Original Article

QJEP

Quarterly Journal of Experimental Psychology
1–25

DOI: 10.1177/17470218231219971
qjep.sagepub.com
Sage

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

Jieying He[1,2], Candice Frances[1], Ava Creemers[1] and Laurel Brehm[1,3]

Abstract

Earlier work has explored spoken word production during irrelevant background speech such as intelligible and unintelligible word lists. The present study compared how different types of irrelevant background speech (word lists vs. sentences) influenced spoken word 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 a similar group with Dutch word lists and sentences. In both experiments, the lexical selection demands in speech production were manipulated by varying name agreement (high vs. low) of the to-be-named pictures. Results showed that background speech, regardless of its intelligibility, disrupted spoken word 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 with the quiet condition was eliminated when planning low name agreement pictures. These findings suggest that any speech, even unintelligible speech, interferes with production, which implies that the disruption of spoken word production is mainly phonological in nature. The disruption by intelligible background speech can be reduced or eliminated via top–down attentional engagement.

Keywords

Irrelevant speech effect; name agreement; speech production

Received: 26 May 2023; revised: 4 September 2023; accepted: 3 October 2023

Introduction

Much of daily conversation, which requires both speech comprehension and production, occurs in the presence of irrelevant external auditory stimulation, including noise from nearby traffic or construction, a television broadcasting in the background, or a colleague talking on the phone. Extensive work has shown that background noise, music, and speech all have detrimental effects on spoken language comprehension (e.g., Eckert et al., 2016). However, very few studies have investigated how speakers plan their speech in the presence of irrelevant background noise, especially irrelevant background speech (e.g., Fargier & Laganaro, 2016, 2019; He, Meyer & Brehm, 2021). Understanding speech production in non-verbal and verbal sources of noise advances our understanding of how speakers cope with auditory disruption when planning their speech. The present study thus investigated how different types of irrelevant background speech (word lists and sentences) influenced spoken word production with varying lexical selection demands, and whether the influence was modulated by the difficulty of speech production.

One irrelevant speech effect, two relevant theories

Previous 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;

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands
[3]Department of Linguistics, University of California, Santa Barbara, Santa Barbara, CA, USA

Corresponding author:
Jieying He, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands.
Email: Jieyingpsy@gmail.com

Schlittmeier et al., 2012) and reading (e.g., Cauchard et al., 2012; Hyönä & Ekholm, 2016; Yan et al., 2018), even when they 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 & Morris, 1992). One major account for the irrelevant speech effect is the involvement of shared mechanisms or representations in both tasks; this is known as the *domain-specific interference-by-similarity account* (e.g., Jones et al., 1993; Martin et al., 1988; Salamé & Baddeley, 1982, 1989). This was first proposed to explain the changing-state effect in serial recall where distractor sequences like A B C D E F G H disrupt more than A A A A A A A A (Hughes, 2014; Hughes et al., 2007; Jones et al., 1993; Jones & Morris, 1992). The effect has been attributed to conflict driven by automatic processing of the irrelevant auditory distractors' order (*interference-by-process account*; e.g., Hughes, 2014; Jones et al., 1993). This interference-by-similarity account resembles the crosstalk account for dual-task processing based on neural resources (Pashler, 1994; *outcome conflict*: Navon & Miller, 1987), claiming that shared or similar representations or processes cause interference in task performance.

Two hypotheses attribute the irrelevant speech effect to different sources that are both important to consider for the effect of background speech on speech production. The *phonological disruption view* (Salamé & Baddeley, 1982, 1989) hypothesises that the irrelevant speech effect results from the similarity in content of phonological codes (e.g., reading and irrelevant background speech), which are both buffered in a phonological memory store (a component of the phonological loop; Baddeley, 2000, 2003). This view predicts that disruption in speaking should occur from the presence of irrelevant background speech, regardless of its content. By contrast, the *semantic disruption view* (Martin et al., 1988) attributes the effect to the shared use of semantic processing (e.g., English reading is disrupted more by English-intelligible- than Russian-unintelligible-background speech). This view predicts that disruption in speaking should be produced by intelligible meaningful speech because meaningless speech does not recruit semantic processing.

In contrast to the domain-specific interference-by-similarity, the *domain-general attention capture* account posits that irrelevant speech or sound 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). When the focus of attention is captured by task-irrelevant sounds, fewer attentional resources are available and task performance is impaired. The attention capture theory has some support in how irrelevant background speech interferes with serial recall performance (e.g., Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015) and reading (e.g., Hyönä & Ekholm, 2016). This attention capture account is compatible with the capacity limitation account for dual-task processing (Pashler, 1994; Ruthruff et al., 2003), which states that the amount of attentional resources available to focal cognitive tasks determines task performance.

There is a similar divide within this domain-general attention capture view with different predictions of the effects of irrelevant background on speech production (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 predicts that irrelevant background speech with varied context (stimulus-*aspecific* variation, e.g., pauses in speech) should interfere more with the focal task than background speech with constant context (e.g., continuous speech). Alternatively, *specific attention capture* can occur when the content of the sound diverts attention (e.g., Eimer et al., 1996; Röer et al., 2013; Wood & Cowan, 1995), which implies that the attention-diverting power is attributable to the stimulus itself (stimulus-*specific* variation). This view predicts irrelevant background speech with rich linguistic representations (e.g., full sentences) should elicit more disruption than that with less linguistic information (e.g., word lists).

Irrelevant speech effects in spoken language production

The earlier work is nearly all conducted on language comprehension, and importantly, similar processes may or may not be relevant for speech production. Prior literature has indicated that speech production and comprehension draw upon similar processes/representations (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 domain-specific interference-by-similarity (Martin et al., 1988; Salamé & Baddeley, 1982, 1989) and domain-general 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. However, it is also important to note that speech production and speech comprehension are also fundamentally different processes, with different goals (production = convert message to output form; comprehension = convert input form to message), and different burdens of attention. This makes it important to systematically investigate the irrelevant speech effect in language production.

Evidence from the picture–word interference (PWI) studies (Glaser & Düngelhoff, 1984; Schriefers et al., 1990) has supported the interference-by-similarity explanation. When naming a picture (e.g., DOG) with a spoken related distractor word (e.g., FOX), naming latencies and error rates increased compared with trials with an

unrelated distractor (e.g., RANK; Damian & Martin, 1999; Schriefers et al., 1990). This suggests that the distractor word activated semantic representations required by the target word, interfering with spoken word production when they are related (see Roelofs, 1992, 2003), which is consistent with the semantic disruption view (Martin et al., 1988). When 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. This, in turn, implies that if what is produced mismatches with what is comprehended, pre-activation of phonological/phonetic representations could also elicit interference, which is consistent with the phonological disruption view (Salamé & Baddeley, 1982, 1989).

Fargier and Laganaro (2016) investigated the roles of both interference-by-similarity and capacity limitation mechanisms by using a dual-task paradigm. Participants named pictures in three listening conditions with varying attentional demand: without distractors (low), while passively listening to distractors (medium), and during a distractor detection task (high). The auditory distractors were either tones (non-verbal stimuli) or syllables (verbal stimuli). Production latencies were longer for syllables relative to tones, and increased for tasks with higher attentional demand. These results suggest that increased representational similarity and attentional demand cause more interference on speech production performance.

To expand on earlier work on interference between single-word production and comprehension (e.g., Fargier & Laganaro, 2016; Glaser & Düngelhoff, 1984; Schriefers et al., 1990), He, Meyer & Brehm, 2021 conducted a study which mainly supports the role of interference-by-similarity in the irrelevant speech effect for speech production. In this study, Dutch speakers named sets of pictures while ignoring Dutch word lists, Chinese word lists, or eight-talker babble (i.e., language-like noise). Irrelevant background speech (Dutch and Chinese word lists) disrupted spoken word production more than eight-talker babble, and Dutch caused more disruption than Chinese word lists. This suggests that more interference on spoken word production is obtained as the representational similarity between speech production and irrelevant background speech increases, consistent with the interference-by-similarity view (Martin et al., 1988; Salamé & Baddeley, 1982, 1989). However, He, Meyer & Brehm, 2021 did not distinguish between phonological and semantic sources of disruption, which might both contribute to interference. This study also does not rule out disruption by attention capture because the irrelevant background speech varied in both aspecific context (pauses in word lists but not in eight-talker babble) and specific linguistic content (information content in word lists but not in eight-talker babble).

Furthermore, because speaking requires attention, task demands may modulate the irrelevant speech effect in language production. He, Meyer & Brehm, 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 one with low name agreement (e.g., the item called *sofa* or *couch;* e.g., Alario et al., 2004; Cheng et al., 2010; Shao et al., 2014; Vitkovitch & Tyrrell, 1995). The effect is caused by both difficulty in object recognition (confusion over what the object should be called) and the demands of lexical selection (the need to select among competing lexical candidates); He, Meyer & Brehm, 2021 used stimuli designed to elicit the latter effect. Irrelevant speech effects were strongest for high name agreement pictures with low lexical selection demands, which suggests that the interference can be eliminated when speech production is more demanding. The finding is consistent with a top–down *attention engagement mechanism* (also referred to as *task engagement*; see Halin et al., 2014; Marsh et al., 2015): difficult speech production may make speakers concentrate harder and reduce processing of irrelevant background speech. This means that to study irrelevant speech effects in speech production, it is also important to consider the production demands.

