03, 95% Cr.I = [0.094, 0.214]), and were even faster for low name agreement pictures (β = 0.225, SE = 0.024, 95% Cr.I = [0.179, 0.272]).

Utterance duration. As shown in Table 5.2 and the top right panel of Figure 5.2, a Bayesian mixed-effect model showed that log-transformed utterance duration was affected by name agreement and the type of background sentences: it was significantly shorter for high name agreement pictures than low name agreement pictures (β = -0.222, SE = 0.023, 95% Cr.I = [-0.268, -0.176]). The log-transformed utterance duration was shorter in funny background sentences than boring background sentences in the varied context block (β = -0.131, SE = 0.015, 95% Cr.I = [-0.161, -0.101]), and it was also shorter for the boring background sentences in the varied context block than in the constant context block (β = 0.198, SE = 0.023, 95% Cr.I = [0.154, 0.243]).

There was an interaction between name agreement and context (β = -0.054, SE = 0.025, 95% Cr.I = [-0.102, -0.005]), showing that the log-transformed utterance duration for the boring background sentences in the varied context block was shorter than in the constant context block for high name agreement pictures (β = 0.171, SE = 0.025, 95% Cr.I = [0.120, 0.220]), and this effect was much larger for low name agreement pictures (β = 0.227, SE = 0.026, 95% Cr.I = [0.177, 0.278]).

Total pause time. As shown in Table 5.2 and the bottom left panel of Figure 5.2, a Bayesian mixed-effect model showed that the results for this measurement patterned in the same way as the log-transformed utterance duration. The log-transformed total pause time was affected by name agreement and the type of background sentences: it was considerably shorter for high name agreement pictures than low name agreement pictures (β = -0.633, SE = 0.065, 95% Cr.I = [-0.761, -0.506]). The log-transformed total pause time was shorter in funny background sentences than boring background sentences in the varied context block (β = -0.296, SE = 0.052, 95% Cr.I = [-0.399, -0.194]), and it was also shorter for the boring background sentences in the varied context block than in the constant context block (β = 0.429, SE = 0.074, 95% Cr.I = [0.283, 0.575]).

There was again an interaction between name agreement and context e (β = -0.203, SE = 0.096, 95% Cr.I = [-0.390, -0.014]), showing that the log-transformed total pause time for the boring background sentences in the varied context block was shorter than in the constant context block for high name agreement pictures (β = 0.322, SE = 0.096, 95% Cr.I = [0.133, 0.512]), and this effect was much larger for low name agreement pictures (β = 0.526, SE = 0.08, 95% Cr.I = [0.369, 0.682]).

Articulation time. As shown in Table 5.2 and the bottom right panel of Figure 5.2, a Bayesian mixed-effect model showed that the log-transformed articulation time was affected by name agreement and the type of background sentences: it was significantly shorter for high name agreement pictures than low name agreement pictures (β = -0.089, SE = 0.022, 95% Cr.I = [-0.133, -0.045]). The log-transformed articulation time was shorter in funny background sentences than boring background sentences in the varied context block (β = -0.070, SE = 0.010, 95% Cr.I = [-0.090, -0.051]), and it was also shorter for the boring background sentences in the varied context block than in the constant context block (β = 0.120, SE = 0.016, 95% Cr.I = [0.089, 0.152]). Unlike as seen for the other time measures, name agreement did not interact with the type of background sentences.

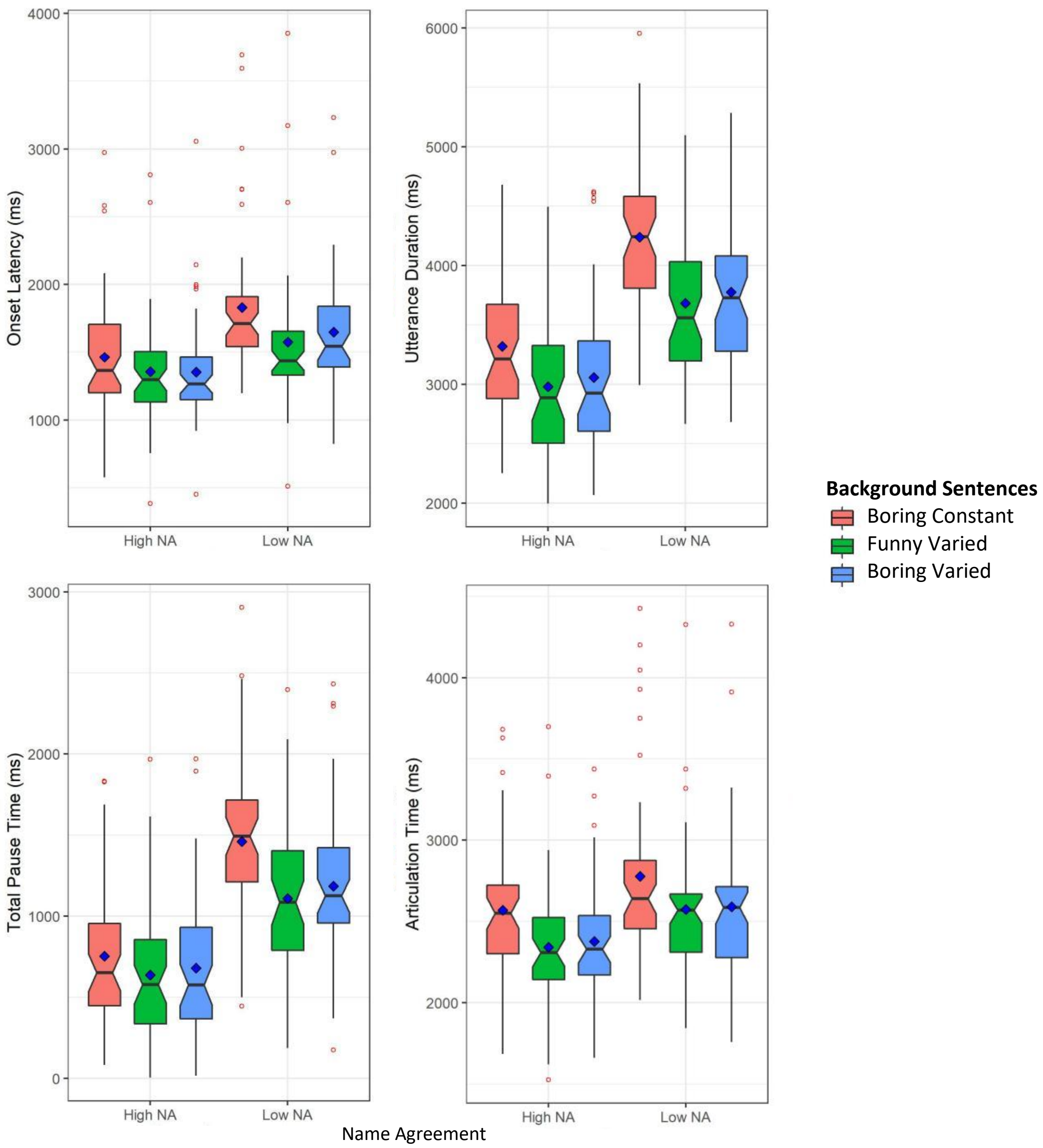

Figure 5.2. Onset latency (top left), utterance duration (top right), total pause time (bottom left), and articulation time (bottom right) split by name agreement (NA: high, low) and the type of background sentences (Boring sentences in the constant context block, funny sentences in the varied context block, boring sentences in the varied context block). Blue squares represent condition means and red points reflect outliers averaged by participants. NA refers to name agreement.

Total chunk number. As shown in Table 5.2 and the left panel of Figure 5.3, a Bayesian mixed-effect model showed that participants grouped their responses in fewer chunks for high name agreement pictures than low name agreement pictures (β = -0.261, SE = 0.03, 95% Cr.I = [-0.321, -0.202]). Total chunk number was also impacted by the type of background sentences: participants grouped their responses in fewer chunks in funny background sentences than boring background sentences in the varied context block (β = -0.134, SE = 0.031, 95% Cr.I = [-0.195, -0.074]), and they also grouped their responses in fewer chunks for the boring background sentences in the varied context block than in the constant context block (β = 0.213, SE = 0.037, 95% Cr.I = [0.139, 0.286]). Name agreement did not interact with the type of background sentences.

First chunk length. As shown in Table 5.2 and the right panel of Figure 5.3, a Bayesian mixed-effect model showed that participants planned more names in their first response chunk for high name agreement pictures than low name agreement pictures (β = 0.259, SE = 0.037, 95% Cr.I = [0.185, 0.332]). First chunk length was also affected by the type of background sentences: participants planned more names in their first response chunk in funny background sentences than boring background sentences in the varied context block (β = 0.182, SE = 0.032, 95% Cr.I = [0.118, 0.245]), and they also planned more names in their first response chunk for the boring background sentences in the varied context block than in the constant context block (β = -0.289, SE = 0.044, 95% Cr.I = [-0.375, -0.203]).

There was an interaction between name agreement and context (β = 0.188, SE = 0.076, 95% Cr.I = [0.040, 0.336]), showing that participants planned more names for high name agreement pictures in their first response chunk for the boring background sentences in the varied context block than in the constant context block (β = -0.199, SE = 0.055, 95% Cr.I = [-0.310, -0.092]), and the magnitude of the context effect was much larger for low name agreement pictures (β = -0.385, SE = 0.059, 95% Cr.I = [-0.501, -0.269]).

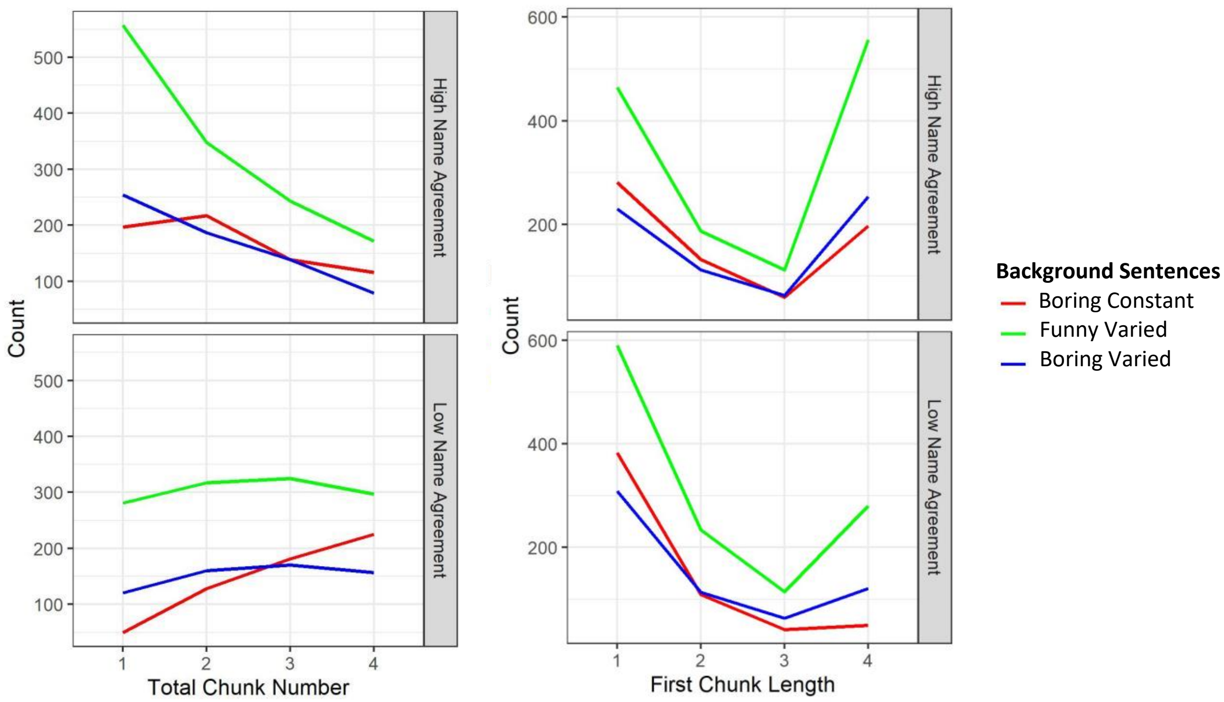

Figure 5.3. Total chunk number (left) and first chunk length (right) split by name agreement (NA: high, low) and the type of background sentences (Boring sentences in the constant context block, funny sentences in the varied context block, boring sentences in the varied context block).

Table 5.2. Results of Bayesian mixed-effect models for all dependent variables.

		Estimate	Est.error	95% Cr. I		Effective samples
				lower	upper	
Accuracy						
Population-level effects	Intercept	2.943	0.199	2.561	3.341	35597
	Name Agreement	**1.194**	**0.28**	**0.651**	**1.757**	**81422**
	FunnyV vs. BoringV	-0.056	0.187	-0.427	0.307	101178
	BoringC vs. BoringV	0.079	0.262	-0.418	0.61	92269
	NA × (FV vs. BV)	-0.588	0.364	-1.311	0.124	120717
	NA × (BC vs. BV)	0.891	0.47	-0.034	1.805	95236
Group-level effects	Participants					
	sd(Intercept)	0.989	0.128	0.768	1.269	30627
	sd(NA)	0.443	0.214	0.043	0.873	18801
	sd(FV vs. BV)	0.16	0.124	0.006	0.459	59334
	sd(BC vs. BV)	0.659	0.274	0.097	1.191	25493
	sd(NA×(FV vs. BV))	0.307	0.235	0.012	0.874	63534
	sd(NA×(BC vs. BV))	0.507	0.368	0.021	1.363	47567
	Items					
	sd(Intercept)	0.555	0.286	0.029	0.993	3610
	sd(NA)	1.004	0.577	0.048	1.945	3607
	sd(FV vs. BV)	0.163	0.124	0.007	0.463	61476
	sd(BC vs. BV)	0.313	0.216	0.013	0.799	37838
	sd(NA×(FV vs. BV))	0.32	0.243	0.013	0.9	63164
	sd(NA×(BC vs. BV))	0.593	0.416	0.025	1.537	38174
Log-transformed onset latency						
Population-level effects	Intercept	7.243	0.036	7.172	7.315	4083
	Name Agreement	**-0.158**	**0.022**	**-0.201**	**-0.115**	**39825**
	FunnyV vs. BoringV	**-0.133**	**0.016**	**-0.164**	**-0.102**	**40994**
	BoringC vs. BoringV	**0.193**	**0.022**	**0.148**	**0.237**	**47484**

	NA × (FV vs. BV)	**0.07**	**0.024**	**0.023**	**0.117**	**68853**
	NA × (BC vs. BV)	**-0.07**	**0.03**	**-0.128**	**-0.012**	**76883**
Group-level effects	Participants					
	sd(Intercept)	0.245	0.026	0.2	0.302	9687
	sd(NA)	0.045	0.011	0.023	0.068	31781
	sd(FV vs. BV)	0.076	0.017	0.044	0.11	32152
	sd(BC vs. BV)	0.127	0.019	0.092	0.168	37732
	sd(NA×(FV vs. BV))	0.048	0.027	0.003	0.105	32189
	sd(NA×(BC vs. BV))	0.088	0.038	0.012	0.161	22718
	Items					
	sd(Intercept)	0.051	0.015	0.021	0.077	6174
	sd(NA)	0.102	0.03	0.041	0.155	5801
	sd(FV vs. BV)	0.016	0.011	0.001	0.042	39530
	sd(BC vs. BV)	0.021	0.015	0.001	0.053	31981
	sd(NA×(FV vs. BV))	0.031	0.023	0.001	0.084	38329
	sd(NA×(BC vs. BV))	0.041	0.029	0.002	0.108	31954
Log-transformed utterance duration						
Population-level effects	Intercept	8.1	0.027	8.048	8.152	5896
	Name Agreement	**-0.222**	**0.023**	**-0.268**	**-0.176**	**44228**
	FunnyV vs. BoringV	**-0.131**	**0.015**	**-0.161**	**-0.101**	**20905**
	BoringC vs. BoringV	**0.198**	**0.023**	**0.154**	**0.243**	**19299**
	NA × (FV vs. BV)	0.016	0.024	-0.031	0.064	95200
	NA × (BC vs. BV)	**-0.054**	**0.025**	**-0.102**	**-0.005**	**106683**
Group-level effects	Participants					
	sd(Intercept)	0.175	0.018	0.144	0.213	11815
	sd(NA)	0.075	0.01	0.056	0.097	41418
	sd(FV vs. BV)	0.07	0.012	0.047	0.096	51994
	sd(BC vs. BV)	0.136	0.018	0.104	0.175	37934
	sd(NA×(FV vs. BV))	0.041	0.026	0.002	0.096	28169

	sd(NA×(BC vs. BV))	0.036	0.025	0.002	0.091	39538
	Items					
	sd(Intercept)	0.048	0.022	0.004	0.083	3687
	sd(NA)	0.097	0.045	0.008	0.167	3761
	sd(FV vs. BV)	0.032	0.018	0.002	0.066	11249
	sd(BC vs. BV)	0.022	0.015	0.001	0.054	29962
	sd(NA×(FV vs. BV))	0.064	0.036	0.004	0.133	11812
	sd(NA×(BC vs. BV))	0.044	0.029	0.002	0.108	29944
Log-transformed total pause time						
Population-level effects	Intercept	6.491	0.077	6.34	6.642	6090
	Name Agreement	**-0.633**	**0.065**	**-0.761**	**-0.506**	**32226**
	FunnyV vs. BoringV	**-0.296**	**0.052**	**-0.399**	**-0.194**	**28808**
	BoringC vs. BoringV	**0.429**	**0.074**	**0.283**	**0.575**	**26904**
	NA × (FV vs. BV)	0.155	0.093	-0.029	0.337	67820
	NA × (BC vs. BV)	**-0.203**	**0.096**	**-0.39**	**-0.014**	**73823**
Group-level effects	Participants					
	sd(Intercept)	0.5	0.053	0.408	0.615	11517
	sd(NA)	0.216	0.038	0.147	0.296	40241
	sd(FV vs. BV)	0.224	0.05	0.129	0.326	36955
	sd(BC vs. BV)	0.409	0.062	0.297	0.54	34161
	sd(NA×(FV vs. BV))	0.298	0.124	0.037	0.531	16709
	sd(NA×(BC vs. BV))	0.174	0.11	0.008	0.411	24421
	Items					
	sd(Intercept)	0.121	0.067	0.005	0.223	1854
	sd(NA)	0.235	0.135	0.01	0.442	1907
	sd(FV vs. BV)	0.077	0.051	0.004	0.188	15450
	sd(BC vs. BV)	0.088	0.057	0.004	0.209	9434
	sd(NA×(FV vs. BV))	0.154	0.103	0.007	0.378	16011
	sd(NA×(BC vs. BV))	0.18	0.114	0.009	0.421	9980

