Title

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

Authors

He, Jieying
Frances, Candice
Creemers, Ava
et al.

Publication Date

2024-01-21

DOI

10.1177/17470218231219971

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

Jieying He[a,b*], Candice Frances[a], Ava Creemers[c], Laurel Brehm[a,d]

[a] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

[b] International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands

[c] Faculty of Humanities, University of Amsterdam, Amsterdam, The Netherlands

[c] Department of Linguistics, University of California, UC Santa Barbara, USA

[*] Correspondence concerning this article should be addressed to Jieying He, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. Email: Jieying.He@mpi.nl

This is a preprint of work accepted in QJEP (https://doi.org/10.1177/17470218231219971)

Abstract

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

Keywords: irrelevant speech effect, name agreement, speech production

1 Introduction

Much of daily conversation 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. It has been shown that both spoken language comprehension (e.g., Eckert et al., 2016; Vasilev et al., 2018) and production (e.g., He et al., 2021) receive interference from irrelevant background noise. However, less is known about how speakers plan their speech in the presence of irrelevant background speech than about how they listen in adverse conditions. Understanding speech production in verbal and non-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 speech production with varying lexical selection demands, and whether the influence was modulated by the difficulty of speech production.

1.1 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; 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 et al., 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 et al., 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).

Critically, two views from this literature on the source of the irrelevant speech effect make different predictions for the effect of background speech on speech production. The *phonological disruption view* (Salamé & Baddeley, 1982, 1989) hypothesizes 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 accounts,* the *domain-general attention capture* view 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. Results showing that irrelevant background speech interferes with serial recall performance (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015) and reading (Hyönä and Ekholm, 2016) support the attention capture theory.

There is a similar divide within this domain-general attention capture view with different predictions of the effects of irrelevant background effects 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).

1.2 Irrelevant speech effects in spoken language production

This 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 interference-by-similarity (Martin et al., 1998; Salamé & Baddeley, 1982, 1989) and attention capture (Buchner et al., 2004; Cowan, 1995; Elliott & Briganti, 2012; Röer et al., 2013, 2015) mechanisms may play roles in the disruption by irrelevant background speech on speech production. 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.

An earlier study by He et al. (2021) supports the role of multiple attention-capturing properties 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 speech production more than eight-talker babble, and Dutch caused more disruption than Chinese word lists. This suggests that more interference on speech 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., 1998; Salamé & Baddeley, 1982, 1989). However, He et al. (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 et al. (2021) also manipulated the difficulty of speech production by varying name agreement (high, low) of to-be-named pictures. Name agreement is the extent to which participants agree on the name of a picture. Previous studies have found that naming a picture with high name agreement (e.g., the item called *banana*) is faster and more accurate than naming one with low name agreement (e.g., the item called *sofa* or *couch;* e.g., Alario et al., 2004; Cheng et al., 2010; Vitkovitch & Tyrell, 1995; Shao et al., 2014). 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 et al. (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 in order to study irrelevant speech effects in speaking, it is also important to consider the production demands.

1.3 Current study

The present study was designed to explore how different types of irrelevant background speech affected spoken language production. Two experiments 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 speech 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 the same results across experiments. By contrast, the semantic disruption view (Martin et al., 1998) predicts that only intelligible background speech should interfere with speech production, predicting more 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 versus 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., 2011; Vitkovitch & Tyrell, 1995; Shao et al., 2014), we predicted that pictures with low name agreement would be named more slowly than those with high name agreement 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 to high name agreement pictures. This is because stimulus-specific disruption requires central attention that taps into cognitive control (Hughes, 2014; Marsh et al., 2018).

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 speech production relative to a quiet condition.
Semantic disruption view (Martin et al., 1998)	Chinese speech (in Exp1) should not disrupt speech 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.

2.1 Experiment 1

2.1.1 Methods

Participants

We recruited 50 native speakers of Dutch who had little Chinese experience, (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-inch or larger laptops (range: 14-24 inches) with Google Chrome, Firefox, Microsoft Edge, or Brave web browsers. Each participant's speech was recorded by a built-in voice recorder 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. 240 pictures from He et al., (2021, Experiment 2; pictures selected from the MultiPic database, Duñabeitia et al., 2018; see Appendix A, Table A1) were used in the present study. Of these, 120 were high name agreement pictures, all with 100% name agreement , 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 neighborhood frequency (PNF), phonological neighborhood size (PNS), orthographic neighborhood frequency (ONF), and orthographic neighborhood size (ONS).

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

On each trial of the experiment, four pictures, all with high name agreement or all with low name agreement, were presented simultaneously in a 1 × 4 grid (size: 10 cm × 40 cm). The pictures per grid were 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. Twenty-four additional pictures (6 picture grids) were selected as practice stimuli from the same database.

Irrelevant background speech. For the Chinese word list condition (see Appendix A, 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 ($t_{(305.91)} = -1.58$, $p > 0.05$). 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 44100 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 seconds and there were similar periods of silence (about 700 ms) between consecutive nouns.

For the Chinese sentence condition (see Appendix A, Table A3), the 20 Chinese word lists were transformed into 20 Chinese sentences by 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 by up to 9.59%) to last 8 seconds.

To check whether participants were listening to the background speech, 19 additional two-syllable Dutch nouns (4 for the practice 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) = -33.98dB) 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 randomized 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 program (van Casteren & Davis, 2006), with the constraints that word lists and sentences sharing the same nouns were presented at least every three trials, and attention check trials were presented at least every five trials.

Procedure

Participants were tested 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 ten trials (six test trials and four attention check trials) was followed by the 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

---

[1] Here is an example of the Experiment 1 for one participant: https://frinexproduction.mpi.nl/image_naming_noise_cn/?stimulusList=List1

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 Appendix A, Table A4). The experiment lasted about 30 minutes.