Current study

The present study was designed to explore how different types of irrelevant background speech affected spoken language production. Given that previous studies have supported the reliability of conducting speech production research online (e.g., Fairs & Strijkers, 2021; He, Meyer, Creemers, & Brehm, 2021; Stark et al., 2022; Vogt et al., 2022), we designed two web-based experiments which focused on teasing apart the variants of the interference-by-similarity and attention capture accounts. To distinguish between the semantic and phonological interference-by-similarity views, we examined disruption by unintelligible (Chinese, Experiment 1) and intelligible background speech (Dutch, Experiment 2) on Dutch spoken word production. The phonological disruption view (Salamé & Baddeley, 1982, 1989) predicts that background speech, regardless of its intelligibility, should disrupt speech production relative to a quiet condition, predicting a similar pattern of results across experiments. By contrast, the semantic disruption view (Martin et al., 1988) predicts that only intelligible background speech should interfere with speech production, predicting more

Table 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 Exp1) and Dutch speech (in Exp2) should disrupt spoken word production relative to a quiet condition.
Semantic disruption view (Martin et al., 1988)	Chinese speech (in Exp1) should not disrupt spoken word production relative to a quiet condition, but Dutch speech (in Exp2) 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)	Exp1: Chinese word lists should be more disruptive than Chinese sentences. Exp2: Dutch word lists may be more disruptive than Dutch sentences.
Specific attention capture view (Eimer et al., 1996)	Exp1: Chinese word lists should have the same disruptive potency as the sentences. Exp2: Dutch word lists may be less disruptive than Dutch sentences.
Attention engagement account (Halin et al., 2014; Marsh et al., 2015)	
Stimulus-aspecific disruption	Interference elicited by Chinese background speech (in Exp1) should not be affected by name agreement.
Stimulus-specific disruption	Interference elicited by Dutch background speech (in Exp2) should be reduced for low name agreement pictures.

disruption in Experiment 2 than Experiment 1. The predictions for each account in the present study are shown in Table 1.

In both experiments, we compared 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."*). This allows us to distinguish between the two attention capture view variants (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015). In Experiment 1, if attention capture is only caused by *aspecific* context variation (e.g., 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 is only caused by *specific* linguistic content (e.g., semantics or syntax), Chinese word lists should cause the same disruption as the Chinese sentences because they are meaningless to our Dutch speakers. Specific and aspecific properties will also elicit similar patterns of disruption in Experiment 2, though these may be modulated by specific linguistic content because Dutch word lists and sentences differ to Dutch speakers in both semantics and syntax. We thus make relatively weak predictions under the attention capture view variants for Experiment 2. See Table 1 for more details.

In both experiments, we also investigated the role of top–down attention engagement by manipulating the name agreement (high vs. low) and therefore, lexical selection demands, 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. Following earlier work (Alario et al., 2004; Cheng et al., 2010; Shao et al., 2014; Vitkovitch & Tyrrell, 1995), we predicted that pictures with low name agreement would be named more slowly than those with high name agreement in both experiments. Interactions between the type of irrelevant background speech and name agreement also show how the irrelevant speech effects are affected by the required attentional demand of speech production. Because stimulus-aspecific disruption occurs automatically, we predicted that any interference present in Experiment 1 would not be affected by name agreement. This is because the stimulus-aspecific disruption is rooted in the automatic processing of the auditory input that escapes cognitive control (Hughes, 2014). By contrast, stimulus-specific disruption is non-automatic, which means that any disruption caused by the attention-capturing properties of intelligible background speech in Experiment 2 might be reduced for low compared with high name agreement pictures. This is because stimulus-specific disruption requires central attention that taps into cognitive control (Hughes, 2014; Marsh et al., 2018).

Experiment 1

Method

Participants. We recruited 50 native speakers of Dutch who had no experience with Chinese (45 females, M_{age} = 25 years, range: 20–35 years) from the participant pool at the Max Planck Institute for Psycholinguistics. Power simulations (see https://osf.io/wuafh/) showed that 50 participants and 144 items (80% of the items in the study named successfully) would provide 95% power to measure a plausibly sized condition difference of 20 ms

(*SD* = 900 ms). 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; Withers, 2017), a web-based platform developed at the Max Planck Institute for Psycholinguistics. Participants used their own laptops with headphones/earphones. We restricted participation to 14-in. or larger laptops (range: 14–24 in.) with Google Chrome, Firefox, Microsoft Edge, or Brave web browsers. Each participant's speech was recorded by a built-in voice recorder in 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 total of 240 pictures from He, Meyer and Brehm (2021), Experiment 2; pictures selected from the MultiPic database, Duñabeitia et al., 2018; see Supplementary Material, 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 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 neighbourhood frequency (PNF), phonological neighbourhood size (PNS), orthographic neighbourhood frequency (ONF), and orthographic neighbourhood 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 10 above-mentioned attributes, and the high and low name agreement picture sets were 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 all from different semantic categories and the first phoneme of each word was unique, 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 (see Supplementary Material, Table A2), 120 additional Dutch nouns were selected from the MultiPic database (Duñabeitia et al., 2018) and translated into Chinese by a native Mandarin Chinese speaker. These 120 Chinese nouns were divided into 20 word lists of 6 nouns and paired with the 20 picture grids. 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, $t_{(305.91)} = -1.58$, $p > .05$. To avoid phonological overlap between picture naming and background speech, we designed the word lists so that the six Chinese nouns per list did not share the first phoneme, and any five consecutive Chinese nouns per list also did not share the first phoneme with the to-be-named pictures in the same ordinal position. To create practice stimuli, 12 additional Dutch nouns were selected from the same database (Duñabeitia et al., 2018) and 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 44,100 Hz. Each word list was 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 s and there were similar periods of silence (about 700 ms) between consecutive nouns. Naming latencies for pictures can be around 1 s (e.g., Shao et al., 2014; Vitkovitch & Tyrrell, 1995), the duration (the difference from speech onset and offset of a word) of a spoken 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 the presence of background speech. Therefore, we estimated that it takes approximately 2 s to name one picture (also see He, Meyer and Brehm 2021)), totaling 8 s per word list.

For the Chinese sentence condition (see Supplementary Material, Table A3), the 20 Chinese word lists were transformed into 20 Chinese sentences by reversing the order of nouns in the list and 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. Again, no five consecutive Chinese nouns per sentence were phonologically related to any to-be-named pictures in the same ordinal position. The two Chinese word lists were also transformed into two Chinese sentences as practice stimuli. The same speaker recorded these in neutral prosody and they were edited in the same fashion as each Chinese word list (by stretching up to a maximum of 9.59%) to last 8 s.

To test the participants' concentration level and compliance to wearing headphones throughout the experiment, 19 additional two-syllable Dutch nouns (4 for the practice

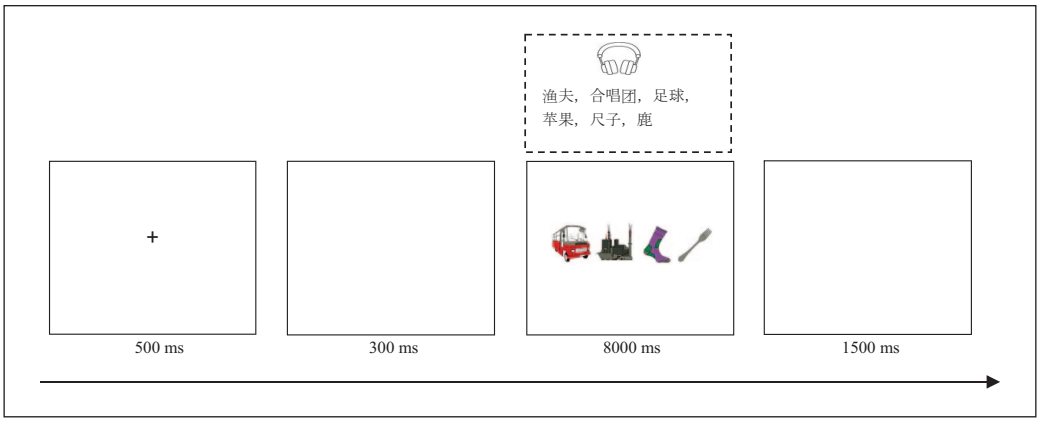

Figure 1. An example trial in which participants named pictures with high name agreement while ignoring a Chinese word list (translation: fisherman, choir, football, apple, ruler, deer).

stage, 15 for the test stage) were selected from Duñabeitia et al. (2018) to be used as attention check stimuli to be repeated back during the experiment. These were recorded by a native Dutch speaker in neutral prosody and matched on intensity, total RMS (root mean square) = −33.98 dB, in Adobe Audition.

Design. The type of unintelligible 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 randomised within experimental blocks and counterbalanced across participants. Items were repeated three times resulting in three blocks containing 60 trials each with one repetition of each background speech condition and 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 stimulus presentation was created for each participant with the Mix programme (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 online[1] 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 imagine 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 them to report their volume values before the test began.

During the experiment, a practice session of 10 trials (six test trials and four attention check trials) was followed by three blocks of experimental trials, each containing 60 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 it during the experiment. They also were asked to fill out a questionnaire asking about their Chinese experience (see Supplementary Material, Table A4). The experiment lasted about 30 min.

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 s. 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 1,500 ms. An example of a test trial is shown in Figure 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.

Analyses. Seven dependent variables were coded to index naming performance. This provides a full description of the many ways production performance can be disrupted. 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 one of those 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 on which all pictures were named correctly and which had no hesitations or self-corrections (hereafter, "fully correct trials"), we calculated four time-based measures. *Onset latency* was defined as the interval from the onset of stimulus presentation to onset of the utterance,

and indexes the beginning stages of speech planning. *Utterance duration* was defined as the interval between the onset of the first picture name and the 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 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 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 quantify the 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 24,000 iterations depending on model convergence (listed in model output tables). We used a warm-up (or burn-in) period of 2,000 iterations in each chain, which means we removed the data based on the first 2,000 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* and *identity* link. Total chunk number and first chunk length had weak normal priors centred at zero with an *SD* of 1, and used family *Poisson* combined with the *log* link. All models were run until the R-hat value for each parameter was 1.00, indicating 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 with 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.