Log-transformed articulation time						
Population-level effects	Intercept	7.808	0.023	7.762	7.854	5084
	Name Agreement	**-0.089**	**0.022**	**-0.133**	**-0.045**	**46868**
	FunnyV vs. BoringV	**-0.07**	**0.01**	**-0.09**	**-0.051**	**32837**
	BorinC vs. BoringV	**0.12**	**0.016**	**0.089**	**0.152**	**25543**
	NA × (FV vs. BV)	-0.02	0.018	-0.055	0.016	43186
	NA × (BC vs. BV)	0.016	0.027	-0.037	0.069	41557
Group-level effects	Participants					
	sd(Intercept)	0.148	0.016	0.121	0.182	12268
	sd(NA)	0.062	0.008	0.048	0.079	33916
	sd(FV vs. BV)	0.049	0.008	0.034	0.067	48058
	sd(BC vs. BV)	0.094	0.012	0.073	0.121	34880
	sd(NA×(FV vs. BV))	0.08	0.016	0.051	0.113	51847
	sd(NA×(BC vs. BV))	0.151	0.022	0.112	0.199	40086
	Items					
	sd(Intercept)	0.046	0.026	0.002	0.085	1775
	sd(NA)	0.095	0.052	0.004	0.171	1742
	sd(FV vs. BV)	0.018	0.01	0.001	0.038	8761
	sd(BC vs. BV)	0.021	0.012	0.001	0.045	7142
	sd(NA×(FV vs. BV))	0.036	0.02	0.002	0.076	8662
	sd(NA×(BC vs. BV))	0.041	0.024	0.002	0.09	7864
Total chunk number						
Population-level effects	Intercept	0.828	0.038	0.754	0.902	12932
	Name Agreement	**-0.261**	**0.03**	**-0.321**	**-0.202**	**54093**
	FunnyV vs. BoringV	**-0.134**	**0.031**	**-0.195**	**-0.074**	**91117**
	BoringC vs. BoringV	**0.213**	**0.037**	**0.139**	**0.286**	**74658**
	NA × (FV vs. BV)	0.05	0.058	-0.063	0.164	116871
	NA × (BC vs. BV)	-0.087	0.065	-0.214	0.041	114873
	Participants					

	sd(Intercept)	0.246	0.027	0.199	0.303	19867
	sd(NA)	0.085	0.023	0.04	0.132	50300
	sd(FV vs. BV)	0.059	0.035	0.003	0.134	28510
	sd(BC vs. BV)	0.126	0.05	0.022	0.223	21350
	sd(NA×(FV vs. BV))	0.055	0.041	0.002	0.154	57812
	sd(NA×(BC vs. BV))	0.075	0.054	0.003	0.2	51941
Group-level effects	Items					
	sd(Intercept)	0.047	0.025	0.003	0.092	6684
	sd(NA)	0.094	0.051	0.005	0.184	6883
	sd(FV vs. BV)	0.024	0.019	0.001	0.069	66063
	sd(BC vs. BV)	0.027	0.021	0.001	0.076	65759
	sd(NA×(FV vs. BV))	0.048	0.037	0.002	0.138	60979
	sd(NA×(BC vs. BV))	0.054	0.041	0.002	0.154	64348
First chunk length						
	Intercept	0.728	0.045	0.64	0.816	9491
	Name Agreement	**0.259**	**0.037**	**0.185**	**0.332**	**61021**
Population-level effects	FunnyV vs. BoringV	**0.182**	**0.032**	**0.118**	**0.245**	**80095**
	BoringC vs. BoringV	**-0.289**	**0.044**	**-0.375**	**-0.203**	**69901**
	NA × (FV vs. BV)	-0.114	0.062	-0.234	0.006	88152
	NA × (BC vs. BV)	**0.188**	**0.076**	**0.04**	**0.336**	**84020**
	Participants					
	sd(Intercept)	0.288	0.033	0.232	0.36	15636
	sd(NA)	0.036	0.026	0.002	0.096	30782
	sd(FV vs. BV)	0.053	0.036	0.002	0.134	26830
Group-level effects	sd(BC vs. BV)	0.154	0.059	0.029	0.268	16280
	sd(NA×(FV vs. BV))	0.064	0.049	0.003	0.181	47825
	sd(NA×(BC vs. BV))	0.103	0.072	0.004	0.267	36798
	Items					
	sd(Intercept)	0.072	0.04	0.003	0.136	3409

sd(NA)	0.145	0.08	0.006	0.273	3432
sd(FV vs. BV)	0.032	0.024	0.001	0.089	47380
sd(BC vs. BV)	0.035	0.027	0.001	0.1	52536
sd(NA×(FV vs. BV))	0.064	0.047	0.002	0.176	47514
sd(NA×(BC vs. BV))	0.07	0.053	0.003	0.198	55527

Note. Models for all dependent variables were run for 24000 iterations. Bolded values indicate effects where the 95% Cr.I does not contain zero. NA refers to name agreement, BC refers to boring sentences in the constant context block, FV refers to funny sentences in the varied context block, BV refers to boring sentences in the varied context block.

5.4 Discussion

This study was designed to explore how the interestingness (funny versus boring) and contextual variation (varied versus constant) of background sentences affected spoken language production, with a focus on how the two factors impact lexical selection in speech planning. As predicted, we obtained consistent name agreement effects on all measures, showing that pictures with low name agreement decreased naming accuracy, slowed down speech planning, and reduced the number of planned utterance units in each response relative to those with high name agreement. Contrary to our predictions that funny sentences (compared with boring sentences) and varied context (compared with constant context) should elicit more interference, we found that funny sentences elicited less interference than boring sentences in the varied context, and boring sentences in the varied context cause less disruptions than those in the constant context. Both interestingness (funny versus boring) and context (varied versus constant) effects reflected on all timing and response chunking measures.

Contrary to our prediction that the effects of interestingness and context should be reduced or eliminated by increased lexical selection demand in speech production, we found that these effects were magnified for low name agreement pictures relative to high name agreement pictures. Specifically, the magnitude of the interestingness effect was larger for low name agreement pictures than high name agreement pictures, with faster onset latencies in the presence of funny sentences than boring sentences in the varied context. The magnitude of the context effect was also larger for low name agreement pictures, showing shorter onset latencies, utterance duration, and total pause time, and longer first chunk length for boring sentences in the varied context than in the constant context. Combined, these findings suggest that the interestingness and contextual variation of background sentences influenced speech production, although in an opposite direction than we predicted, and the influence is modulated by the difficulty of speech production.

5.4.1 Lexical selection demand affects speech production

We found consistent name agreement effects on all measures, which suggests that lexical selection demand affects naming accuracies, planning speed, and the number of planned utterance units in each response. This pattern of results replicated the findings of earlier work using classic single- and multi-picture naming paradigms (e.g., Alario et al., 2004; Shao et al., 2014) and a speaking-while-listening paradigm (He et al., 2021, Chapter 4). As shown in previous studies (He et al., 2021, Chapter 4), name agreement effects on time measures (onset latencies, utterance duration, total pause time, and articulation time) reflect that the demand of lexical selection affects processing before and after starting to speak. This further implies that speakers select and retrieve picture names during the whole process of planning the sequence of picture names. This is in line with the claim that speakers plan speech incrementally because they cannot retrieve all the picture names before articulation and instead have to coordinate the planning and articulation of successive words (e.g., Levelt et al., 1999; Roelofs, 1998; Wheeldon & Lahiri, 1997).

Name agreement effects were also obtained on the measure of response chunking, such that participants grouped their speech responses in relatively more but shorter chunks for low name agreement pictures. This indicates that speakers change the way they plan for low name agreement pictures: they opt to plan their speech with less temporal overlap, resulting in more and shorter separate response chunks. Combined, this means that we replicated the findings in He et al. (Chapter 4) that name agreement affected naming accuracy, time measures, and response chunking measures, which in turn implies that name agreement effects in a speaking-while-listening task are stable across stimuli and groups of participants.

5.4.2 The interestingness of background sentences affects speech production

The present study showed consistent interestingness effects (funny versus boring sentences) on all measures except naming accuracies. Notably, participants took less time and made fewer but longer response chunks when naming pictures in the presence of funny sentences than boring sentences in the varied context. This pattern of results is the opposite of our prediction under the specific attention capture view (Eimer et al., 1996). We expected that in the varied context funny sentences should cause more disruption than boring sentences due to their increased linguistic complexity (thematically, semantically, phonologically, and syntactically). Below we discuss possible interpretations for the reversed pattern.

One possibility is that funny sentences cause less interference because they not only elicit a passive bottom-up response but also active top-down cognitive control, whereas boring sentences only evoke a passive bottom-up response. That is, funny sentences are rich in linguistic information, which has great potency to divert attention away from the picture naming task and then impair speech production performance. It is possible that speakers realize the greater potential of attentional capture by funny sentences, and they increase their top-down cognitive control to concentrate more on the speech production task. This, in turn, leads to faster speech planning and more overlapping responses in the presence of funny sentences. However, the increased top-down cognitive control does not occur when planning speech in the presence of boring sentences because speakers have enough attentional resources to process speech planning and background speech.

Alternatively, the relatively reduced interference elicited by funny sentences could be because the auditory disruption on speech production is mainly caused by concrete nouns involved in background speech. In the present study, the average number of concrete nouns in each funny sentence (four concrete nouns) is lower than in the boring sentences (six concrete nouns) because funny sentences have been long enough (i.e., to match total number of words with boring sentences) that cannot contain more nouns. If the disruption is indeed caused by the concrete nouns, then funny sentences would produce less disruption on speech production than boring sentences. This also would explain the absence of an effect of word lists versus sentences in He et al. (Chapter 4), as their word lists and sentences all contained six concrete nouns, perhaps then causing the same degree of auditory disruption on speech production. One potential argument against this, however, is that if only concrete nouns are responsible for the disruption, then boring sentences in the constant context should have the same potential to elicit interference as those in the varied context; we found instead boring sentences in the constant context were more disruptive than those in the varied context.

Another possibility for the relatively reduced disruption by funny sentences is that they are so amusing that speakers attempt to listen to them. To this end, speakers might have tried to finish picture naming as fast as possible by speeding up planning and making their responses more overlapping, leaving more time to listen to the funny sentences. Because participants clicked on the mouse to end the trials and we did not record mouse-clicking latencies, we do not know how much of each background sentence they listened to on any trial. In addition, because the present study only matched the two types of sentences on the number of words and duration (8s), other differences at multiple levels, such as linguistic richness, prosodic pattern, speech rate etc., may also lead to different disruptive effects. More research with well-matched funny and boring sentences is needed to explore how the interestingness affects speech production.

As a follow-up of He et al. (Chapter 4), the current study obtained considerable interestingness effects on all measures except accuracy, although not in the predicted way. This in turn suggests that the lack of difference between word lists and sentences in He et al. (Chapter 4) was because the background stimuli were too boring and uniform (i.e., word lists had a regular acoustic pattern and sentences had uniform syntactic structure) and has relatively impoverished attentional capture properties. Combined, the consistent interestingness effects in the present study suggest that different types of background sentences have different potentials to interfere with speech production via different underlying mechanisms, such as bottom-up attention capture and / or top-down cognitive control.

5.4.3 The contextual variation of background sentences influences speech production

The current study found that the contextual variation of background sentences affected speech production, showing decreased planning speed and shorter planned utterance units for boring sentences in the constant context than in the varied context. This finding is contrary to our prediction under the habituation account (Cowan, 1995, 1999) that boring sentences should cause less disruption in the constant context than in the varied context because speakers can adapt to constant background speech. The reversed pattern of context effect implies that adaptation to background speech with constant properties (e.g., uniform syntactic structure) may not occur in auditory disruption on speech production. This also rules out the possibility that the relatively weak disruption in He et al. (Chapter 4) happened because speakers adapt to background speech with regular acoustic patterns (such as word lists) and uniform syntactic structures (i.e., boring sentences) easily and then reduce disruption on speech production performance.

One possible mechanism for the reversed context effect, where the boring sentences in the constant context elicited more interference than in the varied context, is in overall attentional control. Perhaps the varied context made speakers stay alert and increase their global attention level on picture naming. In turn, this resulted in better performance with faster speech planning and larger response chunks. In other words, speakers may employ top-down strategies to shield against auditory disruption in the varied context. Once they realize background speech varies, they can increase top-down cognitive control and concentrate more on speech production to escape possible auditory disruption.

We also cannot rule out the possibility that reduced interference by boring sentences in the varied context compared to the constant context arose because of a practice or learning effect. In the present study, the constant context block (only contains boring sentences) always preceded the varied context block (includes boring and funny sentences). This order of the blocks was used to exclude that participants in constant blocks following varied ones might still wait for entertaining sentences to occur and therefore listen particularly attentively. Speakers may have habituated to the boring sentences after finishing the constant context block, allowing them to block off only the boring sentences in the following varied context block more easily, resulting in better performance in the presence of boring sentences in the varied context. Alternatively, speakers may have learned the task requirements very well and become better at ignoring background sentences in the varied context block, resulting in faster speech planning and increased overlapping responses.

5.4.4 Modulation of name agreement on auditory disruption by background sentences

This study found that the level of interference by background sentences was modulated by the lexical selection demands of speech production. The magnitude of the interestingness effect (funny versus boring) and the context effect (varied versus constant) was larger for low name agreement pictures than high name agreement pictures. For difficult—low name agreement—picture naming, the interestingness effect was only obtained for planning before speaking (i.e., only on onset latency), while the context effect was detected during the whole process of naming (i.e., on all timing measures except articulation time, and also on first chunk length). The finding is inconsistent with our prediction following the attention engagement account (Halin et al., 2014; Marsh et al., 2015) that both effects should be reduced when naming difficult (i.e., low name agreement) pictures because increased speech production difficulty should make speakers concentrate more on speech production and reduce the processing of background speech. Our pattern of results implies that attention engagement may not work for difficult speech production when background speech is funny and varied enough.

The relatively large interestingness and context effects for difficult—low name agreement—picture naming are consistent with the *load theory of attention* (Lavie, 2005; Lavie & Dalton, 2014). This load theory assumes that load on executive control functions renders them unavailable to actively maintain stimulus-processing priorities throughout task performance, and then increases interference from irrelevant background distractors. In this case, difficult picture naming (i.e., low name agreement pictures) should impose higher demands on cognitive control and then increase auditory distraction, leading to larger effects of interestingness and context.