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

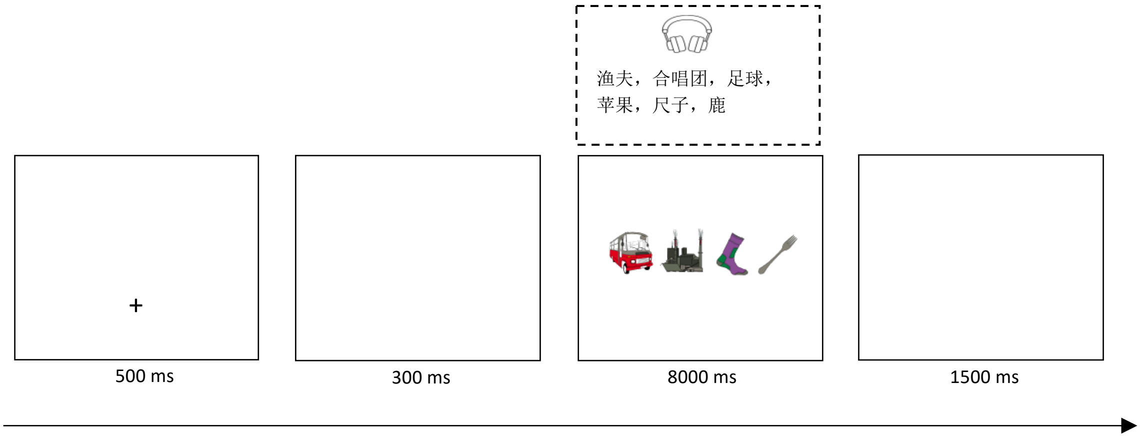

Figure 1. *An example trial in which participants named pictures with high name agreement while ignoring a Chinese word list (translation: fisherman, choir, football, apple, ruler, deer).*

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 where all pictures were named correctly and 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, 2018). Predictors were name agreement (high/low) and the type of background speech (Chinese word list/Chinese sentence/quiet). Name agreement (high/low) was contrast coded with (0.5, -0.5). Two contrasts were made for the type of background speech: the first was coded with (0.25, 0.25, -0.5) to compare the two Chinese speech conditions (word list and sentence) with the quiet condition, and the second was coded with (0.5, -0.5, 0) to compare the Chinese word list and Chinese sentence conditions. The random effect structure for the models included random intercepts for participants and items, and random slopes for name agreement and the type of background speech by participants and items. Separate models were fitted for each dependent measure. All models had four chains and each chain had 24000 iterations depending on model convergence (listed in model output tables). We used a warm-up (or burn-in) period of 2000 iterations in each chain, which means we removed the data based on the first 2000 iterations in order to correct the initial sampling bias.

All models used weak, widely spread priors that would be consistent with a range of null to moderate effects. The model of accuracy used family *bernoulli* combined with a *logit* link, with a student-*t* prior with 1 degree of freedom and a scale parameter of 2.5. The models of log-transformed onset latency, log-transformed utterance duration, and log-transformed articulation time used a weak normal prior with an SD of 0.2, and the model of log-transformed total pause time used a weak normal prior with an SD of 1. These models were performed using the family *gaussian* and *identity* link. Total chunk number and first chunk length had weak normal priors centered at zero with an SD of 1, and used family *possion* combined with the *log* link. All models were run until the R-hat value for each 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 to the uncertainty around it. We also reported the posterior probability of all weak effects, indicating the proportion of samples with a value equal to or above the beta estimate.

2.1.2 Results

Six participants were removed from further analyses: three did not run the experiments successfully due to a bad internet connection, two gave no responses on attention check trials, and one had too much Chinese experience as indicated by their responses on the Chinese experience questionnaire. The data from the remaining 44 participants was checked for errors, removing from analysis any trials with implausible names (e.g., *koekje* ‘cookie’ named as *virus*), hesitations (e.g., *komkommer* ‘cucumber’ named as *kom...komkommer*), self-corrections (e.g., *komkommer* ‘cucumber’ misnamed as *courgette...komkommer* ‘courgette...cucumber’), and any trials where objects were omitted or named in the wrong order. The exclusion of these inaccurate trials resulted in a loss of 13.7% of the data (range by participants: 1.1% - 30% of removed trials). Then, any onset latencies below 200 ms were removed from this analysis, resulting in a loss of 0.47% of the data. Any total pause times below 20 ms were also removed from this analysis, resulting in a loss of 12.98% of the data. Finally, any data points more than 2.5 standard deviations below or above the mean values were removed for each time measure (1.87% for log-transformed onset latency, 0.86% for log-transformed utterance duration, 0.97% for log-transformed total pause time, and 1.33% for log-transformed articulation time). Descriptive statistics appear in Table 2.

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)	1246 (462)	1279 (522)	1198 (408)	1434 (579)	1413 (539)	1345 (486)
Utterance duration (ms)	2868 (790)	2868 (771)	2791 (765)	3475 (1062)	3482 (1025)	3392 (970)
Total pause time (ms)	685 (621)	662 (590)	645 (582)	1078 (860)	1043 (790)	1040 (805)
Articulation time (ms)	2309 (431)	2332 (429)	2246 (392)	2518 7(498)	2536 (522)	2450 (476)
Total chunk number	1.9 (1.0)	1.9 (1.0)	1.9 (1.0)	2.3 (1.1)	2.4 (1.1)	2.4 (1.1)
First chunk length	2.7 (1.3)	2.7 (1.3)	2.8 (1.3)	2.3 (1.3)	2.2 (1.2)	2.2 (1.2)

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

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

Accuracy. Participants produced sensible responses on 86% of the naming trials. As shown in Table 3, a Bayesian mixed-effect model showed that accuracy was considerably lower for low name agreement pictures than high name agreement pictures (β= 0.099, SE = 0.025, 95% Cr.I = [0.051, 0.147]), but it was not influenced by the type of background speech. Name agreement and the type of background speech did not interact.