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 were 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 *SD*s 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 appear in Table 2.

Attention check. 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 3, a Bayesian

Table 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)	1,246(462)	1,279 (522)	1,198 (408)	1,434 (579)	1,413 (539)	1,345 (486)
Utterance duration (ms)	2,868(790)	2,868 (771)	2,791(765)	3,475 (1,062)	3,482(1,025)	3,392 (970)
Total pause time (ms)	685(621)	662 (590)	645 (582)	1,078 (860)	1,043 (790)	1,040 (805)
Articulation time (ms)	2,309(431)	2,332 (429)	2,246 (392)	2,518 (498)	2,536 (522)	2,450 (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. NA: name agreement.

mixed-effect model showed that accuracy was considerably lower for low name agreement pictures than high name agreement pictures (β = .099, *SE* = .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 3 and the left panel of Figure 2, a Bayesian mixed-effect model showed that log-transformed onset latency was affected by name agreement: it took participants longer to plan names for low name agreement pictures than high name agreement pictures (β = −.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 = .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 3 and the right panel of Figure 2, a Bayesian mixed-effect model showed that the log-transformed utterance duration was longer for low name agreement pictures than high name agreement pictures (β = −.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 3 and the left panel of Figure 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, −0.46]), 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 3 and the right panel of Figure 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 = −.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.

Total chunk number. As shown in Table 3 and the left panel of Figure 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 = −.252, *SE* = −0.025, 95% Cr.I = [−0.301, −0.203]). There was no interaction between name agreement and the type of background speech.

First chunk length. As shown in Table 3 and the right panel of Figure 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 = .218, *SE* = 0.025, 95% Cr.I = [0.168, 0.258]). 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.

Interim discussion

This experiment provides support for phonological disruption and specific attention capture impacting speech production. Consistent with the phonological disruption view

Table 3. Results of Bayesian mixed-effect models for all dependent variables in Experiment 1.

	Estimate	Est. error	95% Cr. I		Effective samples
			Lower	Upper	
Accuracy					
Population-level effects					
Intercept	0.863	0.017	0.83	0.895	32,170
Name agreement	**0.099**	**0.025**	**0.051**	**0.147**	**59,697**
Speech vs. quiet	0	0.014	–0.028	0.029	107,958
Word List vs. Sentence	0.003	0.011	–0.019	0.025	131,954
NA × (S vs. Q)	–0.02	0.028	–0.076	0.036	107,878
NA × (WL vs. S)	0.001	0.022	–0.042	0.045	134,552
Group-level effects					
Participants					
sd(Intercept)	0.075	0.009	0.06	0.095	27,257
sd(NA)	0.043	0.01	0.024	0.064	54,647
sd(S vs. Q)	0.016	0.012	0.001	0.043	48,050
sd(WL vs. S)	0.012	0.009	0.001	0.033	56,746
sd(NA × (S vs. Q))	0.021	0.016	0.001	0.061	69,866
sd(NA × (WL vs. S))	0.023	0.017	0.001	0.065	55,462
Items					
sd(Intercept)	0.058	0.02	0.016	0.092	6,156
sd(NA)	0.117	0.04	0.033	0.184	6,086
sd(S vs. Q)	0.05	0.018	0.011	0.085	20,580
sd(WL vs. S)	0.03	0.018	0.002	0.066	16,829
sd(NA × (S vs. Q))	0.099	0.037	0.023	0.17	22,166
sd(NA × (WL vs. S))	0.06	0.036	0.003	0.133	17,133
Log-transformed onset latency					
Population-level effects					
Intercept	7.133	0.028	7.078	7.188	5,293
Name agreement	**–0.122**	**0.014**	**–0.149**	**–0.095**	**48,510**
Speech vs. quiet	*0.064*	*0.038*	*–0.011*	*0.138*	*49,911*
Word list vs. sentence	–0.002	0.037	–0.074	0.071	47,960
NA × (S vs. Q)	–0.006	0.07	–0.144	0.132	50,854
NA × (WL vs. S)	–0.014	0.069	–0.15	0.122	56,068
Group-level effects					
Participants					
sd(Intercept)	0.177	0.02	0.143	0.223	10,270
sd(NA)	0.029	0.011	0.005	0.051	18,616
sd(S vs. Q)	0.077	0.015	0.049	0.109	31,488
sd(WL vs. S)	0.05	0.013	0.024	0.077	24,869
sd(NA × (S vs. Q))	0.035	0.025	0.001	0.091	27,704
sd(NA × (WL vs. S))	0.048	0.027	0.003	0.105	21,254
Items					
sd(Intercept)	0.029	0.012	0.004	0.049	2,331
sd(NA)	0.058	0.024	0.008	0.098	2,319
sd(S vs. Q)	0.173	0.095	0.008	0.311	1,284
sd(WL vs. S)	0.177	0.1	0.006	0.316	1,181
sd(NA × (S vs. Q))	0.345	0.189	0.016	0.622	1,222
sd(NA × (WL vs. S))	0.325	0.202	0.011	0.626	1,228
Log-transformed utterance duration					
Population-level effects					
Intercept	8.021	0.023	7.974	8.066	6,414
Name agreement	**–0.191**	**0.02**	**–0.231**	**–0.151**	**39,748**
Speech vs. quiet	0.029	0.026	–0.022	0.08	54,056
Word list vs. sentence	–0.003	0.022	–0.046	0.04	51,599

(Continued)

Table 3. (Continued)

	Estimate	Est. error	95% Cr. I		Effective samples
			Lower	Upper	
NA × (S vs. Q)	0.018	0.05	–0.081	0.117	56,494
NA × (WL vs. S)	0.005	0.044	–0.081	0.091	49,868
Group-level effects					
Participants					
sd(Intercept)	0.142	0.016	0.115	0.178	12,242
sd(NA)	0.064	0.009	0.047	0.084	35,908
sd(S vs. Q)	0.014	0.01	0.001	0.036	35,029
sd(WL vs. S)	0.01	0.007	0	0.026	45,776
sd(NA × (S vs. Q))	0.019	0.014	0.001	0.054	49,185
sd(NA × (WL vs. S))	0.04	0.02	0.004	0.081	31,111
Items					
sd(Intercept)	0.04	0.023	0.002	0.074	1,565
sd(NA)	0.081	0.045	0.004	0.148	1,643
sd(S vs. Q)	0.125	0.055	0.015	0.21	3,193
sd(WL vs. S)	0.111	0.036	0.037	0.173	5,059
sd(NA × (S vs. Q))	0.251	0.109	0.032	0.422	3,182
sd(NA × (WL vs. S))	0.222	0.073	0.072	0.346	4,698
Log-transformed total pause time					
Population-level effects					
Intercept	6.274	0.081	6.115	6.432	7,041
Name agreement	**–0.574**	**0.058**	**–0.687**	**–0.46**	**43,884**
Speech vs. quiet	0.009	0.07	–0.127	0.147	67,063
Word list vs. sentence	0.017	0.064	–0.108	0.143	58,586
NA × (S vs. Q)	0.039	0.134	–0.224	0.304	69,382
NA × (WL vs. S)	0.033	0.126	–0.216	0.283	62,853
Group-level effects					
Participants					
sd(Intercept)	0.508	0.058	0.41	0.635	13,162
sd(NA)	0.177	0.033	0.116	0.247	43,499
sd(S vs. Q)	0.122	0.052	0.017	0.222	26,954
sd(WL vs. S)	0.067	0.04	0.004	0.152	31,799
sd(NA × (S vs. Q))	0.078	0.06	0.003	0.223	53,517
sd(NA × (WL vs. S))	0.126	0.08	0.006	0.298	32,126
Items					
sd(Intercept)	0.107	0.063	0.004	0.204	2,282
sd(NA)	0.222	0.124	0.01	0.409	2,251
sd(S vs. Q)	0.293	0.14	0.023	0.518	3,763
sd(WL vs. S)	0.292	0.102	0.078	0.469	6,780
sd(NA × (S vs. Q))	0.59	0.279	0.049	1.038	3,738
sd(NA × (WL vs. S))	0.579	0.205	0.151	0.935	6,811
Log-transformed articulation time					
Population-level effects					
Intercept	7.768	0.019	7.731	7.805	5,872
Name agreement	**–0.085**	**0.02**	**–0.125**	**–0.046**	**46,351**
Speech vs. quiet	**0.038**	**0.014**	**0.01**	**0.066**	**61,569**
Word list vs. sentence	–0.007	0.012	–0.031	0.017	64,224
NA × (S vs. Q)	0.007	0.027	–0.046	0.06	66,049
NA × (WL vs. S)	–0.003	0.024	–0.05	0.044	62,948
Group-level effects					
Participants					
sd(Intercept)	0.108	0.013	0.087	0.136	11,302
sd(NA)	0.053	0.007	0.041	0.069	28,988
sd(S vs. Q)	0.029	0.008	0.011	0.045	20,619

(Continued)

Table 3. (Continued)