The discrepancy in results between the current study and He et al. (2021, Chapter 4) makes it difficult to support either the load theory of attention (Lavie, 2005; Lavie & Dalton, 2014) or attention engagement account (Halin et al., 2014; Marsh et al., 2015), and thus requires an explanation taking into account other properties of the background speech. That is, when background speech has a regular acoustic pattern (e.g., word lists with similar pauses between two consecutive words) and uniform syntactic structure (e.g., boring sentence), the auditory disruption can be reduced (or eliminated) by high lexical selection demand via an attention engagement mechanism (Halin et al., 2014; Marsh et al., 2015). By contrast, when background speech is funny and varied enough, auditory distraction is increased when the lexical selection demand of speech production is high due to high load on cognitive control (Lavie, 2005; Lavie & Dalton, 2014). That is, difficult picture naming imposes higher demand on cognitive control, which makes speakers unavailable to actively maintain speech planning, which increased the interference by background speech.

Combined with previous research (He et al., 2021, Chapter 4), we detected a reversal of modulations by speech production difficulty on irrelevant speech effects. This implies that there are different ways of processing of different types of background speech (e.g., simple word lists/sentences versus complex sentences), and different compensatory mechanisms speakers may use to avoid interference from the background speech. More research is required to determine why demanding speech production increases auditory distraction from certain types of background speech (e.g., funny sentences), but reduces disruption in the presence of other types of background speech (e.g., word lists and boring sentences).

5.4.5 Outlook

While the present study provides some insights into how the interestingness and contextual variation of background sentences interfere with speech production, more work is needed to reveal how speakers plan their production in the presence of background speech. For example, a comparison between funny and boring sentences matched on specific content and acoustic variation would be the key to understanding our results. This is because funny sentences often concern topics that are particularly interesting, meaningful, have varied acoustic properties, or have complex thematic and syntactic structure, which results that the manipulation of the degree of interestingness may be not perfect. Thus, carefully controlled background sentences are needed before firm conclusions can be reached about the interestingness effect on speech production.

While lexical selection demands (in the form of name agreement) modulate background speech effects (e.g., He et al., 2021, Chapter 4), it is unclear whether the same pattern would obtain for other aspects of speech production such as object recognition, phonological encoding, and phonetic encoding. This provides a fruitful direction in for future studies on encoding difficulty and speech production. Finally, the speech production task in existing studies (e.g., He et al., 2021, Chapter 4) using speaking-while-listening paradigm is relatively easy—naming sets of pictures. To increase ecological validity, future studies should utilize more demanding tasks such as phrase, sentence, or dialogue production in the presence of background speech. This would reveal how speakers plan their speech in real-world settings such as on a train or in a restaurant.

5.5 Conclusions

This study showed that the interestingness and contextual variation of background sentences influenced speech production, and that the influences were modulated by the difficulty of speech production, although in a reversal of predictions. The reversed pattern of results implies that funny and varied background speech may evoke speakers' top-down cognitive control and then cause less interference, and that the interestingness and context effects are magnified when speech production is difficult due to high load on cognitive control. An implication of the current study is that speakers are able to manage disruption from background speech by changing when and how they plan their speech.

Acknowledgments

We would like to thank Maarten van den Heuvel and Thijs Rinsma for programming; Annelies van Wijngaarden and Sophie Slaats for translating and recording materials; Elizardo Laclé, Elsa Opheij, Marije Veeneman, Marjolein van Hoften, Marius Wolsink, and Charlotte Borgers for data coding.

Appendices

Appendix A. Stimuli used in the current study

Table A1. 224 pictures used in the present study.

Picture Grid	*Picture 1*	*Picture 2*	*Picture 3*	*Picture 4*	*Picture Grid*	*Picture 1*	*Picture 2*	*Picture 3*	*Picture 4*
Pictures with high name agreement									
1	koelkast	pijl	gevangenis	dolfijn	15	microfoon	koning	bloem	stier
2	leeuw	kruiwagen	driehoek	tomaat	16	kokosnoot	steen	gitaar	egel
3	harp	radio	knie	paprika	17	roos	kroon	trechter	ballon
4	vlinder	trap	batterij	cactus	18	honing	slak	weegschaal	rug
5	kiwi	zaag	vliegtuig	bezem	19	muis	drumstel	tandarts	parachute
6	schaap	waaier	glas	baard	20	ananas	spiegel	robot	zaklamp
7	konijn	doedelzak	ster	handschoen	21	broek	schilderij	kangoeroe	tunnel
8	pijp	hamer	duim	berg	22	ketting	sleutel	dobbelsteen	rechter
9	eekhoorn	keuken	orkest	banaan	23	stopcontact	ezel	diamant	arm
10	geit	schaduw	horloge	kompas	24	clown	eiland	schildpad	bril
11	skelet	kaars	pompoen	vlieger	25	fruit	vlag	aansteker	lepel
12	heks	aardappel	vleermuis	boog	26	kikker	wasmachine	bokser	trompet
13	zwembad	masker	bijbel	kanon	27	sok	bus	fabriek	vork

14	rups	schaar	kraan	puzzel	28	papegaai	helikopter	riem	toetsenbord
Pictures with low name agreement									
1	trui	baksteen	jager	klauw	15	koekje	garage	cirkel	mossel
2	lade	schedel	foto	melk	16	camping	pruik	sneeuw	ballerina
3	duif	nagel	kerkhof	speer	17	munt	strand	kameel	lamp
4	troon	parel	engel	viool	18	kleed	tram	doodskist	garnaal
5	vogelkooi	snoepje	kasteel	brievenbus	19	inktvis	staart	perzik	herder
6	kerk	schoolbord	bank	walrus	20	hersenen	ijsberg	kwast	sigaret
7	soldaat	vis	gorilla	kruk	21	gymzaal	leraar	handdoek	worst
8	hagedis	armband	kogel	rimpels	22	museum	tuinslang	kegel	druif
9	ijsje	paus	badkuip	spuit	23	soep	koningin	trein	buik
10	varken	broekzak	naald	wasbak	24	olie	antenne	piano	knuffel
11	gevangene	brug	hengel	driewieler	25	planeet	motor	gang	litteken
12	vinger	magneet	plas	zanger	26	domino	badkamer	wortels	komkommer
13	hoorn	blad	raam	jurk	27	koffie	elf	put	schatkist
14	monster	rivier	pion	goochelaar	28	prullenbak	meloen	schelp	ridder

Table A2. 56 boring sentences used in the present study.

No.	*Boring sentences*
1	De dierenarts en het terras bevinden zich links van het hotel, en het vuilnis en de rem bevinden zich rechts van de fluit.
2	De tak en de raket bevinden zich links van het schild, en de prinses en de beker bevinden zich links van de veer.
3	De kers en de zonnebloem bevinden zich rechts van het luipaard, en de boot en de medaille bevinden zich links van de fee.
4	De plant en het scheermes bevinden zich rechts van het oog, en de accordeon en de kaas bevinden zich rechts van de uil.
5	Het hek en het dienblad bevinden zich links van de slang, en de ring en de kubus bevinden zich rechts van de watermeloen.
6	Het bot en de pop bevinden zich links van de molen, en de tas en de veter bevinden zich links van de krokodil.
7	De zeep en de vulkaan bevinden zich rechts van de mand, en de rekenmachine en de kano bevinden zich links van het paard.
8	De boom en de kast bevinden zich rechts van de ambulance, en de schoen en het vierkant bevinden zich rechts van de gier.
9	De friet en de regenjas bevinden zich links van het lijf, en de mist en de doek bevinden zich rechts van het apparaat.
10	De rijst en de dierentuin bevinden zich links van de wolken, en de armen en de zaal bevinden zich links van de halsband.
11	De boter en de doos bevinden zich rechts van de houthakker, en het kopje en de wolf bevinden zich links van het vliegveld.
12	De kabel en de wieg bevinden zich rechts van het aquarium, en de zuster en de rok bevinden zich rechts van de vlieg.
13	De boon en de fakkel bevinden zich links van de camera, en de spijker en het nijlpaard bevinden zich rechts van de koffer.
14	Het altaar en het muntje bevinden zich links van de dame, en het gazon en de frisdrank bevinden zich links van het vel.
15	De kip en de visser bevinden zich rechts van de olijf, en de haak en het cadeau bevinden zich links van de pet.
16	Het koor en het hert bevinden zich rechts van de portemonnee, en het bord en de lijst bevinden zich rechts van de zwemmer.
17	De glimlach en het tapijt bevinden zich links van de vruchten, en de wagen en de mantel bevinden zich rechts van het hokje.
18	Het slot en de wandelstok bevinden zich links van de vogel, en de knoop en de theepot bevinden zich links van de appel.

19	Het handvat en de tweeling bevinden zich rechts van het zand, en de fontein en de eetkamer bevinden zich links van de vingers.
20	De brief en de liniaal bevinden zich rechts van de tank, en het zout en de ventilator bevinden zich rechts van de adelaar.
21	De voeding en de grot bevinden zich links van de onderbroek, en de lantaarnpaal en de winkel bevinden zich rechts van het rietje.
22	De zonsopgang en de tent bevinden zich links van de woonkamer, en de rechterhand en de noten bevinden zich links van de loodgieter.
23	De taart en het vuurwerk bevinden zich rechts van de oprit, en de rol en de waterval bevinden zich links van de zetel.
24	De zolder en de verf bevinden zich rechts van de tekening, en de lift en de druiven bevinden zich rechts van het rek.
25	De rozen en de telescoop bevinden zich links van de vrucht, en de zaklantaarn en de woestijn bevinden zich rechts van de neus.
26	De indiaan en de rits bevinden zich links van de eieren, en het graf en de tepel bevinden zich links van de vleugels.
27	De oorbel en de tong bevinden zich rechts van de vrachtwagen, en de walnoot en de lerares bevinden zich links van de hoek.
28	De danser en de lucifer bevinden zich rechts van het hol, en het ijs en de vuilnisbak bevinden zich rechts van de muts.
29	Het hok en de goal bevinden zich links van de adem, en de deken en de flits bevinden zich rechts van het mannetje.
30	Het voorhoofd en het toilet bevinden zich links van de worp, en het ondergoed en het restaurant bevinden zich links van de zeeman.
31	De zak en de apotheek bevinden zich rechts van het gordijn, en de luier en het eitje bevinden zich links van de regenboog.
32	De nagels en de tuin bevinden zich rechts van de vlam, en de achterdeur en de lakens bevinden zich rechts van de geur.
33	De parfum en de serveerster bevinden zich links van het goud, en de wortel en de voetbal bevinden zich rechts van de koe.
34	De draad en de tractor bevinden zich links van het vest, en de machine en het feestje bevinden zich links van het achterwerk.
35	De tanden en het veld bevinden zich rechts van de danseres, en de ham en de regen bevinden zich links van de zwembroek.
36	De map en de tamboerijn bevinden zich rechts van de jas, en het vlees en de postzegel bevinden zich rechts van de kam.
37	De kruik en de trommel bevinden zich links van de borstel, en het ontbijt en het nest bevinden zich rechts van de stoel.
38	De fotograaf en het roer bevinden zich links van de verwarming, en de tafel en de kas bevinden zich links van de gans.
39	De drank en de muzikant bevinden zich rechts van de traan, en de jacht en het gebit bevinden zich links van de handschoenen.

40	De citroen en het vuur bevinden zich rechts van de spons, en de brandblusser en de kurk bevinden zich rechts van het potlood.
41	De jam en het fluitje bevinden zich links van het lint, en het riool en de voordeur bevinden zich rechts van het gebouw.
42	De hartslag en het jasje bevinden zich links van de ladder, en de zweep en de donder bevinden zich links van de cake.
43	De jojo en de laarzen bevinden zich rechts van het circus, en de vloer en het zonlicht bevinden zich links van de oren.
44	De voet en de klaver bevinden zich rechts van het zwaard, en de schilder en de televisie bevinden zich rechts van de peer.
45	De rots en de lippenstift bevinden zich links van de vaas, en de tovenaar en de zegel bevinden zich rechts van het gaatje.
46	De ijskast en de halsketting bevinden zich links van het venster, en het wapen en het deksel bevinden zich links van de frietjes.
47	Het rijtje en het watje bevinden zich rechts van het gras, en het applaus en de zakdoek bevinden zich links van de emmer.
48	Het leer en de zwaai bevinden zich rechts van de gootsteen, en de tenen en het dessert bevinden zich rechts van de reus.
49	De rolstoel en de cowboy bevinden zich links van het dorp, en de inkt en de tennisbal bevinden zich rechts van de hak.
50	De grond en de mosterd bevinden zich links van de druppels, en het lintje en de vuurtoren bevinden zich links van het oerwoud.
51	De tuinman en de rotsen bevinden zich rechts van de linkerhand, en de achtertuin en de computer bevinden zich links van de gordel.
52	Het tasje en de wekker bevinden zich rechts van de afwas, en de enkel en de lasagne bevinden zich rechts van de zeemeermin.
53	De meubels en de tang bevinden zich links van de voetstappen, en het zaad en de laars bevinden zich rechts van de duivel.
54	De timmerman en het gips bevinden zich links van de wanten, en de muziek en het hooi bevinden zich links van de vijver.
55	Het leger en de yoghurt bevinden zich rechts van de grasmaaier, en het duister en de toeter bevinden zich links van het hoofdje.
56	Het gelach en de lap bevinden zich rechts van de inbreker, en het juweel en het ziekenhuis bevinden zich rechts van de reep.

Table A3. 56 funny sentences used in the present study.

No.	*Funny sentences*
1	De stoel danste samen met de lamp in de woonkamer terwijl de rest van het huis aan het slapen was.
2	De lampen aten geitenkaas op hun broodje tijdens de lunch in de wei van de autoband.
3	De deur klom de steile heuvel op met de zelfgemaakte asbak vlak achter zich aan.
4	De koffiekan zwom moeiteloos naar de overkant van het kanaal tijdens de jaarlijkse wedstrijd in Nederland.
5	De olifant knuffelt de grote mieren die elke dag in zijn verblijf verschijnen meerdere malen.
6	De giraf leest een goed boek over de bruine bananen die de wereld veroverden in 10 dagen.
7	Het leger van aardbeien vecht tegen het leger van spijkerbroeken bij het vallen van de nacht.
8	De grote bomen zingen het Nederlandse volkslied ook dit jaar uit volle borst mee op koningsdag.
9	Het krukje loopt mank naar de winkel van de buurman voor een nieuwe houten poot.
10	De mol fietst op zijn gloednieuwe fiets door het drassige weiland tijdens de opkomst van de zon.
11	De hamster kocht nieuwe bergschoenen voor zijn lange reis naar de hoofdstad van Sri Lanka.
12	Het dak luistert aandachtig naar de ruziemakende buren op de vierde verdieping van het flatgebouw.
13	Het theelepeltje verbrandt zich aan het hete water uit de zwarte fluitketel op het fornuis.
14	De sleutelbos ruimt de inhoud van de handtas van de chaotische studente voor de derde keer op.
15	De grote vuilnisbak drinkt de resten van de lege colafles op bij het bezoek aan de huisarts.
16	Het lammetje breit een warme deken voor de wintermaanden van het wol van zijn familie.
17	De drukke rotonde smeert zijn lunch om het onderweg naar zijn werk op te eten.
18	De wolk springt in de lucht van blijdschap omdat het zonnetje de hele dag al schijnt.
19	De blinde kraaien kijken naar een horrorfilm in de bioscoop met popcorn in hun handen.

20	De cactus versiert de taart voor zijn eigen verjaardagsfeest van morgen met veel slagroom.
21	De handdoek zwom samen met de gele badeendjes in het zwembad voor vier uur lang.
22	De koptelefoon drinkt een flesje water na het lopen van een halve marathon.
23	De groene basketbal zweeft boven het hoofd van de engel zonder vleugels.
24	Het meisje vliegt samen met haar kleine hondje naar de grootste toren van de maan.
25	Een groepje kevers huilt vanwege het emotionele lied dat ze op de radio hoorden.
26	Het drumstel bouwt een houten kast met de onderdelen uit het grote Ikea pakket.
27	Het boek sliep samen met de computer in een tent onder de heldere sterrenhemel.
28	De egel knuffelde de ballon die hij van zijn beste vriend kreeg voor zijn verjaardag.
29	Het mes streelt de zwarte zwerf kat die iedere dag rond het avondeten langskomt.
30	De verjaardagskaart belt de koningin van Engeland elk jaar om 9 uur 's ochtends om haar te feliciteren.
31	De stapel bussen maakte zich zorgen dat ze nooit gebruikt worden voor een excursie.
32	De startkabel schrok van de kleine man met de grote hoed van stro op zijn hoofd.
33	Het anker zwemt samen met de moersleutel over het roze kanaal heen midden in de kerstvakantie.
34	De duizendpoot schaatst met de courgette bij de recent gerenoveerde schaatsbaan in Duitsland.
35	Het waterflesje spoelt zijn mond na het eten van een onsmakelijk dessert tijdens het etentje met de gootsteen.
36	Spiegels roddelen dagelijks over de vele twijfelachtige kledingkeuzes die ze die dag hebben gezien.
37	De leguaan had hoogtevrees en ging daarom liever niet mee in het reuzenrad met de dennenappels.
38	De flamingo knuffelde verschillende gazelles nadat hij weer een drukke dag had gehad op kantoor.
39	De vlinders applaudisseerden enthousiast voor het geslaagde optreden van het symfonie-orkest.