Onset latency. As shown in Table 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= -0.122, SE = 0.014, 95% Cr.I = [-0.149, -0.095]). There was moderate evidence for the first contrast (Chinese vs. Quiet) of background speech, showing that the log-transformed onset latencies in the two Chinese speech conditions (word list and sentence) were slower than in the quiet condition (β= 0.064, SE = 0.038, 95% Cr.I = [-0.011, 0.138]). Note that while the 95 % Cr.I contains zero, the point estimate is high relative to the error around it, and 96% of the posterior distribution around the estimated effect is above zero. Name agreement and the type of background speech did not interact.

Utterance duration. As shown in Table 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= -0.191, SE = 0.02, 95% Cr.I = [-0.231, -0.151]), but it was not influenced by the type of background speech. Again, name agreement and the type of background speech did not interact.

Total pause time. As shown in Table 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, -0.460]), but it did not vary with the type of background speech. Name agreement and the type of background speech did not interact.

Articulation time. As shown in Table 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= -0.085, SE = 0.02, 95% Cr.I = [-0.125, -0.046]), and it was reliably longer in the two Chinese speech conditions (word list and sentence) than in the quiet condition (β= 0.038, SE = 0.014, 95% Cr.I = [0.01, 0.066]). Again, name agreement did not interact with the type of background speech.

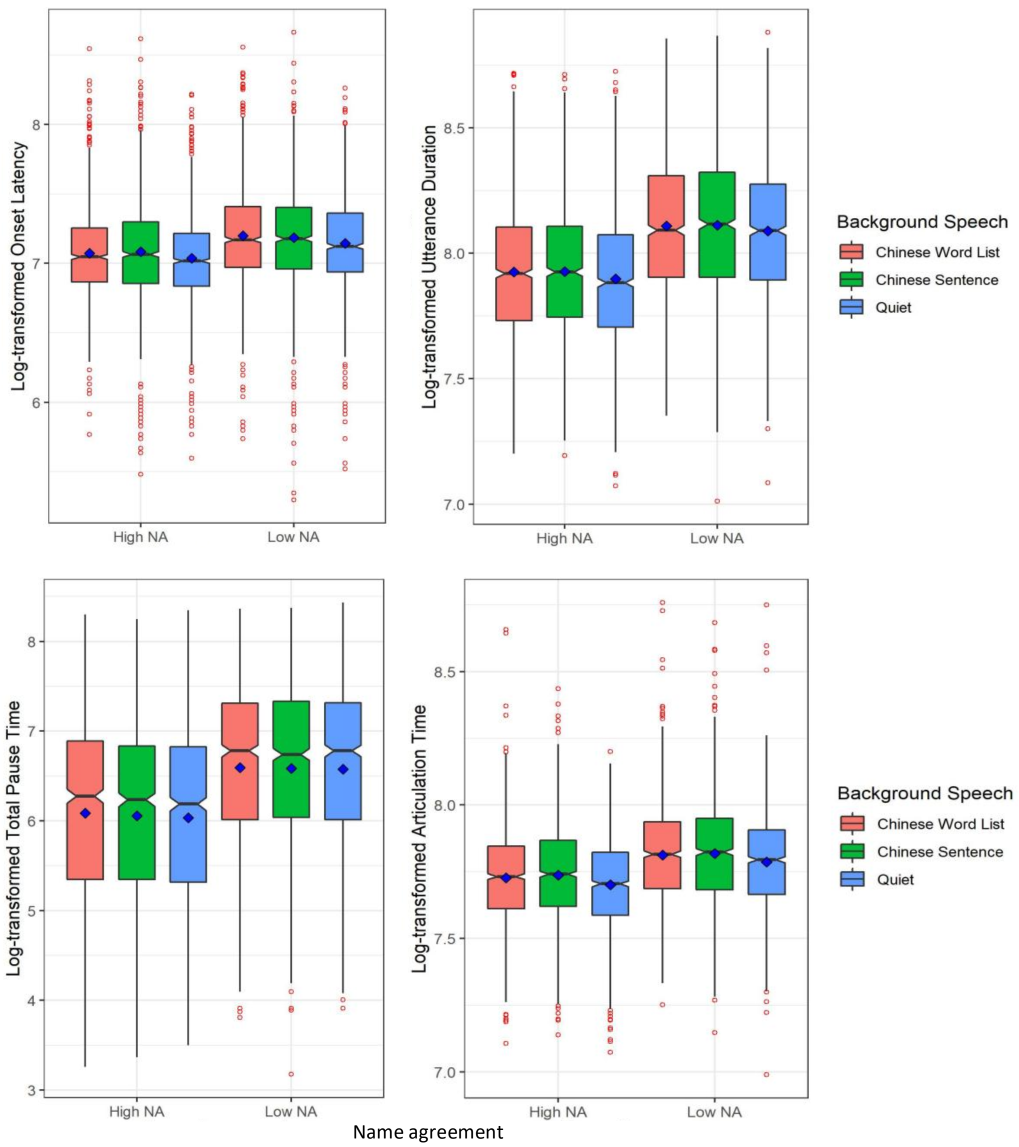

Figure 2. *Log-transformed Onset latency (top left), log-transformed utterance duration (top right), log-transformed total pause time (bottom left), and log-transformed articulation time (bottom right) split by name agreement (NA: high, low) and the type of background speech (Chinese word list, Chinese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

Total chunk number. As shown in Table 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= -0.241, SE = 0.022, 95% Cr.I = [-0.284, -0.197]). 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= 0.209, SE = 0.024, 95% Cr.I = [0.162, 0.256]). First chunk length was not affected by the type of background speech and there was no interaction between name agreement and the type of background speech.