	Estimate	Est. error	95% Cr. I		Effective samples
			Lower	Upper	
sd(WL vs. S)	0.008	0.005	0	0.02	35,991
sd(NA × (S vs. Q))	0.014	0.011	0.001	0.039	41,441
sd(NA × (WL vs. S))	0.021	0.014	0.001	0.051	21,175
Items					
sd(Intercept)	0.042	0.026	0.001	0.078	1,378
sd(NA)	0.083	0.051	0.003	0.157	1,380
sd(S vs. Q)	0.06	0.036	0.002	0.113	1,763
sd(WL vs. S)	0.055	0.029	0.003	0.098	1,923
sd(NA × (S vs. Q))	0.121	0.071	0.005	0.225	1,729
sd(NA × (WL vs. S))	0.106	0.059	0.005	0.195	1,932
Total chunk number					
Population-level effects					
Intercept	0.715	0.041	0.635	0.795	9,365
Name agreement	**–0.252**	**0.025**	**–0.301**	**–0.203**	**52,559**
Speech vs. quiet	–0.016	0.035	–0.085	0.053	74,601
Word list vs. sentence	–0.017	0.029	–0.074	0.040	79,456
NA × (S vs. Q)	0.014	0.070	–0.123	0.152	77,761
NA × (WL vs. S)	0.009	0.058	–0.105	0.123	78,972
Group-level effects					
Participants					
sd(Intercept)	0.256	0.030	0.206	0.321	15,391
sd(NA)	0.062	0.021	0.020	0.104	46,312
sd(S vs. Q)	0.023	0.018	0.001	0.067	62,627
sd(WL vs. S)	0.020	0.016	0.001	0.058	63,929
sd(NA × (S vs. Q))	0.049	0.037	0.002	0.139	64,075
sd(NA × (WL vs. S))	0.043	0.033	0.002	0.122	61,696
Items					
sd(Intercept)	0.035	0.020	0.002	0.073	8,804
sd(NA)	0.070	0.040	0.004	0.146	7,966
sd(S vs. Q)	0.124	0.058	0.012	0.229	9,285
sd(WL vs. S)	0.102	0.043	0.014	0.183	13,656
sd(NA × (S vs. Q))	0.246	0.116	0.020	0.458	9,163
sd(NA × (WL vs. S))	0.202	0.087	0.025	0.365	13,743
First chunk length					
Population-level effects					
Intercept	0.863	0.042	0.781	0.946	11,967
Name agreement	**0.218**	**0.025**	**0.168**	**0.268**	**96,798**
Speech vs. quiet	–0.012	0.034	–0.077	0.055	95,932
Word list vs. sentence	0.013	0.030	–0.046	0.072	92,168
NA × (S vs. Q)	–0.030	0.067	–0.162	0.101	95,948
NA × (WL vs. S)	–0.027	0.060	–0.145	0.091	95,897
Group-level effects					
Participants					
sd(Intercept)	0.262	0.031	0.210	0.330	19,220
sd(NA)	0.022	0.016	0.001	0.061	50,297
sd(S vs. Q)	0.025	0.019	0.001	0.069	64,357
sd(WL vs. S)	0.023	0.018	0.001	0.065	61,516
sd(NA × (S vs. Q))	0.047	0.036	0.002	0.135	64,675
sd(NA × (WL vs. S))	0.043	0.033	0.002	0.122	63,963
Items					
sd(Intercept)	0.047	0.025	0.003	0.090	5,967
sd(NA)	0.094	0.050	0.005	0.179	5,836

(Continued)

Table 3. (Continued)

	Estimate	Est. error	95% Cr. I		Effective samples
			Lower	Upper	
sd(S vs. Q)	0.124	0.053	0.015	0.221	11,407
sd(WL vs. S)	0.116	0.042	0.028	0.195	19,228
sd(NA × (S vs. Q))	0.249	0.106	0.031	0.442	13,355
sd(NA × (WL vs. S))	0.230	0.085	0.051	0.389	18,080

NA: name agreement; WL: word list; S: sentence; Q: quiet.
Models for all dependent variables were run for 24,000 iterations. Bolded values indicate effects where the 95% Cr.I does not contain zero.

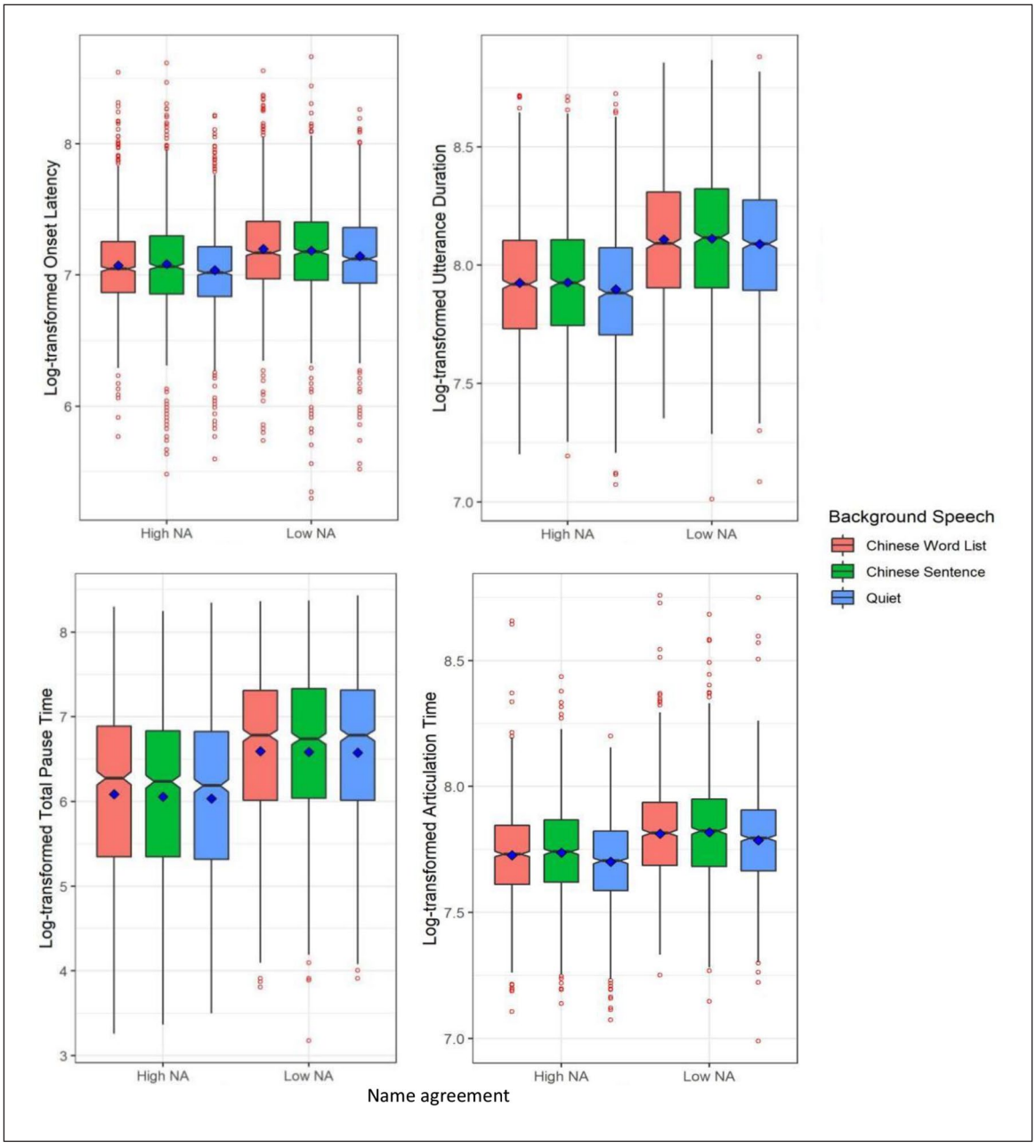

Figure 2. Log-transformed onset latency (top-left), log-transformed utterance duration (top-right), log-transformed total pause time (bottom-left), and log-transformed articulation time (bottom-right) split by name agreement (NA: high, low) and the type of background speech (Chinese word list, Chinese sentence, Quiet) in Experiment 1. Blue squares represent condition means and red points reflect outliers.