40	De deur rent een marathon in Frankrijk en wint een gouden medaille op de olympische spelen.
41	De kokosnoot schrijft een liefdesbrief aan de hoge berg op het onbewoonde eiland.
42	Het ontbijtbordje raapt midden in de nacht de kruimels van de pas geveegde houten vloer op.
43	Het onzekere fruitvliegje fluistert een groot geheim in het linkeroor van de geschilde peer.
44	De toets geniet van de studenten die gespannen aan het schrijven zijn op een zonnige maandagmiddag.
45	Stofzuigers eten in het weekend graag zand uit de speeltuin van de stoomlocomotief als ontbijt.
46	De perzik vermijdt het om naar de lokale sportschool van de grote beer te gaan.
47	De tegel voer met zijn blauwe speedboot door de smalle rivieren van het grote Amazonegebied.
48	Bananen vinden het lekker om als drankje bij het avondeten een eiwit-smoothie te nemen.
49	De mier verschoonde zijn vieze beddengoed voor de tweede keer in minder dan een week tijd.
50	De lantaarnpaal keek met open ogen naar de mensen aan het einde van de straat.
51	De Leeuw danste vol zelfvertrouwen de hele avond lang op liedjes van de leeuwenkoning-film.
52	De zeester applaudisseerde uitbundig toen de broer van zijn beste vriend als eerste over de finish was.
53	De krab verkocht zijn restaurant voor een hele hoge prijs op de laatste dag van de zomer.
54	De blauwe aap leende zijn rode laarzen vorige week dinsdag uit aan de bushalte.
55	De kapstok droeg een bruine manteljas en een rode sjaal onderweg naar zijn vriend de koelkast.
56	De vork bracht een nacht door in het huis van de blije buurman op bevrijdingsdag.

6 | General Discussion

Speaking often takes place in noisy settings such as in a restaurant or on a train. The existing literature has shown that irrelevant background stimuli may disrupt focal cognitive task performance in, for example, serial recall (e.g., Colle & Welsh, 1976; Hughes et al., 2007) and reading (Cauchard et al., 2012; Hyönä & Ekholm, 2016). This effect is referred to as the irrelevant sound effect. Two types of theories have been proposed to account for the irrelevant sound effect: the domain-specific interference-by-similarity account (Jones & Macken, 1993; Martin et al., 1988; Salamé & Baddeley, 1982, 1989) and the domain-general attention capture account (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015). While the first account assumes that focal task interference is caused by shared/similar representations activated by background stimuli, the latter hypothesizes that background stimuli disrupt focal task performance because they divert attention away from the task. However, little work has investigated the influence of background stimuli, especially irrelevant background speech, on the processing of producing spoken language. One goal of the present dissertation was to explore whether and how speech production was affected by irrelevant background speech. Moreover, prior studies have found that the interference elicited by background noise or speech can be reduced or eliminated by an increase in focal task difficulty (e.g., Halin et al., 2014; Hughes, 2014; Marsh et al., 2015). This finding has been explained via the attention engagement account (Halin et al., 2014; Marsh et al., 2015), which assumes that a difficult focal task makes individuals concentrate harder on it, such that the involuntary processing of background information is reduced. Therefore, a second goal of the present dissertation was to investigate whether the difficulty of speech production modulated the interference elicited by irrelevant background speech.

6.1 Summary of the empirical studies

The same speech production task was used in all experiments reported in the present thesis—namely multiple picture naming with a variation of name agreement between pictures. Naming performance was assessed by several dependent variables: naming accuracy, onset latency, utterance duration, total pause time, articulation time, total chunk number, and first chunk length (where response chunks were defined as the number of pictures named without an intervening pause).

The goals of the studies presented in Chapter 2 were two-fold. I assessed how speech production was influenced by the representational similarity between background speech and the speech planned by the participants and the attention demand of concurrent listening (i.e., instructions to divide attention or focus attention on main task). Additionally, I tested whether the influence of background speech was modulated by the difficulty of speech production (indexed by name agreement: high versus low). To this end, two experiments were conducted where native Dutch speakers named sets of pictures with high or low name agreement in Dutch in the presence of different types of background speech. Experiment 1 showed that word lists (Dutch and Chinese) were more disruptive than language-like noise (eight-talker babble), and intelligible word lists (Dutch) caused more disruption than unintelligible word lists (Chinese). Experiment 2 showed that speech production performance was worse in the divided-attention condition than the focused- attention condition. The effects of representational similarity (in Experiment 1) and attention demand (in Experiment 2) were absent when naming pictures with low name agreement, where production difficulty was high.

The studies presented in Chapter 4 further explored how different types of irrelevant background speech (word lists versus sentences) influenced speech production relative to a quiet control condition, and whether this influence depended on the intelligibility of the background speech. Two web-based experiments were conducted in which native Dutch speakers named sets of pictures in Dutch with high or low name agreement while ignoring different types of background speech. There were three main findings. Both unintelligible (Chinese, Experiment 1) and intelligible (Dutch, Experiment 2) background speech impaired speech production performance relative to a quiet control condition. There was no difference in the effects of segmented (word lists) and continuous (sentences) background speech in either experiment. Similar to Chapter 2, the disruption caused by intelligible (Dutch) background speech was eliminated when planning low name agreement pictures.

Chapter 5 continued exploring the irrelevant speech effects in speech production by introducing relatively large differences in the background speech: interestingness (funny versus boring) and contextual variation (varied versus constant). This chapter also focused on whether the irrelevant speech effects were modulated by the difficulty of speech production. In a web-based experiment, native Dutch speakers named sets of pictures in Dutch with high or low name agreement while ignoring background sentences in a constant context (only with boring sentences) followed by a varied context (with boring and funny sentences). Here, there were three main findings. Funny sentences caused less disruption on speech production performance than boring sentences. Boring sentences interfered less with speech production in the varied context than in the constant context. Finally, a reversed pattern from Chapters 2 and 4 was obtained: the interestingness and context effects were larger for low than high name agreement pictures.

Finally, Chapter 3 assessed the feasibility of conducting spoken language production research in an online environment. Two previously studied effects related to lexical selection—name agreement and semantic context—were examined. Results showed a stable name agreement effect and a relatively weak semantic context effect in a modified blocked-cyclic naming paradigm, replicating the findings of lab-based studies. However, no interaction between name agreement and semantic context was obtained.

6.2 Lexical selection demand affects speech production performance

As mentioned above, this dissertation investigated how lexical selection demand affected speech production performance by manipulating name agreement (high vs low) of to-be-named pictures. Stable and consistent name agreement effects were obtained on almost all dependent measures, showing that demanding lexical selection decreased naming accuracy and planning speed, and reduced the number of planned utterance units in each response. The name agreement effect on multiple time measures (onset latency, utterance duration, total pause time, and articulation time) is consistent with previous research with classic single- and multiple-picture naming paradigms (e.g., Alario et al., 2004; Shao et al., 2014), which suggests that lexical selection is performed throughout the planning a sequence of picture names, rather than only before the onset of articulation. This is in line with the claim that speakers plan speech incrementally (see Levelt et al., 1999; Roelofs, 1998; Wheeldon & Lahiri, 1997). That is, speakers do not retrieve all picture names before starting to speak but rather have to coordinate the planning and articulation of successive words. Moreover, name agreement effects on response chunking measures (total chunk number and first chunk length)—new dependent variables defined in this dissertation—suggest that lexical selection demand impacts the way speakers group their responses. In particular, planning names of difficult pictures (i.e., low name agreement pictures) with less temporal overlap results in more separate response chunks.

Importantly, this dissertation shows that name agreement effects are stable across different statistical methods (Bayesian and frequentist lme models and bayes factors), error-coding criteria for speech responses (accepting predetermined responses versus any resonable response), experimental paradigms (a speaking-while-listening paradigm in Chapters 2 & 4, & 5, a modified blocked-cyclic naming paradigm in Chapter 3), picture stimuli, groups of populations, and also experimental platforms (a lab-based study for Chapter 2, but web-based studies for Chapters 3, 4, & 5). This suggests that name agreement is a variable that robustly indexes the difficulty of speech production (in terms of lexical selection demand) and is therefore well suited to study how the difficulty of speech production interacts with the irrelevant speech effect.

In addition, the investigation of the name agreement effect combined with the semantic context effect in Chapter 3 provides some insights into a central controversy regarding lexical selection in speech production—namely, whether or not lexical selection is competitive. On the one hand, models with lexical competition (e.g., Abdel Rahman & Melinger, 2009; Howard et al., 2006) predict that increasing the number of activated lemmas during lexical selection for low name agreement pictures would increase the semantic context effect. On the other hand, models not assuming lexical competition (e.g., Oppenheim et al., 2010) predict that name agreement should not interact with semantic context. Chapter 3 showed no interaction between name agreement and semantic context on any dependent variable, which implies that lexical selection may be achieved via non-competitive mechanisms (Oppenheim et al., 2010). Alternatively, there might not have been enough power to detect a subtle interaction. Therefore, more research is required before firm conclusions can be reached about the underlying mechanism of lexical selection in spoken language production.

6.3 Irrelevant background speech affects spoken language production

This dissertation mainly explored how spoken language production was affected by irrelevant background speech with different properties, in particular intelligible versus unintelligible speech, word lists versus sentences, and boring versus funny sentences appearing in constant versus varied context. This section discusses the implications of the findings for theories concerning the impact of background speech on speech production.

In two studies (Experiment 1 in Chapter 2 & Experiment 1 in Chapter 4), background speech that was unintelligible to participants (Chinese word lists) was shown to interfere with speech production relative to non-speech baselines—eight-talker babble or quiet. This is consistent with the phonological disruption view (Baddeley, 1982, 1989) which predicts that any speech sound (intelligible or not) interferes with focal tasks because of similar or shared phonological representations or processes. In this case, unintelligible background speech (Chinese word lists) gained access to the phonological loop component of working memory that speech production required, resulting in disruption. However, the non-speech sound (eight-talker babble) is filtered out by the phonological loop (Salamé & Baddeley, 1987), causing no interference with speech production.

However, the interference elicited by unintelligible (Chinese) word lists may also be caused by stimulus-aspecific variation between background speech conditions. For instance, Chinese word lists are a segmented speech stream with pauses between consecutive words, while eight-talker babble is a continuous sound stream. This difference in stimulus-aspecific variation may also contribute to the degree of disruption on speech production. To test this possibility, Experiment 1 in Chapter 4 manipulated the stimulus-aspecific variation (the presence or absence of pauses) of irrelevant background speech by comparing unintelligible segmented speech (Chinese word lists) and continuous speech (Chinese sentences). No systematic difference was found between the two background speech conditions on any dependent measure. This finding suggests that the interference by unintelligible background speech is caused by phonological similarity, but not by the variation in stimulus-aspecific properties. This goes against the aspecific attention capture view (Eimer et al., 1996), which states that stimulus-aspecific variation (e.g., the context in which it occurs) diverts attention away from the focal task and causes disruption.

The same pattern of results was obtained for the comparison between intelligible segmented speech (Dutch word lists) and continuous speech (Dutch sentences) in Experiment 2 of Chapter 4: No systematic difference between Dutch word lists and sentences was obtained. The results for intelligible background speech are harder to interpret than unintelligible background speech, as the stimulus-aspecific variation (the presence or absence of pauses) of the Dutch background speech is confounded with stimulus-specific properties (e.g., semantic and syntactic information). This leads to the possibility that the absence of a difference between the two background speech conditions arose because the disruption by the presence/absence of pauses in Dutch word lists (via aspecific attentional capture) canceled out the interference by richer linguistic information of semantic/syntactic integration in Dutch sentences (via specific attentional capture). Nevertheless, together, the results indicate that domain-general attention capture (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015), regardless of whether it is aspecific or specific, plays an important role in the irrelevant speech effects in speech production.

This dissertation also compared the interference elicited by unintelligible (Chinese) and intelligible (Dutch) background speech. Specifically, Experiment 1 in Chapter 2 demonstrated that speech production was disrupted more by intelligible (Dutch word lists) than unintelligible (Chinese word lists) background speech. A similar result was found in the comparison across the two experiments in Chapter 4, showing that Dutch background speech (word lists & sentences) elicited more interference with speech production than Chinese background speech. There are a number of probable reasons for this pattern of results. For one, the degree of phonological similarity was different between experiments: intelligible (Dutch) background speech uses the same phonological inventory as the words the speakers had to produce, whereas the unintelligible (Chinese) background speech only partially overlapped with the Dutch phonological inventory. This is consistent with the phonological disruption view (Salamé & Baddeley, 1987), which states that the interference elicited by background speech is phonological in nature. Additionally, the access to semantic information was different between experiments: processing the meaning of background speech was possible for intelligible (Dutch) but not for the unintelligible (Chinese) background speech. This could have led to interference at the conceptual/semantic or syntactic levels. In short, stronger interference from intelligible (Dutch) than unintelligible (Chinese) background speech supports the domain-specific interference-by-similarity account, which assumes that interference is caused by shared representations/processes between focal cognitive tasks and background speech (e.g., Jones et al.,1993; Martin et al., 1988; Salamé & Baddeley, 1982, 1989).

6.4 Speakers can control the processing of background speech

Two studies in the present dissertation demonstrated that speakers have the ability to control the processing of irrelevant background speech. Specifically, Experiment 2 in Chapter 2 demonstrated impaired speech production in the divided-attention condition relative to the focused-attention condition, showing that participants were hindered more by picture naming while attending to background speech for a later memory test than while ignoring background speech. This suggests that deliberate attention allocation, due to the demand of the background speech task, interferes with speech production performance.

More interestingly, Chapter 5 compared the influence of funny versus boring background sentences on speech production, and found that funny sentences elicited less disruption than boring sentences. The reversed pattern of disruption was surprising, and the interpretation of this finding is complicated by the fact that the two types of irrelevant background sentences differed in many ways such as semantic/phonological information, prosodic pattern etc. Nevertheless, the results suggest that speakers may have increased top-down cognitive control to manage the disruption when background speech was engaging. That is, they may have realized the great potential for attention capture in funny sentences due to their variation in stimulus-aspecific properties and specific linguistic information, and have concentrated harder on speech production, leading to less interference.

Chapter 5 also compared the impact of boring sentences appearing in constant blocks (boring sentences) and varied blocks (boring sentences intermixed with funny sentences). In this comparison, the linguistic properties of the boring sentences were kept constant and only the interestingness of the entire block of sentences was varied. The results showed that boring sentences in the constant context caused more interference with speech production than in the varied context. While the practice effects cannot be ruled out because the varied block always followed the constant block, the results do still reflect the involvement of top-down cognitive control, namely, once speakers realize background speech varies, they focus on the speech production task harder to escape potential disruption.

6.5 Modulation of the irrelevant speech effect by production difficulty

This dissertation also explored whether the irrelevant speech effect was modulated by the difficulty of speech production by manipulating the lexical selection demands indexed by name agreement (high versus low). Chapter 2 showed that the interference elicited by representational similarity and attention demand was eliminated when lexical selection demand was high (i.e., when naming low name agreement pictures). A similar pattern of results was found in Chapter 4: producing low name agreement pictures shielded from disruption by intelligible background speech (relative to a quiet control condition), but not by unintelligible background speech. These findings are consistent with the attention engagement account

(Halin et al., 2014; Marsh et al., 2015): difficult tasks can become immune to disruption because of the concentration they require.

However, Chapter 5 showed a reversed pattern: the effects of interestingness and context were magnified for low compared to high name agreement pictures. This finding is not in line with the predictions of the attention engagement account (Halin et al., 2014; Marsh et al., 2015). Namely, these effects should be reduced or eliminated for low name agreement pictures. Instead, it is consistent with the load theory of attention (Lavie, 2005; Lavie & Dalton, 2014). This theory claims that increasing load on executive control functions renders individuals unable to actively maintain stimulus-processing priorities throughout task performance, increasing interference by irrelevant background distractors.

The discrepancy in the results seen in Chapters 2 and 4 versus Chapter 5 requires an explanation that takes the properties of the background speech into account. When background speech is simple and relatively uninteresting (e.g., has a regular acoustic pattern or simple syntactic structure), interference is reduced or eliminated by high lexical selection demands via an attention engagement mechanism (Halin et al., 2014; Marsh et al., 2015). By contrast, when background speech is varied and interesting, interference is increased when the lexical selection demand of speech production is high. Therefore, these findings highlight that lexical selection demands modulate interference differently according to the type of background speech.