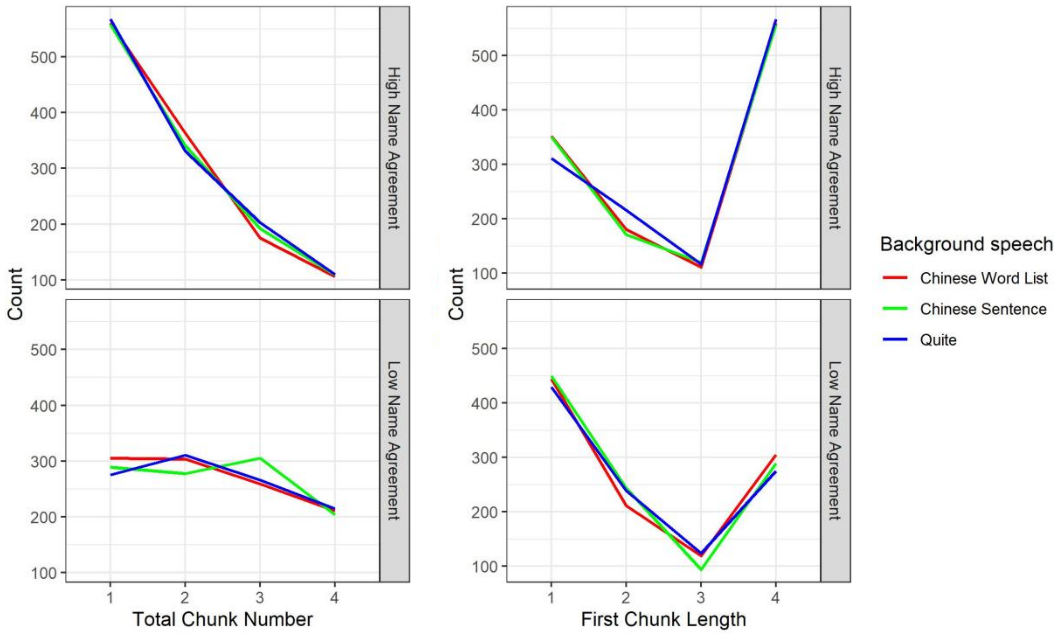

Figure 3. *Total chunk number (left) and first chunk length (right) split by name agreement (NA: high, low) and the type of background speech (Chinese word list, Chinese sentence, Quiet) in Experiment 1.*

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

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

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

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

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

	sd(Intercept)	0.256	0.030	0.206	0.321	15391
	sd(NA)	0.062	0.021	0.020	0.104	46312
	sd(Svs.Q)	0.023	0.018	0.001	0.067	62627
	sd(WLvs.S)	0.020	0.016	0.001	0.058	63929
	sd(NA×(Svs.Q))	0.049	0.037	0.002	0.139	64075
	sd(NA×(WLvs.S))	0.043	0.033	0.002	0.122	61696
	Items					
	sd(Intercept)	0.035	0.020	0.002	0.073	8804
	sd(NA)	0.070	0.040	0.004	0.146	7966
	sd(Svs.Q)	0.124	0.058	0.012	0.229	9285
	sd(WLvs.S)	0.102	0.043	0.014	0.183	13656
	sd(NA×(Svs.Q))	0.246	0.116	0.020	0.458	9163
	sd(NA×(WLvs.S))	0.202	0.087	0.025	0.365	13743
First chunk length						
	Intercept	0.863	0.042	0.781	0.946	11967
	Name Agreement	**0.218**	**0.025**	**0.168**	**0.268**	**96798**
Population-level effects	Speech vs. Quiet	-0.012	0.034	-0.077	0.055	95932
	Word List vs. Sentence	0.013	0.030	-0.046	0.072	92168
	NA × (S vs. Q)	-0.030	0.067	-0.162	0.101	95948
	NA × (WL vs. S)	-0.027	0.060	-0.145	0.091	95897
	Participants					
	sd(Intercept)	0.262	0.031	0.210	0.330	19220
	sd(NA)	0.022	0.016	0.001	0.061	50297
	sd(Svs.Q)	0.025	0.019	0.001	0.069	64357
Group-level effects	sd(WLvs.S)	0.023	0.018	0.001	0.065	61516
	sd(NA×(Svs.Q))	0.047	0.036	0.002	0.135	64675
	sd(NA×(WLvs.S))	0.043	0.033	0.002	0.122	63963
	Items					
	sd(Intercept)	0.047	0.025	0.003	0.090	5967

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

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

2.1.3 Interim Discussion

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

2.2 Experiment 2

Experiment 1 demonstrated clear phonological disruption and specific attention capture effects on unintelligible background speech. However, these patterns may not generalize 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., 1998; 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.

2.2.1 Methods

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 Appendix C, Table C1), the 120 nouns from Experiment 1 were used in Dutch, and matched with picture names on word frequency, number of syllables, number of phonemes, age-of-acquisition, and word prevalence. To pair with the set of 20 picture grids, these 120 Dutch nouns were divided into 20 word lists of 6 nouns, each list matched on word frequency and number of syllables. To equate the amount of semantic and phonological overlap across trials between speech planning and auditory background speech, we made sure that 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 the 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 seconds each with similar silent periods (about 700 ms) between consecutive nouns, by stretching by up to 9.38%.

For the Dutch sentence condition (see Appendix C, Table C2), the 20 Dutch word lists were transformed into 20 Dutch sentences 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 seconds

---

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

each by stretching by up to 14.29%. The same 19 attention catch trials (15 as test stimuli, 4 as practice stimuli) from Experiment 1 were also included. All auditory files were matched on intensity (total RMS = -33.98dB) in Adobe Audition.