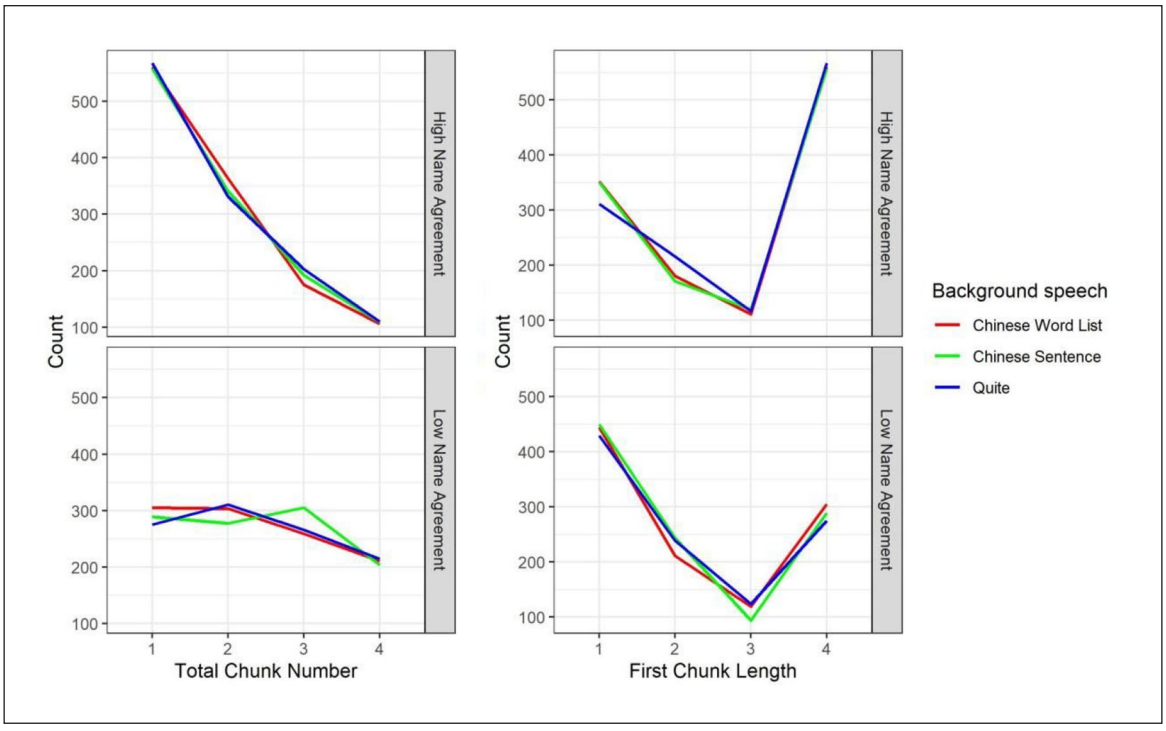

Figure 3. Total chunk number (left) and first chunk length (right) split by name agreement (NA: high, low) and the type of background speech (Chinese word list, Chinese sentence, Quiet) in Experiment 1.

(Salamé & Baddeley, 1982, 1989), 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. Consistent with the specific attention capture view (Eimer et al., 1996), there was no difference between the Chinese word list and Chinese sentence conditions on any dependent measures. Finally, name agreement had a main effect on all dependent measures (as in Alario et al., 2004; He et al., 2021; Shao et al., 2014), but did not interact with the type of Chinese background speech, consistent with the automatic stimulus-aspecific disruption proposal by Hughes (2014).

Experiment 2

Experiment 1 demonstrated clear phonological disruption and specific attention capture effects on unintelligible background speech. However, it is unclear whether these patterns generalise to intelligible background speech. Thus, we extended our investigation to an intelligible-background-speech context by replacing Chinese speech with Dutch speech in Experiment 2. Here, both the phonological and semantic disruption views (Martin et al., 1988; Salamé & Baddeley, 1982, 1989) predict that Dutch speech (word lists and sentences) should disrupt speech production relative to a quiet condition. The aspecific attention capture view (Eimer et al., 1996) predicts there may be more interference in the Dutch word list condition (because of pauses it contains), while the specific attention capture view (Eimer et al., 1996) predicts there may be more disruption in the Dutch sentence condition (due to richer representation recruitment); combined, we make relatively weak predictions under the attention capture variants. Finally, 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 that planning low name agreement pictures would reduce the processing—and thus interference—of Dutch background speech.

Method

Participants. We recruited 47 native Dutch speakers (33 females, M_{age} = 26 years, range: 18–39 years) from the same participant pool as Experiment 1. 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 provide 96% power to measure an interaction between the type of background speech and name agreement on the measurement of utterance duration of 20 ms or smaller (SD = 900 ms) for low name agreement pictures and 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. As in Experiment 1.

Irrelevant background speech. For the Dutch word lists (see Supplementary Material, Table B1), the 120 nouns from Experiment 1 were used in Dutch, and matched with picture names on WF, 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 WF 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 six Dutch nouns per word list were neither semantically nor phonologically related to each other, as described in Experiment 1. 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[2] in neutral prosody and further edited as the Chinese word lists were to last 8 s each with similar silent periods (about 700 ms) between consecutive nouns, by stretching by up to 9.38%.

For the Dutch sentence condition (see Supplementary Material, Table B2), the 20 Dutch word lists were transformed into 20 Dutch sentences as in Experiment 1 by reversing the order of the nouns and then 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 two Dutch word lists were also translated into two Dutch sentences as practice stimuli. The same female native Dutch speaker recorded these sentences in neutral prosody. Sentences were edited to last 8 s 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 = −33.98 dB) in Adobe Audition.

Design. The design was identical to Experiment 1.

Procedure. The procedure was identical to Experiment 1 except that participants did not fill out the questionnaire of Chinese experience.[3]

Analyses. The analysis was the same as Experiment 1.

Results

Six participants were removed from further analyses: one had no audio recordings, three had no responses for attention check trials, one had also participated in Experiment 1, and one had extremely poor-quality audio recordings. The data from the remaining 41 participants were checked for errors as described in Experiment 1. 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 points below 20 ms were also removed for the total pause time measure, resulting in a loss of 12.17% of the data. Finally, any data points more than 2.5 *SDs* 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.

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

Accuracy. Participants produced the intended responses on 87% of the naming trials. As shown in Table 5, a Bayesian mixed-effect model showed that accuracy was lower for low name agreement pictures than high name agreement pictures ($\beta = 1.061$, $SE = 0.223$, 95% Cr.I = [0.630, 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 5 and the left panel of Figure 4, a Bayesian mixed-effect model confirmed that log-transformed onset latency was longer when planning names for low name agreement pictures than high name agreement pictures ($\beta = -.128$, $SE = 0.014$, 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 ($\beta = .076$, $SE = 0.04$, 95% Cr.I = [−0.003, 0.155]). While the 95 % Cr.I contains zero, 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 5 and the right panel of Figure 4, a Bayesian mixed-effect model showed that the log-transformed utterance duration was longer for low name agreement pictures than high name agreement pictures ($\beta = -.215$, $SE = 0.022$, 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

Table 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)	1,304 (496)	1,300 (493)	1,195 (362)	1,451 (568)	1,486 (611)	1,392 (492)
Utterance duration (ms)	2,864 (859)	2,871 (872)	2,690 (776)	3,481 (1,028)	3,463 (1,078)	3,474 (1,087)
Total pause time (ms)	771 (759)	726 (745)	632 (636)	1,090 (877)	1,072 (903)	1,160 (909)
Articulation time (ms)	2,260 (393)	2,274 (415)	2,172 (387)	2,484 (467)	2,482 (482)	2,392 (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)

NA: name agreement. Standard deviations are given in parentheses. All time and chunking measures reflect fully correct trials only.

quiet condition (β = .05, *SE* = 0.031, 95% Cr.I = [−0.012, 0.111]). Here, the 95% Cr.I contains zero but 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 5 and the left panel of Figure 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 = −.599, *SE* = 0.072, 95% Cr.I = [−0.741, −0.458]), but it did not vary with the type of background speech. There was moderate evidence for the interaction of name agreement and the first contrast (Dutch speech vs. Quiet) of background speech (β = .28, *SE* = 0.173, 95% Cr.I = [−0.06, 0.621]). While the 95% Cr.I contains zero, 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 = .394, *SE* = 0.171, 95% Cr.I = [0.058, 0.727]), but not for low name agreement pictures.

Articulation time. As shown in Table 5 and the right panel of Figure 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 longer to articulate names of low name agreement than high name agreement pictures (β = −.093, *SE* = 0.020, 95% Cr.I = [−0.133, −0.054]), and articulation time was longer in the two Dutch speech conditions (word list and sentence) than in the quiet condition (β = .054, *SE* = 0.016, 95% Cr.I = [0.023, 0.085]). There was no interaction between name agreement and the type of background speech.

Total chunk number. As shown in Table 5 and Figure 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* = 0.030, 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 5 and the right panel of Figure 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 = .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.

Interim discussion

The results of Experiment 2 were remarkably similar to those of Experiment 1. Consistent with the phonological disruption view (Salamé & Baddeley, 1982, 1989), the presence of background speech, now in the participants' native language, increased onset latencies and articulation time, and also had a weak impact on utterance durations. There was no difference between the Dutch word list and Dutch sentence conditions on any dependent measures. We also found main effects of name agreement on all dependent measures, and a weak modulation of name agreement on the processing of background speech, such that Dutch background speech increased the total pause time during planning of high, but not low, name agreement pictures. This is consistent with earlier work by He, Meyer and Brehm (2021) and suggests that stronger attentional engagement in the more difficult low name agreement condition leads to less interference from background speech.

General discussion

In two experiments, we explored how different types of unintelligible (Experiment 1) and intelligible (Experiment 2) background speech affected spoken

Table 5. Results of Bayesian mixed-effect models for all dependent variables in Experiment 2.