Combined, the evidence from these chapters (Chapter 2, 4 & 5) paints an interesting picture for the irrelevant speech effects in speaking. Different properties of irrelevant background speech have different disruptive potentials on speech production. For example, representational similarity and attention demand in Chapter 2; stimulus-specific variation and intelligibility in Chapter 4; and interestingness and contextual variation in Chapter 5. This disruption is caused via domain-specific interference-by-similarity (mainly similarity in phonology) and domain-general attention capture (mainly specific attention capture) mechanisms. Speakers use compensatory mechanisms (e.g., increased top-down cognitive control) to avoid interference when background speech is funny and varied. Moreover, the processing of irrelevant background speech is modulated by the difficulty of speech production. Some of the interference elicited by background speech can be reduced or eliminated via increased attention engagement (Halin et al., 2014; Marsh et al., 2015) in response to difficult speech production, up until the influence of background speech increases due to difficult speech production imposing a high load on excutive control and making speakers unable to actively maintain speech planning (Lavie, 2005; Lavie & Dalton, 2014).

6.6 Recommendations for future research

The findings reported in this dissertation have addressed some of the questions related to speaking in noise. For instance, they addressed how spoken language production is influenced by different types of irrelevant background speech, and how this influence is modulated by the difficulty of speech production. This dissertation also provides some insights into how to conduct spoken language production research online. However, answers to these questions result in new questions. There are many avenues for future research to build upon the present work.

First, this dissertation shows that irrelevant background speech influences speech production via domain-specific interference-by-similarity and domain-general attention capture mechanisms. However, more research should be conducted to explore how speech production is affected by irrelevant background speech with multiple properties such as varying semantic/phonological information, syntactic structure, prosodic patterns, speaking speed, speakers' accents, and so on. This line of investigation would improve our understanding of how speakers plan their speech in the presence of irrelevant background speech, and add evidence to enrich current theories of irrelevant speech effects.

Second, the present dissertation focuses on a simple kind of speech production, namely, lists of picture names, and demonstrates that irrelevant background speech influences production performance, modulated by lexical selection demand. This raises a question of whether these findings can be generalized to larger units of speech such as phrase or sentence production. Future studies, therefore, should utilize more demanding tasks such as phrase, sentence, or dialogue production in the presence of background speech. Because interlocutors often communicate with each other by using complex linguistic structures (e.g., phrases or sentences), investigating the influence of irrelevant background speech on this kind of production would reveal how speakers plan their speech in real-world settings.

Third, this dissertation suggests that the direction of the modulation of irrelevant speech effects by speech production difficulty depends on the properties of the background speech. Auditory disruption is reduced in difficult, compared to easier picture naming, when background speech is simple and relatively uninteresting but is increased in difficult picture naming when background speech is varied and interesting. This pattern gives rise to some open questions. In particular, one may ask which properties of the background speech and of the speech planning task determine whether increasing the production difficulty leads to an increase or a reduction of the impact of the background speech. Thus, more research is required to investigate how interference by irrelevant background speech is modulated by the difficulty of spoken language production. This investigation would inform theories of the coordination of speech production and (involuntary) speech comprehension.

A fourth topic that should be explored further pertains to manipulations of speech production difficulty. The present dissertation manipulated the difficulty of lexical selection by varying name agreement and found that lexical selection demands modulated irrelevant speech effects. However, it is unclear whether other manipulations of lexical selection difficulty, such as semantic relationships between production and background speech (related versus unrelated), also affect irrelevant speech effects. It is also unclear whether the same pattern would arise for difficulty modulated by other processes involved in speech production such as object recognition (e.g., comparing intact and blurry pictures), phonological encoding (e.g., manipulating word frequency of picture names), or phonetic encoding (e.g., varying word length of picture names). Thus, more research looking beyond the name agreement manipulation would provide important insights into how auditory interference is affected by the difficulty of certain processes in speech production.

Finally, this dissertation focused on young adults and showed that they use some compensatory mechanisms—namely, attention engagement and increased cognitive control—to shield against the interference elicited by irrelevant background speech. The compensatory mechanism may vary across participant populations resulting in different patterns of auditory disruption from background speech. For example, children or older adults may show larger auditory disruption than young adults due to their poorer attentional control and their weaker ability to filter out task-irrelevant stimuli (e.g., Dulaney & Rogers, 1994; Kray et al., 2005). Therefore, a final future direction is to explore how background speech affects speech production in different populations such as children or older people. This investigation would contribute to a better understanding of how domain-general cognitive abilities, such as attentional control ability, play a role in speech production against background noise.

6.7 Conclusions

Given that spoken communication often takes place in noisy environments, it is essential to understand how background noise affects not only speech comprehension but also speech production. This dissertation focused on the question of how spoken language production is affected by irrelevant background speech, and whether this influence is modulated by the difficulty of speech production. I found that various properties of irrelevant background speech (e.g., representational similarity, attention demand, intelligibility, interestingness, and contextual variation) have different disruptive potentials on speech production. Disruption is caused by both domain-specific interference-by-similarity (mainly similarity in phonology) and domain-general attention capture (mainly specific attention capture) mechanisms. Speakers also use compensatory mechanisms (e.g., increased top-down cognitive control) to avoid interference when background speech is funny and varied. Combined, the results of this dissertation highlight that the potential disruption of spoken language production by irrelevant background speech is caused by several underlying mechanisms and strategies that speakers use to shield against it. Thus, the present dissertation provides important insights into how speakers plan and produce utterances in the presence of background speech, and contributes to our understanding of the coordination of deliberate spoken language production and involuntary comprehension.

Research Data Statement

All stimuli, data, and the scripts for statistical analyses supporting the studies in this thesis are openly available from [OSF] of Jieying He. Below are the links for each chapter:

Chapter 2: https://osf.io/zxy9d/

Chapter 3: https://osf.io/6jg4p/

Chapter 4: https://osf.io/wuafh/

Chapter 5: https://osf.io/kz4gh/

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. https://doi.org/10.1080/01690960802597250

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers, 36*(1), 140–155. https://doi.org/10.3758/BF03195559

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior research methods, 53*(4), 1407–1425. https://doi.org/10.3758/s13428-020-01501-5

Araújo, S., Huettig, F., & Meyer, A. (2020). What underlies the deficit in rapid automatized naming (RAN) in adults with dyslexia? Evidence from eye movements. *Scientific Studies of Reading*. *25*(6), 534–549. https://doi.org/10.1080/10888438.2020.1867863

Aydelott, J., Jamaluddin, Z., & Nixon Pearce, S. (2015). Semantic processing of unattended speech in dichotic listening. *The Journal of the Acoustical Society of America*, *138*(2), 964–975. https://doi.org/10.1121/1.4927410

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience*, *4*(10), 829–839. https://doi.org/10.1038/nrn1201

Badri, R., Siegel, J. H., & Wright, B. A. (2011). Auditory filter shapes and high-frequency hearing in adults who have impaired speech in noise performance despite clinically normal audiograms. *The Journal of the Acoustical Society of America*, *129*(2), 852–863. https://doi.org/10.1121/1.3523476

Ball, L. J., Threadgold, E., Solowiej, A., & Marsh, J. E. (2018). Can intrinsic and extrinsic metacognitive cues shield against distraction in problem solving? *Journal of cognition*, *1*(1), 15–15. https://doi.org/10.5334/joc.9

Banbury, S., & Berry, D. C. (1998). Disruption of office-related tasks by speech and office noise. *British journal of psychology*, *89*(3), 499–517. https://doi.org/10.1111/j.2044-8295.1998.tb02699.x

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. https://doi.org/10.1016/j.jml.2012.11.001

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart Pictures: Effects of age of acquisition, frequency, and name agreement. *The Quarterly Journal of Experimental Psychology Section A*, *50*(3), 560–585. https://doi.org/10.1080/783663595

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

Belke, E. (2017). The role of task-specific response strategies in blocked-cyclic naming. *Frontiers in psychology, 7*, Article 1955. https://doi.org/10.3389/fpsyg.2016.01955

Belke, E., & Meyer, A. S. (2007). Single and multiple object naming in healthy ageing. *Language and Cognitive Processes*, *22*(8), 1178–1211. https://doi.org/10.1080/01690960701461541

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology Section A*, *58*(4), 667–692. https://doi.org/10.1080/02724980443000142

Bell, R., Dentale, S., Buchner, A., & Mayr, S. (2010). ERP correlates of the irrelevant sound effect. *Psychophysiology, 47*(6), 1182–1191. https://doi.org/10.1111/j.1469-8986.2010.01029.x

Bell, R., Röer, J. P., Dentale, S., & Buchner, A. (2012). Habituation of the irrelevant sound effect: Evidence for an attentional theory of short-term memory disruption. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1542–1557. https://doi.org/10.1037/a0028459

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science, 31*(5), 733–764. https://doi.org/10.1080/03640210701530748

Boersma, P., & Weenink, D. (2009). *Praat: doing phonetics by computer* (Version 5.1.05) [Computer program]. http://www.praat.org/

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology*, *93*(1), 89–114. https://doi.org/10.1348/000712602162463

Boyle, R., & Coltheart, V. (1996). Effects of irrelevant sounds on phonological coding in reading comprehension and short term memory. *The Quarterly Journal of Experimental Psychology Section A*, *49*(2), 398–416. https://doi.org/10.1080/713755630

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*:e9414. https://doi.org/10.7717/peerj.9414

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica*, *86*(1), 117–128.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 441–458. https://doi.org/10.1037/xhp0000159

Buchner, A., Rothermund, K., Wentura, D., & Mehl, B. (2004). Valence of distractor words increases the effects of irrelevant speech on serial recall. *Memory & Cognition*, *32*(5), 722–731. https://doi.org/10.3758/BF03195862

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. In B. Bel & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 Conference* (pp. 199–202). Aix-en-Provence, France: Laboratoire Parole et Langage.

Cauchard, F., Cane, J. E., & Weger, U. W. (2012). Influence of background speech and music in interrupted reading: An eye-tracking study. *Applied Cognitive Psychology*, *26*(3), 381–390. https://doi.org/10.1002/acp.1837

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*(6), 2156–2160. https://doi.org/10.1016/j.chb.2013.05.009

Chein, J. M., & Fiez, J. A. (2010). Evaluating models of working memory through the effects of concurrent irrelevant information. *Journal of Experimental Psychology: General*, *139*(1), 117–137. https://doi.org/10.1037/a0018200

Cheng, X., Schafer, G., & Akyürek, E. G. (2010). Name agreement in picture naming: An ERP study. *International Journal of Psychophysiology*, *76*(3), 130–141. https://doi.org/10.1016/j.ijpsycho.2010.03.003

Clark, H. H. (1996). *Using language*. Cambridge university press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

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. https://doi.org/10.1037/0096-1523.32.1.104

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*, *15*(1), 17–31. https://doi.org/https://doi.org/10.1016/S0022-5371(76)90003-7

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. https://doi.org/10.1037/0278-7393.34.4.886

Cooke, M., Barker, J., Lecumberri, M. L. G., & Wasilewski, K. (2011). Crowdsourcing for word recognition in noise. In P. Cosi, R. De Mori, G. Di Fabbrizio, & R. Pieraccini (Eds.), *Proceedings of Interspeech 2011* (pp. 3049–3052). Grenoble, France: International Speech Communication Association.

Cowan, N. (1995). Verbal Working Memory: A View with a Room. [Working Memory and Language, S. E. Gathercole &A. D. Baddeley]. *The American Journal of Psychology*, *108*(1), 123–129. https://doi.org/10.2307/1423105

Cowan, N. (1999). An Embedded-Processes Model of working memory (*Models of working memory: Mechanisms of active maintenance and executive control.* (pp. 62–101). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.006

Cuetos, F., Ellis, A. W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, Instruments, & Computers*, *31*(4), 650–658. https://doi.org/10.3758/BF03200741

Cullington, H. E., & Zeng, F. G. (2008). Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects. *The Journal of the Acoustical Society of America*, *123*(1), 450–461. https://doi.org/10.1121/1.2805617

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 416–431. https://doi.org/10.1037/0278-7393.29.3.416

Damian, M. F., & Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of object names. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1372–1384. https://doi.org/10.1037/0278-7393.31.6.1372

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. https://doi.org/10.1037/0278-7393.25.2.345

Damian, M. F., Vigliocco, G., & Levelt, W. J. M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, *81*(3), B77–B86. https://doi.org/10.1016/S0010-0277(01)00135-4

de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, *48*(1), 1–12. https://doi.org/10.3758/s13428-015-0567-2

Dell′Acqua, R., Lotto, L., & Job, R. (2000). Naming times and standardized norms for the Italian PD/DPSS set of 266 pictures: Direct comparisons with American, English, French, and Spanish published databases. *Behavior Research Methods, Instruments, & Computers*, *32*(4), 588–615. https://doi.org/10.3758/BF03200832

Denckla, M. B., & Rudel, R. G. (1976). Rapid ‘automatized’ naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, *14*(4), 471–479. https://doi.org/10.1016/0028-3932(76)90075-0

Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(2), 470–484. https://doi.org/10.1037/0278-7393.20.2.470

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Dupoux, E., Kouider, S., & Mehler, J. (2003). Lexical access without attention? Explorations using dichotic priming. *Journal of Experimental Psychology: Human perception and performance*, *29*(1), 172–184. https://doi.org/10.1037/0096-1523.29.1.172

Duquesnoy, A. J., & Plomp, R. (1980). Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis. *The Journal of the Acoustical Society of America*, *68*(2), 537–544. https://doi.org/10.1121/1.384767

Egan, J. J. (1972). Psychoacoustics of the Lombard voice response. *Journal of Auditory Research*. *12*(4), 318–324.

Eckert, M. A., Teubner-Rhodes, S., & Vaden, K. I., Jr. (2016). Is Listening in Noise Worth It? The Neurobiology of Speech Recognition in Challenging Listening Conditions. *Ear and hearing*, *37 Suppl 1*(Suppl 1), 101S–110S. https://doi.org/10.1097/AUD.0000000000000300

Eggemeier, F. T., Crabtree, M. S., & LaPointe, P. A. (1983). The Effect of Delayed Report on Subjective Ratings of Mental Workload. *Proceedings of the Human Factors Society Annual Meeting*, *27*(2), 139–143. https://doi.org/10.1177/154193128302700205

Eimer, M., Nattkemper, D., Schröger, E., & Prinz, W. (1996). Chapter 5 Involuntary attention. In O. Neumann & A. F. Sanders (Eds.), *Handbook of Perception and Action* (Vol. 3, pp. 155–184). Academic Press. https://doi.org/https://doi.org/10.1016/S1874-5822(96)80022-3

Elliott, E. M., & Briganti, A. M. (2012). Investigating the role of attentional resources in the irrelevant speech effect. *Acta Psychologica*, *140*(1), 64–74. https://doi.org/https://doi.org/10.1016/j.actpsy.2012.02.009

Fairs, A. (2019). *Linguistic dual-tasking: Understanding temporal overlap between production and comprehension* [Doctoral dissertation, Radboud University]. Radboud Repository. https://hdl.handle.net/2066/203906

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

Fairs, A., & Strijkers, K. (2021). Can we use the internet to study speech production? Yes we can! Evidence contrasting online versus laboratory naming latencies and errors. *Plos one*, *16*(10), e0258908. https://doi.org/10.1371/journal.pone.0258908

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

Fargier, R., & Laganaro, M. (2019). Interference in speaking while hearing and vice versa. *Scientific reports*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-41752-7

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. https://doi.org/10.1037/0278-7393.28.6.1187

Ferreira, F., & Swets, B. (2017). The production and comprehension of resumptive pronouns in relative clause “island” contexts. In A. Cutler (Ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones* (pp. 263–278). Routledge.

Fischer, R., & Plessow, F. (2015). Efficient multitasking: parallel versus serial processing of multiple tasks. *Frontiers in Psychology*, *6*, Article 1366. https://doi.org/10.3389/fpsyg.2015.01366

Garrett, M. F. (1975). The analysis of sentence production. *Psychology of learning and motivation, 9*, 133–177. https://doi.org/10.1016/S0079-7421(08)60270-4

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. https://doi.org/10.1080/016909697386646

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. https://doi.org/10.1037/0096-1523.10.5.640

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, *15*(2), 103–113. https://doi.org/10.1177/002383097201500201

Goodwin, C. (1981). Conversational organization. *Interaction between speakers and hearers*.