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].

Analysis

The analysis was the same as Experiment 1.

2.2.2 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 was 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 standard deviations below or above the mean values were removed for the time measures (1.61% for log-transformed onset latency, 0.85% for log-transformed utterance duration, 1.01% for log-transformed total pause time, and 1.18% for log-transformed articulation time).

---

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

Descriptive statistics of all dependent variables are shown in Table 4.

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

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

Attention Check. The mean accuracy for attention check responses was 98% (range by participants: 73% - 100%), showing that participants indeed 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 (β= 0.921, SE = 0.195, 95% Cr.I = [0.543, 1.309]), 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 (β= -0.128, SE = 0.014, 95% Cr.I = [-0.155, -0.1]). There was moderate evidence for the first contrast of background speech (Dutch speech vs. Quiet), such that the log-transformed onset latencies in the two Dutch speech conditions (word list and sentence) were slower than in the quiet condition (β= 0.076, SE = 0.04, 95% Cr.I = [-0.003, 0.155]). 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 (β= -0.216, SE = 0.018, 95% Cr.I = [-0.251, -0.182]). There was moderate evidence for the first contrast of background speech (Dutch speech vs. Quiet), such that the log-transformed utterance durations in the two Dutch speech conditions (word list and sentence) were slower than in the quiet condition (β= 0.076, SE = 0.04, 95% Cr.I = [-0.003, 0.155]). 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= -0.599, 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= 0.28, 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= 0.394, SE = 0.171, 95% Cr.I = [0.058, 0.727]), but not for low name agreement pictures.

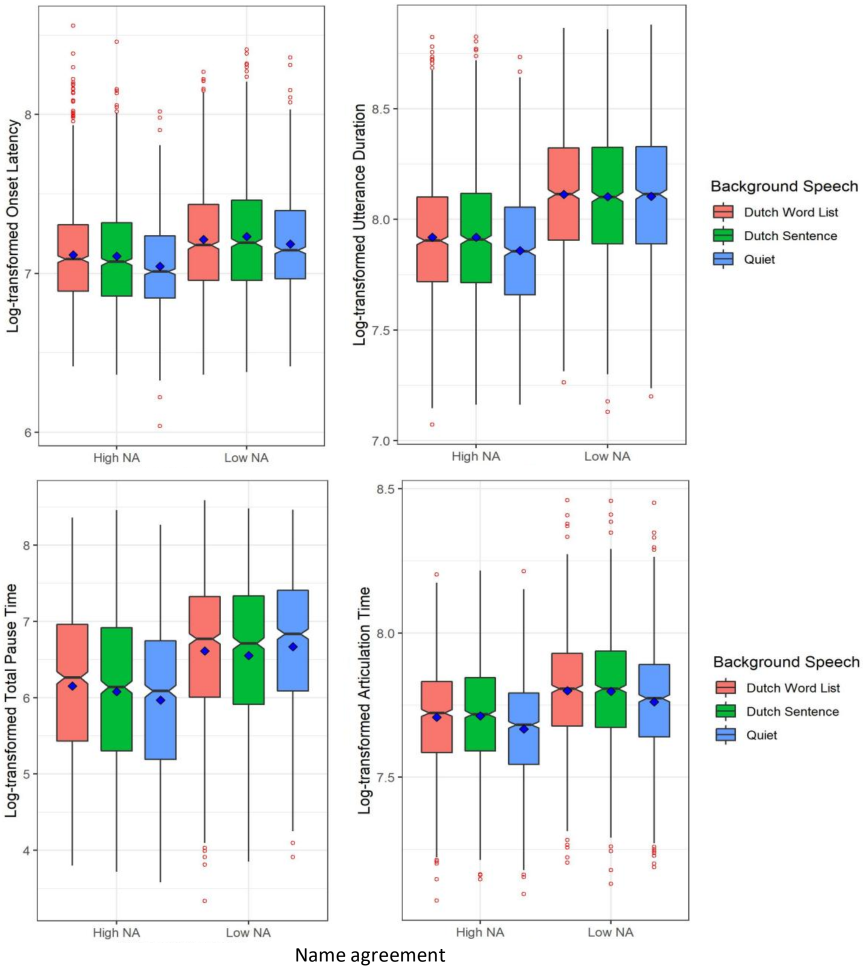

Figure 4. *Log-transformed onset latency (top left), log-transformed utterance duration (top right), log-transformed total pause time (bottom left), and log-transformed articulation time (bottom right) split by name agreement (NA: high, low) and the type of background speech (Dutch word list, Dutch sentence,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

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= 0.047, SE = 0.009, 95% Cr.I = [0.03, 0.064]), and articulation time was longer in the two Dutch speech conditions (word list and sentence) than in the quiet condition (β= 0.047, SE = 0.009, 95% Cr.I = [0.03, 0.064]). 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= -0.254, SE = 0.026, 95% Cr.I = [-0.306, -0.204]). 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= 0.228, SE = 0.025, 95% Cr.I = [0.178, 0.278]). First chunk length was not impacted by the type of background speech. Again, name agreement did not interact with the type of background speech.