	Estimate	Est.error	95% Cr. I		Effective samples
			Lower	Upper	
Accuracy					
Population-level effects					
Intercept	2.295	0.165	1.974	2.628	29,013
Name agreement	**1.061**	**0.223**	**0.630**	**1.506**	**79,513**
Speech vs. quiet	–0.043	0.142	–0.328	0.230	118,039
Word list vs. sentence	0.016	0.123	–0.231	0.256	109,284
NA × (S vs. Q)	–0.134	0.275	–0.669	0.412	118,838
NA × (WL vs. S)	0.063	0.246	–0.416	0.553	112,914
Group-level effects					
Participants					
sd(Intercept)	0.812	0.103	0.634	1.038	28,016
sd(NA)	0.317	0.135	0.043	0.582	25,107
sd(S vs. Q)	0.171	0.123	0.007	0.455	45,424
sd(WL vs. S)	0.125	0.093	0.005	0.345	54,483
sd(NA × (S vs. Q))	0.220	0.169	0.008	0.630	64,394
sd(NA × (WL vs. S))	0.236	0.178	0.009	0.663	53,301
Items					
sd(Intercept)	0.478	0.265	0.020	0.868	2,980
sd(NA)	0.901	0.531	0.034	1.714	3,066
sd(S vs. Q)	0.340	0.189	0.021	0.715	19,407
sd(WL vs. S)	0.315	0.187	0.017	0.692	18,572
sd(NA × [S vs. Q])	0.652	0.371	0.039	1.394	21,918
sd(NA × (WL vs. S))	0.601	0.366	0.030	1.338	18,389
Log-transformed onset latency					
Population-level effects					
Intercept	7.161	0.028	7.105	7.216	5,610
Name agreement	**–0.128**	**0.014**	**–0.155**	**–0.1**	**60,813**
Speech vs. quiet	*0.076*	*0.04*	*–0.003*	*0.155*	*61,479*
Word list vs. sentence	–0.004	0.046	–0.096	0.086	65,617
NA × (S vs. Q)	0.04	0.074	–0.104	0.187	64,085
NA × (WL vs. S)	0.022	0.086	–0.147	0.19	66,181
Group-level effects					
Participants					
sd(Intercept)	0.171	0.02	0.136	0.217	12,128
sd(NA)	0.024	0.011	0.003	0.044	22,175
sd(S vs. Q)	0.05	0.014	0.021	0.078	26,754
sd(WL vs. S)	0.028	0.014	0.002	0.054	20,076
sd(NA × (S vs. Q))	0.027	0.02	0.001	0.074	39,897
sd(NA × (WL vs. S))	0.026	0.018	0.001	0.067	39,453
Items					
sd(Intercept)	0.029	0.016	0.001	0.053	1,183
sd(NA)	0.059	0.031	0.003	0.107	1,196
sd(S vs. Q)	0.184	0.106	0.008	0.339	1,012
sd(WL vs. S)	0.233	0.117	0.016	0.405	2,193
sd(NA × (S vs. Q))	0.376	0.213	0.015	0.68	1,029
sd(NA × (WL vs. S))	0.454	0.237	0.029	0.807	2,111
Log-transformed utterance duration					
Population-level effects					
Intercept	8.012	0.028	7.957	8.067	4,298
Name agreement	**–0.215**	**0.022**	**–0.257**	**–0.172**	**34,356**
Speech vs. quiet	*0.050*	*0.031*	*–0.012*	*0.111*	*48,720*

(Continued)

Table 5. (Continued)

	Estimate	Est.error	95% Cr. I		Effective samples
			Lower	Upper	
Word list vs. sentence	0.005	0.024	–0.042	0.052	54,738
NA × (S vs. Q)	0.070	0.060	–0.047	0.187	50,417
NA × (WL vs. S)	–0.007	0.047	–0.100	0.085	58,527
Group-level effects					
Participants					
sd(Intercept)	0.171	0.021	0.136	0.216	11,188
sd(NA)	0.073	0.011	0.054	0.097	31,638
sd(S vs. Q)	0.045	0.014	0.014	0.072	16,224
sd(WL vs. S)	0.008	0.006	0.000	0.023	55,147
sd(NA × (S vs. Q))	0.039	0.027	0.002	0.097	21,573
sd(NA × (WL vs. S))	0.019	0.014	0.001	0.054	45,545
Items					
sd(Intercept)	0.044	0.023	0.002	0.078	1,561
sd(NA)	0.085	0.046	0.004	0.155	1,554
sd(S vs. Q)	0.151	0.065	0.021	0.253	2,658
sd(WL vs. S)	0.112	0.059	0.006	0.200	1,808
sd(NA × (S vs. Q))	0.301	0.130	0.040	0.504	2,617
sd(NA × (WL vs. S))	0.225	0.119	0.012	0.401	1,766
Log-transformed total pause time					
Population-level effects					
Intercept	6.298	0.09	6.12	6.476	8,463
Name agreement	**–0.599**	**0.072**	**–0.741**	**–0.458**	**50,058**
Speech vs. quiet	0.055	0.086	–0.114	0.224	74,556
Word list vs. sentence	0.059	0.068	–0.075	0.194	8,7601
NA × (S vs. Q)	*0.28*	*0.173*	*–0.06*	*0.621*	*74,891*
NA × (WL vs. S)	–0.006	0.137	–0.275	0.263	88,114
Group-level effects					
Participants					
sd(Intercept)	0.542	0.065	0.432	0.687	16,813
sd(NA)	0.28	0.042	0.207	0.373	38,849
sd(S vs. Q)	0.078	0.051	0.004	0.188	27,262
sd(WL vs. S)	0.035	0.027	0.001	0.099	55,607
sd(NA × (S vs. Q))	0.28	0.12	0.035	0.51	25,088
sd(NA × (WL vs. S))	0.117	0.078	0.005	0.29	35,367
Items					
sd(Intercept)	0.125	0.067	0.007	0.227	2,808
sd(NA)	0.249	0.134	0.014	0.455	2,789
sd(S vs. Q)	0.401	0.163	0.067	0.665	4,686
sd(WL vs. S)	0.297	0.168	0.012	0.549	2,653
sd(NA × (S vs. Q))	0.786	0.326	0.123	1.322	4,524
sd(NA × (WL vs. S))	0.589	0.337	0.024	1.099	2,693
Log-transformed articulation time					
Population-level effects					
Intercept	7.744	0.021	7.704	7.785	8,367
Name agreement	**–0.093**	**0.020**	**–0.133**	**–0.054**	**63,460**
Speech vs. quiet	**0.054**	**0.016**	**0.023**	**0.085**	**97,570**
Word list vs. sentence	–0.003	0.013	–0.029	0.022	100,970
NA × (S vs. Q)	0.010	0.030	–0.048	0.069	103,634
NA × (WL vs. S)	0.000	0.026	–0.050	0.051	101,332
Group-level effects					
Participants					
sd(Intercept)	0.120	0.014	0.096	0.152	16,082
sd(NA)	0.055	0.008	0.042	0.071	33,143

(Continued)

Table 5. (Continued)

	Estimate	Est.error	95% Cr. I		Effective samples
			Lower	Upper	
sd(S vs. Q)	0.031	0.007	0.018	0.046	24,300
sd(WL vs. S)	0.007	0.005	0.000	0.018	43,960
sd(NA × (S vs. Q))	0.033	0.017	0.002	0.067	20,736
sd(NA × (WL vs. S))	0.017	0.011	0.001	0.041	37,705
Items					
sd(Intercept)	0.042	0.025	0.001	0.078	1,772
sd(NA)	0.083	0.051	0.003	0.156	1,798
sd(S vs. Q)	0.066	0.040	0.002	0.124	1,927
sd(WL vs. S)	0.058	0.035	0.002	0.108	2,217
sd(NA × (S vs. Q))	0.130	0.080	0.004	0.247	1,977
sd(NA × (WL vs. S))	0.116	0.069	0.004	0.217	2,209
Total chunk number					
Population-level effects					
Intercept	0.728	0.041	0.647	0.808	8,660
Name agreement	**–0.266**	**0.030**	**–0.325**	**–0.208**	**41,811**
Speech vs. quiet	–0.003	0.037	–0.077	0.071	73,370
Word list vs. sentence	0.015	0.030	–0.045	0.074	77,365
NA × (S vs. Q)	0.070	0.075	–0.078	0.217	74,377
NA × (WL vs. S)	0.014	0.061	–0.105	0.133	79,264
Group-level effects					
Participants					
sd(Intercept)	0.246	0.030	0.196	0.312	15,554
sd(NA)	0.086	0.022	0.045	0.132	47,199
sd(S vs. Q)	0.024	0.019	0.001	0.070	62,041
sd(WL vs. S)	0.020	0.015	0.001	0.057	68,947
sd(NA × (S vs. Q))	0.051	0.040	0.002	0.148	61,109
sd(NA × (WL vs. S))	0.040	0.031	0.002	0.114	70,155
Items					
sd(Intercept)	0.047	0.026	0.002	0.092	4,816
sd(NA)	0.094	0.052	0.005	0.184	4,829
sd(S vs. Q)	0.140	0.066	0.012	0.257	7,236
sd(WL vs. S)	0.102	0.057	0.005	0.204	6,819
sd(NA × (S vs. Q))	0.278	0.132	0.023	0.512	7,343
sd(NA × (WL vs. S))	0.201	0.114	0.010	0.407	6,661
First chunk length					
Population-level effects					
Intercept	0.858	0.045	0.767	0.948	8,363
Name agreement	**0.237**	**0.027**	**0.183**	**0.291**	**74,876**
Speech vs. quiet	–0.008	0.043	–0.092	0.076	64,681
Word list vs. sentence	–0.022	0.036	–0.093	0.048	70,214
NA × (S vs. Q)	–0.090	0.085	–0.257	0.078	65,380
NA × (WL vs. S)	–0.005	0.072	–0.146	0.137	70,142
Group-level effects					
Participants					
sd(Intercept)	0.272	0.034	0.214	0.346	17,057
sd(NA)	0.030	0.021	0.001	0.079	35,240
sd(S vs. Q)	0.026	0.019	0.001	0.073	58,663
sd(WL vs. S)	0.021	0.016	0.001	0.060	67,790
sd(NA × (S vs. Q))	0.059	0.044	0.002	0.164	54,199
sd(NA × (WL vs. S))	0.040	0.031	0.002	0.115	72,032
Items					
sd(Intercept)	0.050	0.027	0.003	0.095	4,599
sd(NA)	0.100	0.053	0.006	0.190	4,610

(Continued)

Table 5. (Continued)

	Estimate	Est.error	95% Cr. I		Effective samples
			Lower	Upper	
sd(S vs. Q)	0.185	0.064	0.049	0.300	8,825
sd(WL vs. S)	0.150	0.063	0.020	0.258	6,981
sd(NA × (S vs. Q))	0.367	0.128	0.093	0.595	9,005
sd(NA × (WL vs. S))	0.301	0.125	0.040	0.519	7,420

NA: name agreement; WL: word list; S: sentence; Q: quiet.
Models for all dependent variables were run for 24,000 iterations. Bolded values indicate effects where the 95% Cr.I does not contain zero; italicised values indicate effects where the beta estimate is twice the estimate of the standard error.