Gordon-Salant, S., & Fitzgibbons, P. J. (2004). Effects of stimulus and noise rate variability on speech perception by younger and older adults. *The Journal of the Acoustical Society of America*, *115*(4), 1808–1817. https://doi.org/10.1121/1.1645249

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279. https://doi.org/10.1111/1467-9280.00255

Halin, N., Marsh, J. E., Haga, A., Holmgren, M., & Sörqvist, P. (2014). Effects of speech on proofreading: can task-engagement manipulations shield against distraction? *Journal of Experimental Psychology: Applied*, *20*(1), 69–80. https://doi.org/10.1037/xap0000002

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. https://doi.org/10.1111/sjop.12246

He, J., Meyer, A. S., & Brehm, L. (2021). Concurrent listening affects speech planning and fluency: the roles of representational similarity and capacity limitation. *Language, Cognition and Neuroscience*, *36*(10), 1258–1280. https://doi.org/10.1080/23273798.2021.1925130

He, J., Meyer, A. S., Creemers, A., & Brehm, L. (2021). Conducting language production research online: A web-based study of semantic context and name agreement effects in multi-word production. *Collabra: Psychology*, *7*(1), 29935. https://doi.org/10.1525/collabra.29935

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

Hellbrück, J., & Liebl, A. (2008). Noise effects on cognitive performance. In Kuwano & Sonoko (Eds), *Recent topics in environmental psychoacoustics* (pp. 153–184). Osaka University Press.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition*, *100*(3), 464–482. https://doi.org/10.1016/j.cognition.2005.02.006

Hughes, R. W. (2014). Auditory distraction: A duplex-mechanism account. *PsyCh Journal*, *3*(1), 30–41. https://doi.org/10.1002/pchj.44

Hughes, R. W., Hurlstone, M. J., Marsh, J. E., Vachon, F., & Jones, D. M. (2013). Cognitive control of auditory distraction: Impact of task difficulty, foreknowledge, and working memory capacity supports duplex-mechanism account. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 539–553. https://doi.org/10.1037/a0029064

Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1050–1061. https://doi.org/10.1037/0278-7393.33.6.1050

Humes, L. E., & Roberts, L. (1990). Speech-recognition difficulties of the hearing-impaired elderly: The contributions of audibility. *Journal of Speech, Language, and Hearing Research*, *33*(4), 726–735. https://doi.org/10.1044/jshr.3304.726

Hyönä, J., & Ekholm, M. (2016). Background Speech Effects on Sentence Processing during Reading: An Eye Movement Study. *PLoS One*, *11*(3), e0152133. https://doi.org/10.1371/journal.pone.0152133

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. https://doi.org/10.1080/14640749208401304

Jones, D., & Morris, N. (1992). Irrelevant speech and serial recall: Implications for theories of attention and working memory. *Scandinavian Journal of Psychology*, *33*(3), 212–229. https://doi.org/10.1111/j.1467-9450.1992.tb00911.x

Jones, D. M., Macken, W. J., & Murray, A. C. (1993). Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Memory & Cognition*, *21*(3), 318–328. https://doi.org/10.3758/BF03208264

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. https://doi.org/10.1080/17470218.2014.964736

Jun, E., Hsieh, G., & Reinecke, K. (2017). Types of motivation affect study selection, attention, and dropouts in online experiments. *Proceedings of the ACM on Human-Computer Interaction*, *1*(CSCW), 1–15. https://doi.org/10.1145/3134691

Kandel, M., Wyatt, C., & Phillips, C. (2021, March 4–6). *Transitioning to online language production: a direct comparison of in-lab and web-based experiments* [Conference presentation abstract]. The 34th Annual CUNY Conference on Human Sentence Processing, Philadelphia, United States. https://www.cuny2021.io/wp-content/uploads/2021/02/CUNY_2021_abstract_131.pdf

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, N.J.: Prentice-Hall.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, *14*(2), 185–209. https://doi.org/10.1016/0010-0277(83)90029-X

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. https://doi.org/10.1016/j.jml.2015.08.001

Konopka, A. E. (2019). Encoding actions and verbs: Tracking the time-course of relational encoding during message and sentence formulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(8), 1486–1510. https://doi.org/10.1037/xlm0000650

Kray, J., Eppinger, B. E. N., & Mecklinger, A. (2005). Age differences in attentional control: An event-related potential approach. *Psychophysiology*, *42*(4), 407–416. https://doi.org/10.1111/j.1469-8986.2005.00298.x

Kristensen, L. B., Wang, L., Petersson, K. M., & Hagoort, P. (2013). The interface between language and attention: Prosodic focus marking recruits a general attention network in spoken language comprehension. *Cerebral Cortex*, *23*(8), 1836–1848. https://doi.org/10.1093/cercor/bhs164

Lane, H., Tranel, B., & Sisson, C. (1970). Regulation of voice communication by sensory dynamics. *The Journal of the Acoustical Society of America*, *47*(2B), 618–624. https://doi.org/10.1121/1.1911937

Lange, E. B. (2005). Disruption of attention by irrelevant stimuli in serial recall. *Journal of Memory and Language*, *53*(4), 513–531. https://doi.org/10.1016/j.jml.2005.07.002

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75–82. https://doi.org/https://doi.org/10.1016/j.tics.2004.12.004

Lavie, N., & Dalton, P. (2014). Load theory of attention and cognitive control. In A. C. Nobre, K. Nobre, & S, Kastner (Eds), *The Oxford handbook of attention* (pp. 56–75). Oxford University Press.

Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of experimental psychology: General, 133*(3), 339–354. https://doi.org/10.1037/0096-3445.133.3.339

Lee, E.-K., Brown-Schmidt, S., & Watson, D. G. (2013). Ways of looking ahead: Hierarchical planning in language production. *Cognition*, *129*(3), 544–562. https://doi.org/10.1016/j.cognition.2013.08.007

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: 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*(1), 1–38. https://doi.org/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. https://doi.org/https://doi.org/10.1016/j.tics.2015.10.010

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. https://doi.org/10.1037/0096-1523.34.3.751

Macken, W. J., Phelps, F. G., & Jones, D. M. (2009). What causes auditory disrtaction?. *Psychonomic Bulletin & Review*, *16*(1), 139–144. https://doi.org/10.3758/PBR.16.1.139

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. https://doi.org/10.1037/a0022330

Marsh, J. E., Hughes, R. W., & Jones, D. M. (2008). Auditory distraction in semantic memory: A process-based approach. *Journal of Memory and Language*, *58*(3), 682–700. https://doi.org/10.1016/j.jml.2007.05.002

Marsh, J. E., Hughes, R. W., & Jones, D. M. (2009). Interference by process, not content, determines semantic auditory distraction. *Cognition*, *110*(1), 23–38. https://doi.org/10.1016/j.cognition.2008.08.003

Marsh, J. E., Ljung, R., Jahncke, H., MacCutcheon, D., Pausch, F., Ball, L. J., & Vachon, F. (2018). Why are background telephone conversations distracting?. *Journal of Experimental Psychology: Applied*, *24*(2), 222–235. https://doi.org/10.1037/xap0000170

Marsh, J. E., Sörqvist, P., & Hughes, R. W. (2015). Dynamic cognitive control of irrelevant sound: Increased task engagement attenuates semantic auditory distraction. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(5), 1462–1474. https://doi.org/10.1037/xhp0000060

Martin, R. C., Wogalter, M. S., & Forlano, J. G. (1988). Reading comprehension in the presence of unattended speech and music. *Journal of Memory and Language*, *27*(4), 382–398. https://doi.org/https://doi.org/10.1016/0749-596X(88)90063-0

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McLeod, P. (1977). A dual task response modality effect: Support for multiprocessor models of attention. *Quarterly Journal of Experimental Psychology*, *29*(4), 651–667. https://doi.org/10.1080/14640747708400639

Meyer, A., Wheeldon, L. R., & Konopka, A. (2012). Effects of speech rate and practice on the allocation of visual attention in multiple object naming. *Frontiers in psychology, 3,* Article 39. https://doi.org/10.3389/fpsyg.2012.00039

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. https://doi.org/10.1016/j.cognition.2008.08.002

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*, *9,* Article 86. https://doi.org/10.3389/fnhum.2015.00086

Momma, S., & Ferreira, V. S. (2019). Beyond linear order: The role of argument structure in speaking. *Cognitive psychology, 114,* Article 101228. https://doi.org/10.1016/j.cogpsych.2019.101228

Moore, B. C. (2012). *An introduction to the psychology of hearing*. Brill.

Mortensen, L., Meyer, A. S., & Humphreys, G. W. (2008). Speech planning during multiple-object naming: Effects of ageing. *Quarterly Journal of Experimental Psychology*, *61*(8), 1217–1238. https://doi.org/10.1080/17470210701467912

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological review*, *86*(3), 214–255. https://doi.org/10.1037/0033-295X.86.3.214

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. https://doi.org/10.1037/0096-1523.13.3.435

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational ideas—Part II. *Language and Linguistics Compass*, *10*(11), 591–613. https://doi.org/10.1111/lnc3.12207

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*(2), 227–252. https://doi.org/10.1016/j.cognition.2009.09.007

Oswald, C. J. P., Tremblay, S., & Jones, D. M. (2000). Disruption of comprehension by the meaning of irrelevant sound. *Memory*, *8*(5), 345–350. https://doi.org/10.1080/09658210050117762

Parmentier, F. B. R., & Beaman, C. P. (2015). Contrasting effects of changing rhythm and content on auditory distraction in immediate memory. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *69*(1), 28–38. https://doi.org/10.1037/cep0000036

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

Pittman, A. L., & Wiley, T. L. (2001). Recognition of speech produced in noise. *Journal of Speech, Language, and Hearing Research, 44* (3), 487–496. https://doi.org/10.1044/1092-4388(2001/038)

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual review of neuroscience*, *13*(1), 25–42. https://doi.org/10.1146/annurev.ne.13.030190.000325

Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology*, *58*(1), 1–23. https://doi.org/10.1146/annurev.psych.58.110405.085516

Quist-Hanssen, S. V., Thorud, E., & Aasand, G. (1978). Noise-induced hearing loss and the comprehension of speech in noise. *Acta Oto-Laryngologica*, *86*(sup360), 90–95. https://doi.org/10.3109/00016487809123483

R Core Team (2018). *R: A language and environment for statistical computing* (Version 3.6.1) [computer software]. http://www.R-project.org.

R Core Team (2020). *R: A language and environment for statistical computing* (Version 4.0.3) [computer software]. http://www.R-project.org

Reimers, S., & Stewart, N. (2015). Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript Web experiments. *Behavior Research Methods*, *47*(2), 309–327. https://doi.org/10.3758/s13428-014-0471-1

Rivenez, M., Darwin, C. J., & Guillaume, A. (2006). Processing unattended speech. *The Journal of the Acoustical Society of America*, *119*(6), 4027–4040. https://doi.org/10.1121/1.2190162

Rivenez, M., Guillaume, A., Bourgeon, L., & Darwin, C. J. (2008). Effect of voice characteristics on the attended and unattended processing of two concurrent messages.

European Journal of Cognitive Psychology, *20*(6), 967–993. https://doi.org/10.1080/09541440701686201

Robison, M. K., & Unsworth, N. (2015). Working memory capacity offers resistance to mind-wandering and external distraction in a context-specific manner. *Applied Cognitive Psychology*, *29*(5), 680–690. https://doi.org/10.1002/acp.3150

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

Roelofs, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: Verb–particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(4), 904–921. https://doi.org/10.1037/0278-7393.24.4.904

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological review*, *110*(1), 88–125. https://doi.org/10.1037/0033-295X.110.1.88

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

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

Roelofs, A., & Piai, V. (2011). Attention Demands of Spoken Word Planning: A Review *Frontiers in Psychology*, *2*, Article 307. https://doi.org/10.3389/fpsyg.2011.00307

Röer, J. P., Bell, R., & Buchner, A. (2013). Self-relevance increases the irrelevant sound effect: Attentional disruption by one's own name. *Journal of Cognitive Psychology*, *25*(8), 925–931. https://doi.org/10.1080/20445911.2013.828063

Röer, J. P., Bell, R., & Buchner, A. (2014). Evidence for habituation of the irrelevant-sound effect on serial recall. *Memory & Cognition*, *42*(4), 609–621. https://doi.org/10.3758/s13421-013-0381-y

Röer, J. P., Bell, R., & Buchner, A. (2015). Specific foreknowledge reduces auditory distraction by irrelevant speech. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(3), 692–702. https://doi.org/10.1037/xhp0000028

Röer, J. P., Bell, R., Marsh, J. E., & Buchner, A. (2015). Age equivalence in auditory distraction by changing and deviant speech sounds. *Psychology and Aging*, *30*(4), 849–855. https://doi.org/10.1037/pag0000055

Ruthruff, E., Pashler, H. E., & Hazeltine, E. (2003). Dual-task interference with equal task emphasis: Graded capacity sharing or central postponement? *Perception & Psychophysics*, *65*(5), 801–816. https://doi.org/10.3758/BF03194816

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, *21*(2), 150–164. https://doi.org/https://doi.org/10.1016/S0022-5371(82)90521-7

Salamé, P., & Baddeley, A. (1987). Noise, unattended speech and short-term memory. *Ergonomics*, *30*(8), 1185–1194. https://doi.org/10.1080/00140138708966007

Salamé, P., & Baddeley, A. (1989). Effects of Background Music on Phonological Short-Term Memory. *The Quarterly Journal of Experimental Psychology Section A*, *41*(1), 107–122. https://doi.org/10.1080/14640748908402355

Sauter, M., Draschkow, D., & Mack, W. (2020). Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain Sciences*, *10*(4), 251. https://doi.org/10.3390/brainsci10040251

Schlittmeier, S. J., Weißgerber, T., Kerber, S., Fastl, H., & Hellbrück, J. (2012). Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength. *Attention, Perception, & Psychophysics*, *74*(1), 194–203. https://doi.org/10.3758/s13414-011-0230-7

Schnoebelen, T., & Kuperman, V. (2010). Using Amazon Mechanical Turk for linguistic research. *Psihologija*, *43*(4), 441–464. https://doi.org/10.2298/PSI1004441S

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

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

Sharp, D. J., Scott, S. K., Mehta, M. A., & Wise, R. J. (2006). The neural correlates of declining performance with age: evidence for age-related changes in cognitive control. *Cerebral Cortex*, *16*(12), 1739–1749. https://doi.org/10.1093/cercor/bhj109

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

Shao, Z., Roelofs, A., Acheson, D. J., & Meyer, A. S. (2014). Electrophysiological evidence that inhibition supports lexical selection in picture naming. *Brain Research*, *1586*, 130–142. https://doi.org/10.1016/j.brainres.2014.07.009

Shao, Z., Roelofs, A., Martin, R. C., & Meyer, A. S. (2015). Selective inhibition and naming performance in semantic blocking, picture-word interference, and color–word Stroop tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1806–1820. https://doi.org/10.1037/a0039363

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. https://doi.org/10.1016/j.cognition.2014.10.008

Smiljanić, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. *The Journal of the Acoustical Society of America*, *130*(6), 4020–4031. https://doi.org/10.1121/1.3652882

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, *28*(4), 516–536. https://doi.org/10.3758/BF03200540

Sokolov, E. N. (1963). *Perception and the conditioned reflex.* Macmillan.

Sörqvist, P., & Marsh, J. E. (2015). How Concentration Shields Against Distraction. *Current Directions in Psychological Science*, *24*(4), 267–272. https://doi.org/10.1177/0963721415577356

Sörqvist, P., Nöstl, A., & Halin, N. (2012). Disruption of writing processes by the semanticity of background speech. *Scandinavian Journal of Psychology*, *53*(2), 97–102. https://doi.org/10.1111/j.1467-9450.2011.00936.x

Sörqvist, P., & Rönnberg, J. (2014). Individual differences in distractibility: An update and a model. *PsyCh Journal*, *3*(1), 42–57. https://doi.org/10.1002/pchj.47

Sperry, J. L., Wiley, T. L., & Chial, M. R. (1997). Word recognition performance in various background competitors. *Journal-American Academy of Audiology*, *8*, 71–80.