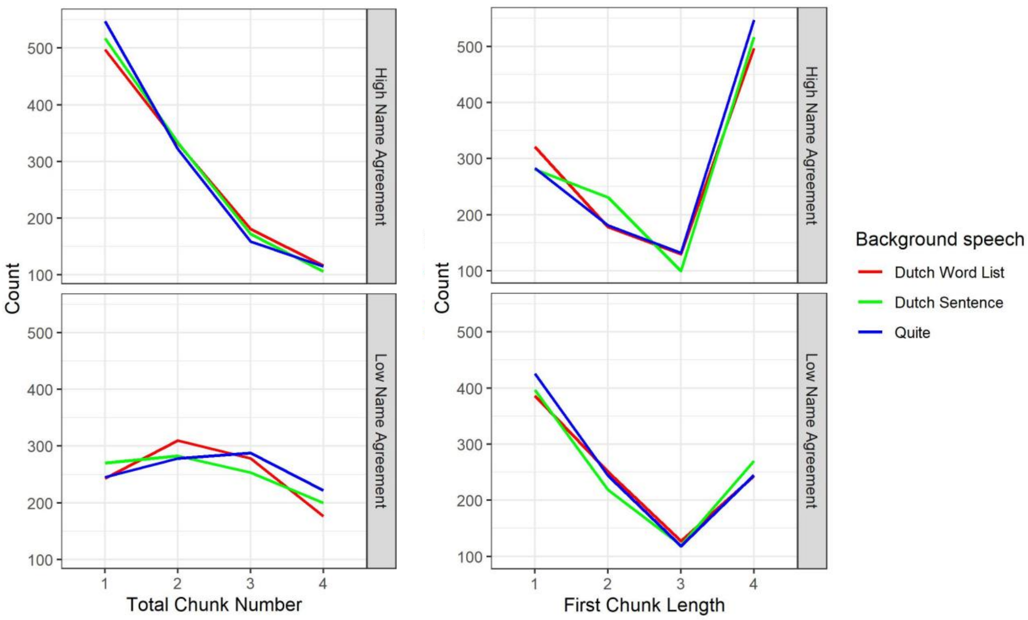

Figure 5. *Total chunk number (left) and first chunk length (right) split by name agreement (NA: high, low) and the type of background speech (Dutch word list, Dutch sentence, Quiet) in Experiment 2.*

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

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

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

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

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

	sd(Intercept)	0.246	0.030	0.196	0.312	15554
	sd(NA)	0.086	0.022	0.045	0.132	47199
	sd(Svs.Q)	0.024	0.019	0.001	0.070	62041
	sd(WLvs.S)	0.020	0.015	0.001	0.057	68947
	sd(NA×(Svs.Q))	0.051	0.040	0.002	0.148	61109
	sd(NA×(WLvs.S))	0.040	0.031	0.002	0.114	70155
Group-level effects	Items					
	sd(Intercept)	0.047	0.026	0.002	0.092	4816
	sd(NA)	0.094	0.052	0.005	0.184	4829
	sd(Svs.Q)	0.140	0.066	0.012	0.257	7236
	sd(WLvs.S)	0.102	0.057	0.005	0.204	6819
	sd(NA×(Svs.Q))	0.278	0.132	0.023	0.512	7343
	sd(NA×(WLvs.S))	0.201	0.114	0.010	0.407	6661
First chunk length						
	Intercept	0.858	0.045	0.767	0.948	8363
	Name Agreement	**0.237**	**0.027**	**0.183**	**0.291**	**74876**
Population-level effects	Speech vs. Quiet	-0.008	0.043	-0.092	0.076	64681
	Word List vs. Sentence	-0.022	0.036	-0.093	0.048	70214
	NA × (S vs. Q)	-0.090	0.085	-0.257	0.078	65380
	NA × (WL vs. S)	-0.005	0.072	-0.146	0.137	70142
	Participants					
	sd(Intercept)	0.272	0.034	0.214	0.346	17057
	sd(NA)	0.030	0.021	0.001	0.079	35240
	sd(Svs.Q)	0.026	0.019	0.001	0.073	58663
Group-level effects	sd(WLvs.S)	0.021	0.016	0.001	0.060	67790
	sd(NA×(Svs.Q))	0.059	0.044	0.002	0.164	54199
	sd(NA×(WLvs.S))	0.040	0.031	0.002	0.115	72032
	Items					
	sd(Intercept)	0.050	0.027	0.003	0.095	4599

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

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

2.2.3 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 presenceof 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 et al. (2021) and suggests that stronger attentional engagement in the more difficult low name agreement condition leads to less interference from background speech.

3 General Discussion

In two experiments, we explored how different types of unintelligible (Experiment 1) and intelligible (Experiment 2) background speech affected spoken language production, with a focus on their impact on lexical selection in speech planning. There were four major findings. First, we obtained consistent name agreement effects on all measures in both experiments, 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 speech 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 background speech 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 Appendix E, Table E1), 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 to 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). 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) (see Appendix E, Table E1), our results further argue against the importance of semantic similarity in disrupting speech planning. Combined with earlier results from He et al., (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 et al. (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, 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 speech production, which goes against the aspecific attention capture view (Eimer et al., 1996) 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 speech 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 two 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 background speech type (word list versus sentence) and block in either experiment (see Appendices, Table B1 for Experiment 1; Table D1 for Experiment 2), which shows that there is no evidence any background speech effect changes with repetition. An alternate possibility, and one we deem more likely, is that the aspecific and specific effects may have canceled each other out. In other words, the disruption by the presence of pauses (aspecific context variation) in Dutch word lists canceled 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) and susceptible to habituation effects over time. This possibility was supported by a follow-up study in He (2023, Chapter 6). This study directly manipulated the relative interestingness (boring versus 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 et al., (2021), though note that the effects appeared on total pause time in Experiment 2 but on onset latencies in He et al., (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 et al., 2021) and the speech production task (naming four pictures in the present study and naming six pictures in He et al., 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 open 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., 1998; 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 matchedclosely 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. Finally, 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.