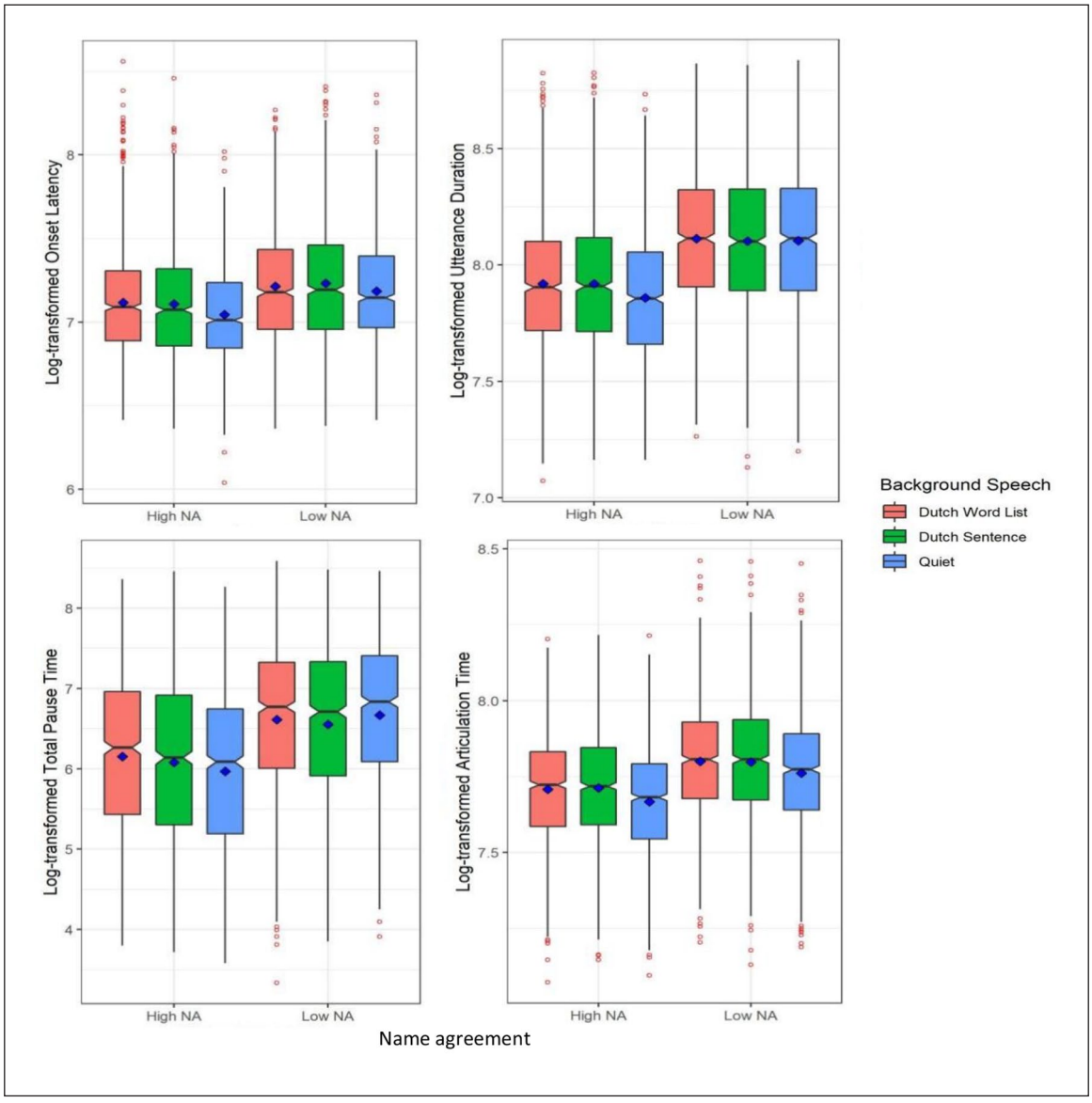

Figure 4. Log-transformed onset latency (top-left), log-transformed utterance duration (top-right), log-transformed total pause time (bottom-left), and log-transformed articulation time (bottom-right) split by name agreement (NA: high, low) and the type of background speech (Dutch word list, Dutch sentence, Quiet) in Experiment 2. Blue squares represent condition means and red points reflect outliers.

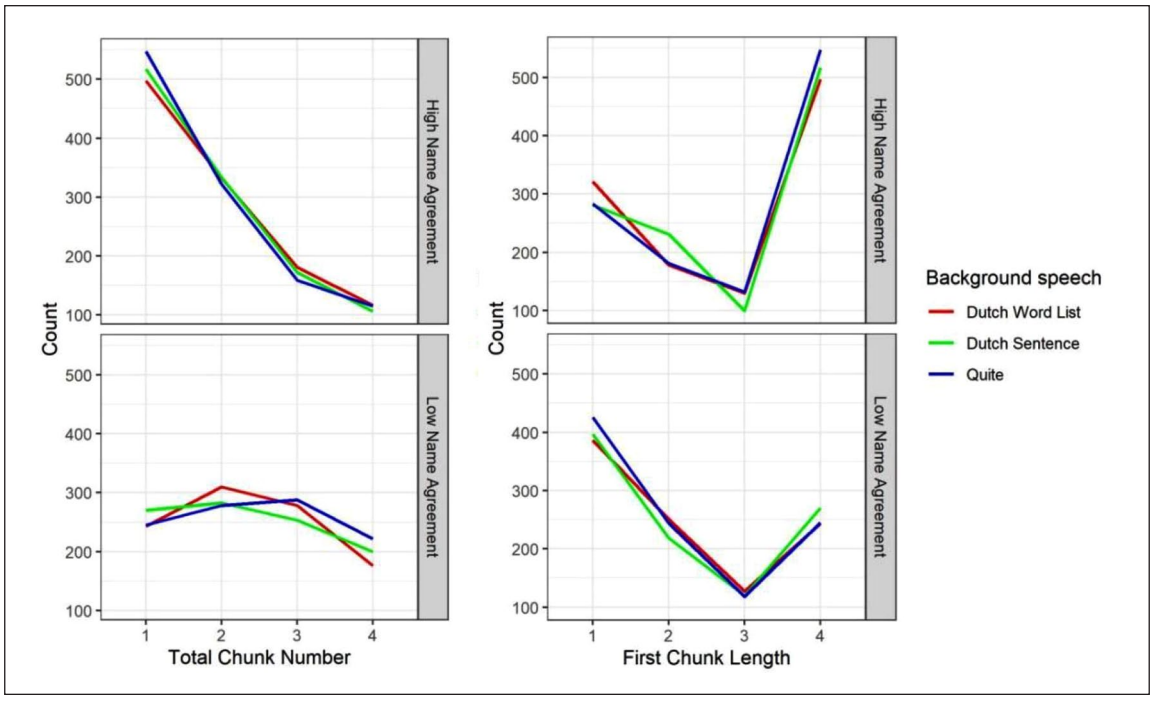

Figure 5. Total chunk number (left) and first chunk length (right) split by name agreement (NA: high, low) and the type of background speech (Dutch word list, Dutch sentence, Quiet) in Experiment 2.

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, with participants producing the names of low name agreement pictures more slowly, with more errors, and in shorter sets ("chunks") than high name agreement pictures. Second, irrelevant background speech in Experiment 1 (Chinese, unintelligible to speakers) and Experiment 2 (Dutch, intelligible to speakers) always disrupted spoken word production relative to a quiet condition. This patterned as increased articulation time and onset latencies in Experiment 1 (Chinese background speech), and increased articulation time, onset latencies, and utterance duration in Experiment 2 (Dutch background speech). Third, no systematic difference between word lists and sentences was found in either experiment. Finally, there were differences in how the two types of irrelevant speech effects were modulated by the difficulty of speech production: the disruptive effects of Dutch background speech in Experiment 2 were strongest when high name agreement pictures were named.

The effect of name agreement (indexing lexical selection demands in production) was remarkably consistent on all measures and experiments (also see Supplementary Material, Table C1), replicating earlier work (e.g., Alario et al., 2004; He et al., 2021; Shao et al., 2014). The name agreement effects on time measures (onset latencies, utterance duration, total pause time, and articulation time) are noteworthy because they show how the demand of lexical selection affects processing before and after speech onset. This finding suggests that speakers retrieve picture names during the whole process of planning a sequence of picture names, indicative of incremental speech planning during which speakers have to coordinate the planning and articulation of successive words (e.g., Levelt et al., 1999; Roelofs, 1998; Wheeldon & Lahiri, 1997). Moreover, the finding that name agreement affected response chunking measures (total pause time, first chunk length) 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 with high name agreement.

In both experiments, irrelevant speech consistently increased onset latencies and articulation time relative to a quiet control condition, which is in line with the phonological disruption view (Salamé & Baddeley, 1982, 1989) under the framework of the interference-by-similarity account (e.g., Hughes, 2014; Jones et al., 1993; also the crosstalk account, Pashler, 1994). This view predicts that any background speech (whether it is intelligible or not) should disrupt speech production due to the similarity of phonological codes between the focal task and background speech. Since Dutch speech (Experiment 2) did not cause more disruption than Chinese speech (Experiment 1) during initial planning and articulation processes (see Supplementary Material, Table C1), our results further

argue against the importance of semantic similarity in disrupting speech planning.