Stark, K., van Scherpenberg, C., Obrig, H., & Abdel Rahman, R. (2022). Web-based language production experiments: Semantic interference assessment is robust for spoken and typed response modalities. *Behavior Research Methods*, 1-27. https://doi.org/10.3758/s13428-021-01768-2

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592. https://doi.org/10.1073/pnas.0903616106

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. https://doi.org/10.1111/1467-9280.00386

Sweeney, R. E., & Ulveling, E. F. (1972). A Transformation for Simplifying the Interpretation of Coefficients of Binary Variables in Regression Analysis. *The American Statistician*, *26*(5), 30–32. https://doi.org/10.1080/00031305.1972.10478949

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. https://doi.org/10.1037/0096-1523.29.1.3

Vachon, F., Labonté, K., & Marsh, J. E. (2017). Attentional capture by deviant sounds: A noncontingent form of auditory distraction?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 622–634. https://doi.org/10.1037/xlm0000330

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*(4), 584–589. https://doi.org/10.3758/BF03193889

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, *121*(1), 519–526. https://doi.org/10.1121/1.2400666

van Paridon, J., & Thompson, B. (2021). subs2vec: Word embeddings from subtitles in 55 languages. *Behavior Research Methods*, *53*, 629–655. https://doi.org/10.3758/s13428-020-01406-3

Vasilev, M. R., Kirkby, J. A., & Angele, B. (2018). Auditory Distraction During Reading: A Bayesian Meta-Analysis of a Continuing Controversy. *Perspectives on Psychological Science*, *13*(5), 567–597. https://doi.org/10.1177/1745691617747398

Vasilev, M. R., Liversedge, S. P., Rowan, D., Kirkby, J. A., & Angele, B. (2019). Reading is disrupted by intelligible background speech: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(11), 1484–1512. https://doi.org/10.1037/xhp0000680

Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming. *The Quarterly Journal of Experimental Psychology Section A*, *48*(4), 822–848. https://doi.org/10.1080/14640749508401419

Vogt, A., Hauber, R., Kuhlen, A. K., & Rahman, R. A. (2022). Internet-based language production research with overt articulation: Proof of concept, challenges, and practical advice. *Behavior Research Methods*, *54*(4), 1954-1975. https://doi.org/10.3758/s13428-021-01686-3

Walker, M. B., & Trimboli, C. (1982). Smooth transitions in conversational interactions. *The Journal of Social Psychology*, *117*(2), 305–306. https://doi.org/10.1080/00224545.1982.9713444

Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, *37*(3), 356–381. https://doi.org/https://doi.org/10.1006/jmla.1997.2517

Withers, P. (2017). *Frinex: Framework for interactive experiments.* https://doi.org/10.5281/zenodo.3522911

Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 255–260. https://doi.org/10.1037/0278-7393.21.1.255

Yan, G., Meng, Z., Liu, N., He, L., & Paterson, K. B. (2018). Effects of irrelevant background speech on eye movements during reading. *Quarterly Journal of Experimental Psychology*, *71*(6), 1270–1275. https://doi.org/10.1080/17470218.2017.1339718

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*, *111*(4), 493–504. https://doi.org/10.1037/pspa0000056

Nederlandse Samenvatting

Conversaties bestaan in alle culturen en zijn de meest voorkomende manier om taal te gebruiken. In het dagelijks leven besteden veel mensen aanzienlijke tijd aan het uitwisselen van ideeën, gedachten en gevoelens met anderen via gesproken conversaties, wat het de meest natuurlijke en basale vorm van taalgebruik maakt. Een belangrijk kenmerk van gesproken conversaties is de regelmatige wisseling van beurten tussen gesprekspartners, oftewel, de rolwisseling tussen sprekers en luisteraars. Dit betekent dat gesproken conversaties vereisen dat er twee taken tegelijk gedaan moeten worden; namelijk spreken (productie) en luisteren (begrip). Bovendien vinden gesproken conversaties plaats in diverse fysieke omgevingen. Zo kunnen ze plaatsvinden in een stille kamer zonder externe bronnen van achtergrondgeluid. Ze kunnen echter ook plaatsvinden in lawaaierige omgevingen, zoals een druk café of een drukke trein, waar op de achtergrond een levendige discussie of een telefoongesprek kan worden opgevangen. Veel mensen geven er de voorkeur aan om gesprekken te voeren op rustige plekken, omdat ze vinden dat lawaaierige omgevingen hun communicatie afleiden en verstoren. Dit roept de vraag op hoe gesproken conversaties worden beïnvloed door irrelevante achtergrondgeluiden, waaronder non-verbale ruis (bijvoorbeeld verkeer of bouwwerkzaamheden) en verbale ruis (bijvoorbeeld achtergrondgesprekken of radio- en televisie-uitzendingen). Om deze vraag te kunnen beantwoorden, is het essentieel om te beschrijven hoe mensen hun eigen spraak plannen en hoe ze de spraak van hun gesprekpartners begrijpen als er irrelevante achtergrondruis aanwezig is. Met dit proefschrift heb ik dit doel bereikt door me te concentreren op de specifieke vraag hoe spraakproductie (spreken) beïnvloed werd door irrelevante achtergrondspraak (verbale achtergrondgeluiden). De tweede vraag in dit proefschrift had betrekking op de observatie dat mensen kunnen omgaan met verstoringen door achtergrondgeluid wanneer zij zich concentreren op cognitieve taken, zoals bij het werken op kantoor of bij het lezen. Met name bij het uitvoeren van moeilijke taken hebben mensen de neiging zich meer op de taak te concentreren en minder hinder te ondervinden van het achtergrondgeluid. Het tweede doel van dit proefschrift was dan ook om te onderzoeken hoe mensen omgaan met de invloed van irrelevante achtergrondspraak wanneer ze spraak plannen met wisselende moeilijkheidsgraad.

In hoofdstuk 2 van dit proefschrift onderzocht ik hoe de spraakproductie werd beïnvloed door verschillende soorten irrelevante achtergrondspraak (met wisselende overeenkomsten tussen hun eigen spraak en de achtergrondspraak - verder: ‘gelijkenis in representatie’) in experiment 1, en wisselende eisen van aandacht in experiment 2) en of deze invloed werd gereguleerd door de moeilijkheidsgraad van de spraakproductie. In experiment 1 benoemden deelnemers – allemaal moedertaalsprekers van het Nederlands - plaatjes met verschillende moeilijkheidsgraden in het Nederlands, terwijl ze Nederlandse woordlijsten moesten negeren (verstaanbare achtergrondspraak; hoge gelijkenis in representatie), Chinese woordlijsten (onverstaanbare achtergrondspraak; matige gelijkenis in representatie), of gebrabbel van acht sprekers door elkaar (taalachtige ruis; lage gelijkenis in representatie). Het bleek dat de prestaties van het benoemen door de deelnemers meer werden verstoord door achtergrondspraak (Nederlandse en Chinese woordlijsten) dan door taalachtige ruis (gebrabbel van acht sprekers), en dat de prestaties van het benoemen meer aangetast werden door verstaanbare (Nederlandse woordlijsten) dan onverstaanbare (Chinese woordlijsten) achtergrondspraak. In experiment 2 werden er door andere moedertaalsprekers van het Nederlands plaatjes benoemd met een wisselende moeilijkheidsgraad in het Nederlands, terwijl ze Nederlandse woordlijsten moesten negeren (weinig aandacht nodig), of terwijl ze aandacht moesten schenken aan de Nederlandse woordlijsten voor een geheugentaak later (veel aandacht nodig). Het bleek dat de prestaties op de spraakproductie slechter waren in de conditie waarbij er veel aandacht nodig was, in vergelijking met de conditie waar weinig aandacht werd gevraagd. Bovendien werden de verstoringen die veroorzaakt werden door de irrelevante achtergrondspraak in beide experimenten verminderd wanneer ze moeilijke plaatjes moesten benoemen. Deze bevindingen suggereren dat zowel een verhoogde gelijkenis in representatie als een verhoogde eis in aandacht van irrelevante achtergrondspraak meer verstoringen veroorzaken in de spraakproductie, maar dat de verstoringen beschermd kunnen worden wanneer er moeilijke spraakproductietaken uitgevoerd moeten worden.

Doordat vanwege de COVID-19-pandemie experimenten in het lab niet door konden gaan, heb ik in hoofdstuk 3 gekeken naar de haalbaarheid van het uitvoeren van experimenten op het gebied van spraakproductie in een online omgeving. Hiervoor werden twee klassieke bevindingen onderzocht die in experimenten in het lab waren uitgevoerd. Het eerste is dat het benoemen van een plaatje met een hoge benoemingsconsensus (een plaatje van een *appel* wordt bijna altijd *appel* genoemd) sneller gaat en nauwkeuriger is dan het benoemen van een plaatje met een lage benoemingsconsensus (een plaatje van een *beker* kan ook een *mok* of *kopje* genoemd worden). Dit wordt het ‘effect van benoemingsconsensus’ genoemd. De tweede bevinding is dat het benoemen van plaatjes langzamer gaat en foutgevoeliger is wanneer er plaatjes benoemd worden binnen dezelfde semantische categorie (homogene context; bijvoorbeeld *dolfijn, vlinder, muis,* leeuw), dan wanneer het plaatjes van verschillende semantische categorieën zijn (heterogene context; bijvoorbeeld *dolfijn, oor, handschoen, kam*). Dit wordt het ‘semantische contexteffect’ genoemd. Moedertaalsprekers van het Nederlands benoemden online plaatjes met gevarieerde benoemingsconsensus (hoog versus laag) en semantische context (homogeen versus heterogeen). De twee effecten werden gerepliceerd in dit online experiment, ter ondersteuning van de haalbaarheid van het uitvoeren van experimenten op het gebied van spraakproductie in een online omgeving.

In hoofdstuk 4 ging ik terug naar het onderzoek naar hoe spraakproductie werd beïnvloed door irrelevante achtergrondspraak. Dit omdat de irrelevante achtergrondspraak uit hoofdstuk 2 (Nederlandse woordlijsten, Chinese woordlijsten en gebrabbel van acht sprekers) verschilde op meerdere aspecten, zoals talige eigenschappen (semantiek en fonologie) en niet-talige eigenschappen (akoestisch patroon en gesegmenteerde versus doorlopende geluidsstroom). Daarom kan de verstoring niet alleen veroorzaakt worden door het verschil in gelijkenis in representatie (oftewel de taalkundige inhoud), maar ook door het akoestische patroon (oftewel niet-talige eigenschappen). In dit hoofdstuk heb ik dus onderzocht of het akoestische patroon (gesegmenteerde versus doorlopende woordlijsten versus zinnen) van onverstaanbare achtergrondspraak (Chinees) de spraakproductie beïnvloedde (experiment 1), en of deze invloed ook optrad wanneer de irrelevante achtergrondspraak verstaanbaar was (Nederlands; experiment 2). In beide experimenten benoemden moedertaalsprekers van het Nederlands plaatjes met variërende moeilijkheidsgraad in het Nederlands, terwijl ze woordlijsten of zinnen moesten negeren (Chinees in experiment 1, Nederlands in experiment 2), of in een stille conditie (zonder achtergrondspraak). Het bleek dat onverstaanbare (Chinese) en verstaanbare (Nederlandse) achtergrondspraak de spraakproductie verslechterde ten opzichte van de stille conditie, maar er werd geen verschil gevonden tussen de woordlijsten en zinnen in beide experimenten. Dit suggereert dat, ongeacht de verstaanbaarheid, de aanwezigheid van irrelevante achtergrondspraak spraakproductie verslechtert, maar dat het akoestische patroon van de achtergrondspraak de verstoring niet reguleert. Net als in hoofdstuk 2 werd de verstoring die veroorzaakt wordt door verstaanbare achtergrondspraak (Nederlands) verminderd wanneer ze moeilijke plaatjes moesten benoemen, wat aangeeft dat sprekers beter met de verstoring kunnen omgaan wanneer ze moeilijke spraak plannen.

Hoofdstuk 5 onderzocht verder hoe spraakproductie beïnvloed werd door verschillende soorten van irrelevante achtergrondzinnen. Omdat de irrelevante achtergrondspraak in hoofdstuk 4 (woordlijsten, zinnen) saai en simpel waren, was het wellicht makkelijker om deze te negeren, aangezien er geen verschillen ontstonden tussen de woordlijsten en de zinnen. Dit hoofdstuk gebruikte relatief grote verschillen in achtergrondzinnen met betrekking tot *interessantheid* (saai versus grappig) en context (constant versus gevarieerd). Moedertaalsprekers van het Nederlands benoemden plaatjes met verschillende moeilijkheidsgraad in het Nederlands terwijl ze de achtergrondzinnen moesten negeren in een conditie met constante context (alleen saaie zinnen), gevolgd door een conditie met een gevarieerde context (saaie en grappige zinnen). De grappige zinnen waren minder storend dan de saaie zinnen, en de saaie zinnen verstoorden de spraakproductie minder in de gevarieerde dan in de constante context. Bovendien kwam er een omgekeerd patroon van de hoofdstukken 2 en 4 naar voren; de verstoring veroorzaakt door achtergrondzinnen was groter voor moeilijke dan voor makkelijke plaatjes. Deze bevindingen suggereren dat de interessantheid en context van irrelevante achtergrondspraak de spraakproductie beïnvloeden, die op zijn beurt wordt gereguleerd door de moeilijkheidsgraad van de spraakproductie.

Samenvattend onderzocht ik in dit proefschrift twee vragen die verband houden met de observatie dat mensen erin slagen om hun spraak te plannen in lawaaierige omgevingen. De eerste vraag was hoe spraakproductie werd beïnvloed door irrelevante achtergrondspraak. Het bleek dat verschillende eigenschappen van irrelevante achtergrondspraak (gelijkenis in representatie, de eis voor aandacht, verstaanbaarheid, interessantheid en context) verschillende mogelijkheden voor verstoring hebben. De tweede vraag was of het effect van de irrelevante achtergrondspraak werd beïnvloed door de moeilijkheidsgraad van de spraakproductie. Er werd gevonden dat mensen enkele strategieën gebruiken om om te gaan met dit effect. Bijvoorbeeld, ze concentreerden zich meer op het benoemen van moeilijke plaatjes, waardoor de verstoring van irrelevante achtergrondspraak verminderd werd. Daarnaast repliceerde een online experiment (hoofdstuk 3) twee bevindingen die eerder waren verkregen in experimenten in het lab – het effect van benoemingsconsensus en het semantische contexteffect – wat de haalbaarheid van het uitvoeren van experimenten op het gebied van spraakproductie in een online omgeving onderschrijft. Dit proefschrift biedt daarom belangrijke inzichten in hoe mensen spraak plannen en produceren in de aanwezigheid van irrelevante achtergrondspraak, en het draagt ook bij aan ons begrip van de verwerking van conversaties die de coördinatie van spreken en luisteren vereist.

English Summary

Conversation exists in all cultures and is the most common way of using language. In daily life, many people spend considerable hours exchanging ideas, thoughts, and feelings with others via spoken conversation which makes it the most natural and basic form of language use. An important characteristic of spoken conversation is the regular exchange of turns between interlocutors, that is, the switch of roles between speakers and listeners. This means spoken conversation requires dual-tasking between speaking (production) and listening (comprehension). Moreover, spoken conversation takes place in many different physical environments. It may occur in a quiet room undisturbed by external sources of background noise. It may also occur in noisy environments, such as a crowded cafeteria or a busy train, where a lively discussion or a phone conversation may be overheard in the background. Many people prefer to have conversations in quiet places because they find noisy environments distracting and disturbing their communication. This raises a question of how spoken conversation is affected by irrelevant background noise that includes non-verbal noise (e.g., traffic or construction) and verbal noise (e.g., background conversations, radio, and television broadcasting). To answer this question, it is essential to describe how people plan their own speech and comprehend their interlocutors' speech in the presence of irrelevant background noise. The present dissertation achieved this goal by focusing on a specific question of how speech production (speaking) was affected by irrelevant background speech (verbal background noise). The second question of this dissertation was related to the observation that people can manage disruptions by background noise when they focus on cognitive tasks such as when working in an office or engaging in reading. In particular, when performing difficult tasks, people tend to concentrate harder on the tasks and reduce disruption from background noise. Thus, the second purpose of this dissertation was to investigate how people manage the influence of irrelevant background speech when they plan speech with varying difficulty.

Chapter 2 of this dissertation explored how speech production was influenced by different types of irrelevant background speech (with varied representational similarity in Experiment 1 and varied attention demand in Experiment 2) and whether this influence was

modulated by the difficulty of speech production. In Experiment 1, participants—all native Dutch speakers—named sets of pictures with varying difficulty in Dutch while ignoring Dutch word lists (intelligible background speech; high representational similarity), Chinese word lists (unintelligible background speech; medium representational similarity), or eight-talker babble (language-like noise; low representational similarity). It was found that participants' naming performance was disrupted more by background speech (Dutch and Chinese word lists) than language-like noise (eight-talker babble), and the naming performance was impaired more by intelligible (Dutch word lists) than unintelligible (Chinese word lists) background speech. In Experiment 2, other native Dutch speakers named sets of pictures with varying difficulty in Dutch while ignoring Dutch word lists (low attention demand) or paying attention to Dutch word lists for a later memory task (high attention demand). It was found that speech production performance was impaired in the high, relative to the low, attention demand condition. Moreover, the disruptions caused by the irrelevant background speech in both experiments were reduced when naming difficult pictures. These findings suggest that increased representational similarity and attention demand of irrelevant background speech cause more disruption in speech production, but the disruption can be shielded against when performing difficult speech production tasks.