4 Conclusion

Two experiments using a speaking-while-listening paradigm showed that irrelevant background speech (regardless of its intelligibility) disrupts speech production relative to a quiet condition, and that intelligible background speech elicits further disruption, due to its intelligibility. 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 may not capture attention and cause a drop on naming performance. Finally, while intelligible background speech had a larger impact on speech production, the impact can be reduced through greater engagement with the task, e.g., increasing the difficulty of speech production. The implication is that when the disruption by background speech occurs in speech production, speakers may be able to manage this disruption by changing when and how they plan their speech.

Acknowledgments

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

References

Ball, L. J., Threadgold, E., Solowiej, A., & Marsh, J. E. (2018). Can Intrinsic and Extrinsic Metacognitive Cues Shield Against Distraction in Problem Solving? *Journal of cognition*, *1*(1), 15. https://doi.org/10.5334/joc.9

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1. 05)[Computer program]. Retrieved May 1, 2009.

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

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

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

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*(Suppl 1), 101s-110s. https://doi.org/10.1097/aud.0000000000000300

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

5822(96)80022-3

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

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/https://doi.org/10.1016/j.jarmac.2014.01.003

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. (In prep). Effects of irrelevant unintelligible and intelligible background speech on spoken language produtcion.

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

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

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., Mac Harries, W. J. k., amp, & C. (1997). Disruption of Short term Recognition Memory for Tones: Streaming or Interference? *The Quarterly Journal of*

Experimental Psychology Section A, *50*(2), 337-357. https://doi.org/10.1080/713755707

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

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

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

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

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

Nairne, J. S. (1990). Similarity and long-term memory for order. *Journal of Memory and Language*, *29*(6), 733-746. https://doi.org/https://doi.org/10.1016/0749-596X(90)90046-3

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational Ideas—Part II. *Language and Linguistics Compass*, *10*(11), 591-613.

Perham, N., Marsh, J. E., & Jones, D. M. (2009). Short Article: Syntax and Serial Recall: How Language Supports Short-Term Memory for Order. *Quarterly Journal of Experimental Psychology*, *62*(7), 1285-1293.

https://doi.org/10.1080/17470210802635599

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. (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

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. (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

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

Salamé, P., & Baddeley, A. (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

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/https://doi.org/10.1016/j.brainres.2014.07.009

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

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

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

Wheeldon, L., & Lahiri, A. (1997). Prosodic Units in Speech Production. *Journal of Memory and Language*, *37*(3), 356-381. https://doi.org/https://doi.org/10.1006/jmla.1997.2517

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

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

Appendices

Appendix A: Stimuli used in Experiment 1

Table A1. 240 pictures used in both Experiments.

Picture Grid	*Picture 1*	*Picture 2*	*Picture 3*	*Picture 4*	*Picture Grid*	*Picture 1*	*Picture 2*	*Picture 3*	*Picture 4*
Pictures with high name agreement									
1	koelkast	pijl	dolfijn	gevangenis	16	spiegel	ananas	robot	zaklamp
2	leeuw	kruiwagen	driehoek	tomaat	17	schilderij	tunnel	kangoeroe	broek
3	harp	radio	knie	paprika	18	sleutel	dobbelsteen	ketting	rechter
4	vlinder	trap	cactus	batterij	19	stopcontact	arm	ezel	diamant
5	zaag	kiwi	vliegtuig	bezem	20	kapper	zebra	aardbei	wolk
6	waaier	schaap	glas	baard	21	schaduw	kompas	geit	horloge
7	ster	konijn	doedelzak	handschoen	22	pompoen	vlieger	kaars	skelet
8	pijp	hamer	berg	duim	23	heks	aardappel	vleermuis	boog
9	eekhoorn	keuken	banaan	orkest	24	masker	bijbel	zwembad	kanon
10	kwal	slager	anker	vuist	25	schaar	rups	kraan	puzzel
11	microfoon	bloem	koning	stier	26	eiland	schildpad	clown	bril
12	kokosnoot	steen	gitaar	egel	27	fruit	vlag	aansteker	lepel
13	roos	trechter	kroon	ballon	28	kikker	wasmachine	bokser	trompet

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

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

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

Table A3. 20 Chinese sentences used in Experiment 1.

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

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

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

---

1) Ben je in een land geweest waar Mandarijn Chinees wordt gesproken? Zo ja, hoeveel maanden?
A. Nooit B. <3 maanden C. 3-6 maanden D. 6-12 maanden E. >12 maanden

2) Ben je bij een gezin geweest waar Mandarijn Chinees wordt gesproken? Zo ja, hoeveel maanden?
A. Nooit B. <3 maanden C. 3-6 maanden D. 6-12 maanden E. >12 maanden

3) Ben je in een school/werkomgeving geweest waar Mandarijn Chinees wordt gesproken? Zo ja, hoeveel maanden?
A. Nooit B. <3 maanden C. 3-6 maanden D. 6-12 maanden E. >12 maanden

4) Gebruik onderstaande schaal, waar 0 "helemaal geen kennis" is, en 10 "vloeiend, alsof het je moedertaal is". Geef aan wat jouw vaardigheidsniveau is op het gebied van het spreken, verstaan en lezen van Mandarijn Chinees.
A. Spreken van Mandarijn Chinees: 0 1 2 3 4 5 6 7 8 9 10

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

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

5) Gebruik onderstaande schaal, waar 0 "helemaal geen kennis" is, en 10 "vloeiend, alsof het je moedertaal is". Geef aan in hoeverre je op dit moment blootgesteld wordt aan Mandarijn Chinees in de volgende situaties.
A. Contact hebben met Chinese vrienden: 0 1 2 3 4 5 6 7 8 9 10

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

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

D. Lezen van Chinese boeken/tijdschriften: 0 1 2 3 4 5 6 78 9 10

Appendix B: Results of block analysis in Experiment 1

Table B1. Results of block analysis in Experiment 1.