Combined with earlier results from He, Meyer and Brehm (2021) who showed that word lists (regardless of intelligibility) interfered with onset latencies relative to a speech-like noise condition (i.e., eight-talker babble), these results also argue against the contribution of low-level acoustic properties shared between speech production and speech-like noise. Thus, these results are most in line with the phonological disruption view (Salamé & Baddeley, 1982, 1989).

We also found that Dutch but not Chinese background speech had a weak effect on utterance duration. This is consistent with He, Meyer and Brehm (2021), where Dutch word lists increased utterance duration relative to Chinese word lists, indicating that intelligible background speech elicits more disruption than unintelligible background speech. This suggests that intelligible background speech specifically interferes with the planning that is done between producing chunks of words, where a speaker needs to multi-task between speaking, planning, and listening. The extra disruption on utterance duration may result from similarity in semantics and/or phonology, or from an attention capture mechanism; further research would be needed to disentangle these possibilities.

In contrast to robust differences between background speech and quiet conditions, we did not observe any difference between the background word lists and sentences in either Experiment 1 or 2. The results of Experiment 1 suggest that the stimulus-aspecific variation of unintelligible background speech does not elicit disruption on spoken word production, which goes against the aspecific attention capture view but seems consistent with specific attention capture view (Eimer et al., 1996).

However, the specific attention capture view (Eimer et al., 1996) also predicts that in Experiment 2, Dutch sentences (richer syntactic/semantic representation) should disrupt spoken word production more than Dutch word lists (weaker syntactic/semantic representation). This was not the case: we did not find any difference between Dutch word lists and sentences on any measures in Experiment 2. This is consistent with three possibilities. First, the lack of a word lists versus sentences effect might be because the stimulus-specific effect indeed exists, but it was too small and attenuated by the repetition of stimuli, which all appeared 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 type of irrelevant background speech (word list vs. sentence) and block in either experiment (see Supplementary Material, Table A5 for Experiment 1; Table B3 for Experiment 2), which shows that there is no evidence any background speech effect changes with repetition. Another possibility, and one we deem more likely, is that the aspecific and specific effects may have cancelled each other out. In other words, the disruption by the presence of pauses (aspecific context variation) in Dutch word lists cancelled interference by richer linguistic information (specific linguistic variation) in Dutch sentences. This possibility could be pursued in future research with larger sources of stimulus-specific interference. Finally, it is possible that the manipulation of stimulus-aspecific variation in Experiment 2 was weak because the background speech stimuli were too uniform and boring (word lists had a regular acoustic pattern, sentences had uniform syntactic structure). Participants might adapt to the regular tempo of word lists and use a strategy to name pictures, which causes weaker interference than we predicted and results in the absence of word lists versus sentence effect. This possibility was supported by a follow-up study in He (2023, Chapter 5). This study directly manipulated the relative interestingness (boring vs. funny) of irrelevant background sentences, and found an interestingness effect such that boring sentences were more disruptive than funny sentences. This suggests that stimulus-aspecific variation in the present experiments could have been weak due to the relative uniformity of the stimuli, and also suggests that attention to background speech may be influenced by a wide variety of other factors.

Consistent with the predictions from the attention engagement account (Halin et al., 2014; Marsh et al., 2015), the interaction between background speech and name agreement was absent in Experiment 1 but present in Experiment 2 on the measure of total pause time. Disruption by Chinese background speech remained unaffected by changes in attention engagement manipulated by name agreement because the processing of unintelligible auditory input is automatic and escapes cognitive control (Hughes, 2014). In contrast, interference by Dutch background speech was reduced by increased attention engagement (on low name agreement), because the processing of intelligible background speech requires central attention that taps into cognitive control (Marsh et al., 2018). This is largely consistent with He, Meyer and Brehm (2021), though note that the effects appeared on total pause time in Experiment 2 but on onset latencies in He, Meyer and Brehm 2021. The inconsistency may be due to small effect sizes or to variations in the baseline task (quiet in the present study and eight-talker babble in He, Meyer and Brehm 2021) and the speech production task (naming four pictures in the present study and naming six pictures in He, Meyer and Brehm 2021). Future work is needed to determine the cause of the difference.

The fact that many facets of irrelevant background speech interfere with speech production leaves many possibilities for future work. We sketch some of these now. First, we saw clear evidence for the phonological but not

semantic disruption view (Martin et al., 1988; Salamé & Baddeley, 1982, 1989). To understand the nature of interference-by-similarity, more work should therefore be done that considers specific relationships (e.g., phonological, semantic) between speaking and background speech, thereby more cleanly assessing the role of shared representations in speaking-while-listening in a targeted way. Second, this study showed more evidence for specific than aspecific attention capture (Eimer et al., 1996), but could not cleanly distinguish between the two. Future comparisons integrating these two desiderata would be interesting. In particular, a further comparison between different types of irrelevant background speech matched closely on specific content and acoustic variation would be more informative about how two variants of attention capture (aspecific and specific) affect speech production performance in the presence of irrelevant background speech. Third, the non-continuous background speech in this study was regularly timed (with a consistent interval of 700 ms between words), which may have led to habituation effects over time. Future studies with irregular timing in the background speech would provide more clarity regarding the aspecific attention capture account. Fourth, the present research used a multi-object naming task that was relatively easy, and therefore not necessarily representative of typical speech production. Given the complex interplay between the demands of speaking, listening, and attention, it would be fruitful to expand this line of research into more naturalistic speech production tasks such as sentence or dialogue production and to assess whether other aspects of speech production difficulty (such as object recognition, phonological encoding, and phonetic encoding) show similar effects to lexical selection difficulty. Finally, this study mostly focused on the two accounts—the interference-by-similarity and the attention capture account—without considering other theoretical interpretations. Future research should consider alternative explanations for the irrelevant speech effect in speaking. For instance, the timing of the interference could have an effect, based on the results from some PWI studies (e.g., Glaser & Düngelhoff, 1984; Schriefers et al., 1990). This would inform us about other theories for irrelevant speech effect.

Conclusion

Two experiments using a speaking-while-listening paradigm showed that irrelevant background speech (regardless of its intelligibility) disrupts spoken word production relative to a quiet condition, and that intelligible background speech elicits further 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 list and sentence conditions in unintelligible background speech suggests that the aspecific properties of background speech (in this case, the presence of pauses) do not affect naming performance by diverting attention away from the task. Finally, while intelligible background speech had a larger impact on spoken word production, the impact can be reduced through greater engagement with the task, for example, 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.

Acknowledgements

This project was part of the doctoral work conducted by the first author under the supervision of A.S. Meyer at the Max Planck Institute for Psycholinguistics. The authors also 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.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Max Planck Society.

ORCID iDs

Jieying He https://orcid.org/0000-0002-2937-5100
Ava Creemers https://orcid.org/0000-0002-7566-0658

Supplemental material

The supplementary material is available at qjep.sagepub.com.

Data availability statement

All stimuli, participant data, and analysis scripts can be found on this paper's project page on the [OSF, https://osf.io/wuafh/].

Notes

1. Here is an example of Experiment 1 for one participant: https://frinexproduction.mpi.nl/image_naming_noise_cn/?stimulusList=List1.
2. This was a different speaker from the one who recorded Dutch words for attention check trials.
3. Here is an example of Experiment 2 for one participant: https://frinexproduction.mpi.nl/image_naming_noise_nl/?stimulusList=List1.

References

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

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in Cognitive Sciences*, *4*(11), 417-423. 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

Boersma, P., & Weenink, D. (2009). *Praat: Doing phonetics by computer* (Version 5.1. 05) [Computer program]. University of Amsterdam.

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. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

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

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

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/10.1016/S0022-5371(76)90003-7

Cowan, N. (1995). Verbal working memory: A view with a room. *American Journal of Psychology*, *108*(1995), 123–155. https://doi.org/10.2307/1423105

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., & 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

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

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), 101s–110s. https://doi.org/10.1097/aud.0000000000000300

Eimer, M., Nattkemper, D., Schröger, E., & Prinz, W. (1996). Involuntary attention. In O. Neumann & A. F. Sanders (Eds.), *Handbook of perception and action* (Vol. 3, pp. 155–184). Academic Press. 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/10.1016/j.actpsy.2012.02.009

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), Article 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), Article 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

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

Halin, N., Marsh, J. E., Hellman, A., Hellström, I., & Sörqvist, P. (2014). A shield against distraction. *Journal of Applied Research in Memory and Cognition*, *3*(1), 31–36. https://doi.org/10.1016/j.jarmac.2014.01.003

He, J. (2023). *Coordination of spoken language production and comprehension: How speech production is affected by irrelevant background speech*. Radboud University.

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/10.1016/j.wocn.2010.08.002

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., 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

Hyönä, J., & Ekholm, M. (2016). Background speech effects on sentence processing during reading: An eye movement study. *PLOS ONE*, *11*(3), Article e0152133. https://doi.org/10.1371/journal.pone.0152133

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.

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

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

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

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

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/10.1016/0749-596X(88)90063-0

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

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.

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

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

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., & 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*, 692–702. https://doi.org/10.1037/xhp0000028

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/10.1016/S0022-5371(82)90521-7

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

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

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

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

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*, *55*(1), 236–262. https://doi.org/10.3758/s13428-021-01768-2

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

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

Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, *37*(3), 356–381. 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