Because the COVID-19 pandemic impeded lab-based studies, Chapter 3 assessed the feasibility of conducting speech production research in web-based settings. To this end, two classic findings observed in lab-based studies were examined. The first is that naming a picture with high name agreement (e.g., a picture of an *apple* is almost always called an *apple*) is faster and more accurate than naming a picture with low name agreement (e.g., a picture of a *sofa* can also be called a *couch*, *loveseat*, or *lounge*), which is referred to as the name agreement effect. The second finding is that it is slower and more error-prone to name pictures from the same semantic category (homogeneous context; e.g., *dolphin, butterfly, mouse, lion*) than from different semantic categories (heterogeneous context; e.g., *dolphin, ear, glove, comb*), which is referred to as the semantic context effect. Native Dutch speakers named sets of pictures with varied name agreement (high versus low) and semantic context (homogeneous versus heterogeneous) online. The two effects were replicated in this web-based study, supporting the feasibility of conducting speech production research in online environments.

Chapter 4 returned to the investigation of how speech production was affected by irrelevant background speech. Because the irrelevant background speech (Dutch word lists, Chinese word lists, and eight-talker babble) used in Chapter 2 differed in multiple aspects, such as linguistic (e.g., semantics and phonology) and non-linguistic properties (e.g., acoustic pattern; segmented versus continuous sound stream). Therefore, the disruption could be caused not only by the difference in representational similarity (i.e., linguistic content) but also in acoustic pattern (i.e., non-linguistic properties). This chapter thus examined whether the acoustic pattern (segmented versus continuous as word lists versus sentences) of unintelligible background speech (Chinese) influenced speech production (Experiment 1), and whether this influence also occurred when irrelevant background speech was intelligible (Dutch; Experiment 2). In both experiments, native Dutch speakers named sets of pictures with varying difficulty in Dutch while ignoring word lists or sentences (Chinese in Experiment 1, Dutch in Experiment 2), or in a quiet condition. It was found that unintelligible (Chinese) and intelligible (Dutch) background speech impaired speech production performance relative to quiet, but no difference between word lists and sentences was found in either experiment. This suggests that, regardless of intelligibility, the presence of irrelevant background speech impairs speech production performance, but the acoustic pattern of the background speech does not modulate disruption. Similar to Chapter 2, the disruption caused by intelligible background speech (Dutch) was reduced when naming difficult pictures, which indicates that speakers can manage the disruption when planning difficult speech.

Chapter 5 further investigated how speech production was influenced by different types of irrelevant background sentences. Because the irrelevant background speech (word lists, sentences) used in Chapter 4 is boring and simple, it may have been easier to ignore, showing no difference between word lists and sentences. This chapter used relatively large differences in background sentences with respect to interestingness (boring versus funny; interestingness) and context (constant versus varied). Native Dutch speakers named sets of pictures with varying difficulty in Dutch while ignoring background sentences in a constant context condition (only with boring sentences) followed by a varied context condition (with boring and funny sentences). Funny sentences were less disruptive than boring sentences, and boring sentences disrupted speech production less in the varied than in the constant context. Moreover, a reversed pattern from Chapters 2 and 4 emerged: the disruption caused by background sentences was larger for difficult than easy picture naming. These findings suggest that the interestingness and context of irrelevant background speech influence speech production, which, in turn, is modulated by the difficulty of speech production.

To conclude, this dissertation investigated two questions related to the observation that people manage to plan their speech in noisy environments. The first question was how speech production was influenced by irrelevant background speech. It was found that various properties of irrelevant background speech (representational similarity, attention demand, intelligibility, interestingness, and context) have different potentials for disruption. The second question was whether the influence of irrelevant background speech was affected by the difficulty of speech production. It was found that people used some strategies to manage the influence, for example, they concentrated harder on difficult picture naming, reducing the disruption from irrelevant background speech. In addition, a web-based study (Chapter 3) replicated two findings previously obtained in lab-based studies—the name agreement effect and the semantic context effect—which supports the feasibility of conducting speech production research in an online environment. This dissertation, therefore, provides important insights into how people plan and produce speech in the presence of irrelevant background speech, and also contributes to our understanding of the processing of conversation requiring the coordination of speaking and listening.

Acknowledgments

A PhD is a long effortful, but also a magical journey. Finishing a PhD is always accompanied by a large group of people who provides you with support and help. Surely this thesis was only possible because of the invaluable guidance and encouragement I received throughout my PhD. I am immensely grateful to have so many people, in Nijmegen and beyond, to say “thank you” to.

Let us start with my supervisory team: **Antje, Laurel**, **Candice**, and **Ava**. Massive thanks to all of you for your dedication to the projects included in my thesis, and for bringing your diverse ideas and skills to train me to be a qualified researcher. **Antje**, you are definitely a wonderful supervisor. Thank you for your generous guidance and support in this thesis: you kept a watchful eye on the quality of my research, you taught me to anticipate and think ahead, and you plugged in the logical flaws in my thinking. Meeting with you was always inspiring and uplifting, and I learned a lot from your rigorous questions. Your comments on theoretical hypotheses, experimental designs, and manuscripts were insightful and incisive, which taught me to become more precise at reasoning and writing arguments. I am really grateful for your “do not feel despair, this is science” emails when I lost my confidence, which gave me timely comfort and encouragement. I would also like to acknowledge your efforts in maintaining an academically successful research group, and in organizing social events for us whenever there was a good enough reason.

Laurel, I was truly lucky to have you as my daily supervisor. I do not think I can put into words how grateful I am. Your expertise in language production and insightful comments/feedback on my ideas and manuscripts were vital for this thesis. I always looked forward to our weekly meetings in your office where we would discuss scientific topics: you helped me straighten out my half-baked ideas and then plant seeds of studies; you challenged me to think deeply and carefully; you had me to consider every possible outcome when designing experiments and multiple interpretations when discussing results; you showed me that working in science is not only about research but also about showing the world that the research we do is important...I deeply appreciate your regular reminders that it was okay not to perform well at times, and your suggestion that it is necessary to keep a work-and-life balance. Apart from being academically inspiring, thank you for always making time to listen when I had concerns, no matter whether it is big or small, and also thank you for feeding me delicious food such as Christmas cookies.

Candice, I am really grateful that you became part of my supervisory team. Working with you was such a pleasure. Thank you for always taking the time to answer my many questions, explain things to me in detail, provide prompt feedback on my manuscripts, and encourage me to look for solutions instead of problems. Your support and input have been crucial for the successful completion of this thesis. I also very much appreciate your patient support and guidance while I was applying for the postdoc position. Outside of research, we talked a lot about light-hearted subjects such as our growth experience and cultural difference. Your presence has been a gift.

Ava, you entered my supervisory team like a fresh breeze. Thank you for being always available and for your critical evaluation of my work. I learned a lot from you about how to conduct scientific research in a structured and thoughtful way. For example, you taught me to be cautious in stimuli preparation by controlling irrelevant variables well, to be strict in data analysis by using appropriate statistical models, and to be persuasive when writing manuscripts by circling back to "this study is important because...". Thank you also for being the kind-hearted person you are because you were always encouraging and positive when I felt frustrated. In addition, I want to sincerely thank the members of my reading committee, **Ardi Roelofs**, **Iva Ivanova**, and **Sara Bögels**, for your speedy but careful reading and evaluation of this thesis.

Next, on to the large body of support: the Psychology of Language (PoL) department. I feel very lucky to have been part of the PoL which is home to a friendly group of researchers. First, I would like to thank my good officemates: **Fan, Aitor, Joe, and Orhun.** Thank you for many fun conversations in and out of the office. Second, I am grateful for having had the opportunity to work among brilliant members of the **Hans Rutger, Falk, Florian, Ruth, Elli, Stan, Olena, Sho, Sandra, Ronny, Caitlin, Cecília, Christina, Morgane, Franziska, Sophie, Veerle, Yuxi, Giulio, Andera, Merel, Amie, Greta, Eirini, Federica, etc.** over the years. Thank you for the great lab meetings in which you gave me constructive and insightful feedback on earlier versions of the research reported in this thesis, and also for the unpredictable lunch conversations and fun parties. It is such a pleasure working with you. Third, a massive thank you to the project coordinator, secretary, and assistants: **Annelies**, thank you for lending your voice to many of my experimental stimuli, and for helping me with a bunch of translations. **Thy**, thank you for doing such a wonderful job in coordinating the work beyond research. Also, big thank you to the research/student assistants who took off some of my workloads by annotating endless recordings of speech production. Surely, I should not forget my paranymphs: **Veerle and Franziska.** Thank you for agreeing on my short notice to be my paranymphs, and for your help in making my defense a special day and having my back. You are such friendly colleagues and caring friends.

Then, I am also hugely grateful to the many people who, behind the scenes, helped me with my research at the MPI. In particular, massive thanks to **Kevin**, who coordinates the IMPRS with so much dedication. Thank you for always being kind and helpful to IMPRSers, and also for your timely encouraging emails when I almost lost hope during the writing of this thesis. No single project at the MPI can be done without you, the technical group. Thank you for being there when needed. Special thanks to **Maarten** and **Thijs** for your assistance in programming my online experiments, and also for your near-instant answers to other bookish questions. I want to express my appreciation to the librarians, **Meggie** and **Karin**, you run the library like no other. I would also like to thank the operation staff, receptionists, and secretaries, you ensure that everything runs smoothly and problems are solved even before we may notice them. In addition, thank you, my participants, for your time and patience.

Last but not least, I met many wonderful people whom we became friends with. Some are still here and some have been left. A special thank you to **Erik,** who is an amazing landlord and listener. Thank you for helping me without hesitation when I was in trouble, such as repairing my bicycle tyre and talking with me when I felt upset. 最特别的感谢给全世界最棒的朋友，**娄昊**。谢谢你成为我可以随时打扰的存在，谢谢你聆听我的无数苦闷与哀伤，也谢谢你陪着我天南地北地看风景。没有你，我的博士生活想必会缺少一份温暖且安心的陪伴。特殊的感谢给一位超棒的师兄，**白帆**。谢谢你在我遇到挫折时毫不犹豫地给予我帮助与支持，也谢谢你和嫂子在无数个周五精心准备的大餐，和你们聊天吃饭的日子会成为我一份美好的回忆。真诚的感谢给一位重要的朋友，**刘艺。**谢谢你与我分享读博旅程中的喜怒哀乐，也谢谢你在制定旅行计划后带着我四处游玩。还有，谢谢我的其他好友，**家昱，晓茹，冰青，翁鑫，丹婷，Fiona, 丁榕，铁梅**等，谢谢你们的存在与陪伴，使得我的博士生活不会那么孤单。最后，谢谢我的家人，谢谢你们给予我坚实且有力的支持。没有你们，我无法如期完成博士论文。

Curriculum Vitae

Jieying He was born in Jiuquan, China in 1993. In 2018, she obtained her master's degree in psycholinguistics, with a specialization in written language production, from Renmin University of China. In the same year, she moved to Nijmegen to start her Ph.D. research at the Psychology of Language department of the Max Planck Institute for Psycholinguistics. There she investigated the coordination of speaking and listening by focusing on the question of how speech production is affected by irrelevant background speech. Jieying will continue her research on language processing as a post-doctoral researcher at the Neurobiology of Language department at the Basque Center on Cognition, Brain and Language in July, 2023.

Publications

He, J., Frances, C., Creemers, A., & Brehm, L. (Under submission). Effects of irrelevant background speech on spoken language production.

He, J., & Zhang, Q. (Under review). Direct retrieval of orthographic representations in Chinese handwritten production: Evidence from a dynamic causal modelling study.

He, J., Brehm, L., & Zhang, Q. (2022). Dissociation of writing processes: A functional magnetic resonance imaging study on the neural substrates for the handwritten production of Chinese characters. *Journal of Cognitive Neuroscience.* 34(12), 2320-2340. https://doi.org/10.1162/jocn_a_01911

He, J., Meyer, A. S., Creemers, A., & Brehm, L. (2021). Conducting language production research online: A web-based study of semantic context and name agreement effects in multi-word production. *Collabra: Psychology*, *7*(1), 29935. https://doi.org/10.1525/collabra.29935

He, J., Meyer, A. S., & Brehm, L. (2021). Concurrent listening affects speech planning and fluency: the roles of representational similarity and capacity limitation. *Language, Cognition and Neuroscience*, *36*(10), 1258-1280. https://doi.org/10.1080/23273798.2021.1925130

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 argumentstructure 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 anAustralian 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 Pluymaekers*
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 informationtheoretical 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 Turkishlearners 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 nonunderstanding. *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*
146. The role of exemplars in speech comprehension. *Annika Nijveld*
147. A network of interacting proteins disrupted in language-related disorders. *Elliot Sollis*

148. Fast speech can sound slow: Effects of contextual speech rate on word recognition. *Merel Maslowski*
149. Reasons for every-day activities. *Julija Baranova*
150. Speech planning in dialogue - Psycholinguistic studies of the timing of turn taking. *Mathias Barthel*
151. The role of neural feedback in language unification: How awareness affects combinatorial processing. *Valeria Mongelli*
152. Exploring social biases in language processing. *Sara Iacozza*
153. Vocal learning in the pale spear-nosed bat, Phyllostomus discolor. *Ella Lattenkamp*
154. Effect of language contact on speech and gesture: The case of Turkish-Dutch bilinguals in the Netherlands. *Elif Zeynep Azar*
155. Language and society: How social pressures shape grammatical structure. *Limor Raviv*
156. The moment in between: Planning speech while listening. *SvetlanaLito Gerakaki*
157. How speaking fast is like running: Modelling control of speaking rate. *Joe Rodd*
158. The power of context: How linguistic contextual information shapes brain dynamics during sentence processing. *René Terporten*
159. Neurobiological models of sentence processing. Marvin Uhlmann
160. Individual differences in syntactic knowledge and processing: The role of literacy experience. *Saoradh Favier*
161. Memory for speaking and listening. *Eirini Zormpa*
162. Masculine generic pronouns: Investigating the processing of an unintended gender cue. *Theresa Redl*
163. Properties, structures and operations: Studies on language processing in the brain using computational linguistics and naturalistic stimuli. *Alessandro Lopopolo*
164. Investigating spoken language comprehension as perceptual inference. *Greta Kaufeld*
165. What was that Spanish word again? Investigations into the cognitive mechanisms underlying foreign language attrition. *Anne Mickan*
166. A tale of two modalities: How modality shapes language production and visual attention. *Francie Manhardt*
167. Why do we change how we speak? Multivariate genetic analyses of language and related traits across development and disorder. *Ellen Verhoef*
168. Variation in form and meaning across the Japonic language family with a focus on the Ryukyuan languages. *John Huisman*
169. Bilingual sentence production and code-switching: Neural network simulations. *Chara Tsoukala*
170. Effects of aging and cognitive abilities on multimodal language production and comprehension in context. *Louise Schubotz*
171. Speaking while listening: Language processing in speech shadowing and translation. *Jeroen van Paridon*

172. Left-right asymmetry of the human brain: Associations with neurodevelopmental disorders and genetic factors. *Merel Postema*
173. Abstract neural representations of language during sentence comprehension: Evidence from MEG and Behaviour. *Sophie Arana*
174. Infants' perception of sound patterns in oral language play. *Laura Hahn*
175. Spoken and written word processing: Effects of presentation modality and individual differences in experience to written language. *Merel Wolf*
176. Development of Spatial Language and Memory: Effects of Language Modality and Late Sign Language Exposure. *Dilay Karadoller*
177. Kata Kolok - Variation and acquisition. *Hannah Lutzenberger*
178. Individual differences in speech production and maximum speech performance. *Chen Shen*
179. Non-native phonetic accommodation in interactions with humans and with computers. *Aurora Troncoso Ruiz*
180. About time: Exploring the role of grammatical aspect in event cognition. *Julia Misersky*
181. Non-native Lombard speech: The acoustics, perception, and comprehension of English Lombard speech by Dutch natives. *Katherine Marcoux*
182. The enlanguaged brain: Cognitive and neural mechanisms of linguistic influence on perception. *Ksenija Slivac*
183. Discovering the units in language cognition: From empirical evidence to a computational model. *Jinbiao Yang*
184. Neural representation of speech segmentation and syntactic structure discrimination. *Fan Bai*
185. Modelling multi-modal language learning: From sentences to words. *Danny Merkx*
186. Disentangling the molecular landscape of genetic variation of neurodevelopmental and speech disorders. *Joery den Hoed*
187. What's on your mind: Mental simulation and aesthetic appreciation during literary reading. *Marloes Mak*
188. Repetita Iuvant? Studies on the role of repetition priming as a supportive mechanism during conversation. *Federica Bartolozzi*
189. Mutual understanding from a multimodal and interactional perspective. *Marlou Rasenberg*
190. Triangles in the brain: The role of hierarchical structure in language use. *Cas Coopmans*
191. Linguistic alignment: The syntactic, prosodic, and segmental phonetic levels. *Lotte Eijk*
192. Speaking in the Brain: How the brain produces and understands language. *Laura Giglio*
193. Coordination of spoken language production and comprehension: How speech production is affected by irrelevant background speech. *Jieying He*