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

	Block 1 vs. Block 2	**0.134**	**0.016**	**0.103**	**0.166**	**46638**
	NA × (S vs. Q)	0.015	0.024	-0.031	0.062	90001
	NA × (WL vs. S)	0.005	0.023	-0.041	0.051	80784
	NA × (Block 12 vs. 3)	**-0.101**	**0.025**	**-0.149**	**-0.052**	**87321**
	NA × (Block 1 vs. 2)	**-0.073**	**0.024**	**-0.12**	**-0.026**	**82973**
Population-level effects	(S vs. Q) × (Block 12 vs. 3)	-0.022	0.053	-0.125	0.083	63183
	(WL vs. S) × (Block 12 vs. 3)	-0.066	0.046	-0.156	0.025	65632
	(S vs. Q) × (Block 1 vs. 2)	0.031	0.049	-0.066	0.127	64491
	(WL vs. S) × (Block 1 vs. 2)	-0.029	0.04	-0.107	0.049	61714
	NA × (S vs. Q) × (Block 12vs.3)	0.033	0.096	-0.156	0.22	69797
	NA × (WL vs. S) × (Block 12vs.3)	-0.005	0.085	-0.171	0.163	73221
	NA × (S vs. Q) × (Block 1vs.2)	-0.048	0.09	-0.224	0.129	74468
	NA × (WL vs. S) × (Block 1vs.2)	0.03	0.073	-0.113	0.173	69539
Log-transformed total pause time						
	Intercept	5.019	0.291	4.447	5.59	4615
	Name Agreement	**-1.429**	**0.241**	**-1.904**	**-0.952**	**14775**
	Speech vs. Quiet	-0.428	0.238	-0.896	0.037	37330
	Word List vs. Sentence	-0.115	0.2	-0.505	0.278	45039
	Block 12 vs. Block 3	**1.131**	**0.22**	**0.699**	**1.562**	**29293**
	Block 1 vs. Block 2	**0.912**	**0.18**	**0.558**	**1.263**	**28534**
	NA × (S vs. Q)	-0.023	0.365	-0.74	0.7	68037
Population-level effects	NA × (WL vs. S)	0.137	0.348	-0.546	0.819	55847
	NA × (Block 12 vs. 3)	-0.05	0.419	-0.871	0.779	54403
	NA × (Block 1 vs. 2)	-0.214	0.302	-0.808	0.378	70396
	(S vs. Q) × (Block 12 vs. 3)	0.569	0.564	-0.544	1.676	57132
	(WL vs. S) × (Block 12 vs. 3)	-0.252	0.566	-1.361	0.864	55234
	(S vs. Q) × (Block 1 vs. 2)	0.118	0.475	-0.813	1.048	59261
	(WL vs. S) × (Block 1 vs. 2)	0.578	0.449	-0.309	1.458	55047
	NA × (S vs. Q) × (Block 12vs.3)	1.233	1.129	-0.994	3.441	48396
	NA × (WL vs. S) × (Block 12vs.3)	-0.12	1.101	-2.281	2.031	56935

NA × (S vs. Q) × (Block 1vs.2)	-0.75	0.935	-2.586	1.093	63045
NA × (WL vs. S) × (Block 1vs.2)	0.981	0.818	-0.619	2.586	59252

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

Appendix C: Stimuli used in Experiment 2

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

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

Table C2. 20 Dutch sentences used in Experiment 2.

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

Appendix D: Results of block analysis in Experiment 2

Table D1. Results of block analysis in Experiment 2.

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

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

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

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

Appendix E: Comparison of two experiments

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

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

	sd(NA × (WL vs. S))	0.381	0.211	0.016	0.687	1817
	sd(NA × Experiment)	0.037	0.021	0.002	0.075	1954
	sd((S vs. Q) × Experiment)	0.124	0.07	0.006	0.23	1526
	sd((WL vs. S) × Experiment)	0.131	0.058	0.012	0.225	3159
	sd(NA × (S vs. Q) × Experiment)	0.25	0.14	0.011	0.461	1548
	sd(NA × (WL vs. S) × Experiment)	0.258	0.117	0.023	0.446	3247
Log-transformed utterance duration						
	Intercept	8.016	0.019	7.979	8.053	4034
	Name Agreement	**-0.204**	**0.019**	**-0.24**	**-0.166**	**31359**
	Speech vs. Quiet	0.039	0.027	-0.014	0.093	38700
	Word List vs. Sentence	0.001	0.022	-0.043	0.045	37920
	Experiment	0.01	0.033	-0.054	0.075	3561
Population-level effects	NA × (S vs. Q)	0.045	0.053	-0.06	0.149	39293
	NA × (WL vs. S)	-0.001	0.044	-0.087	0.085	38949
	NA × Experiment	0.024	0.018	-0.011	0.059	21478
	(S vs. Q) × Experiment	*-0.02*	*0.015*	*-0.05*	*0.009*	*62382*
	(WL vs. S) × Experiment	-0.007	0.013	-0.032	0.017	69948
	NA × (S vs. Q) × Experiment	**-0.055**	**0.027**	**-0.109**	**-0.001**	**69610**
	NA × (WL vs. S) × Experiment	0.012	0.026	-0.038	0.062	65325
	Participant_sd (Intercept)	0.153	0.012	0.131	0.179	7187
	sd(Name Agreement)	0.067	0.007	0.054	0.081	28946
	sd(Speech vs. Quiet)	0.026	0.011	0.003	0.046	11714
	sd(Word List vs. Sentence)	0.008	0.005	0	0.019	33445
Group-level effects	sd(NA × (S vs. Q))	0.02	0.014	0.001	0.054	24533
	sd(NA × (WL vs. S))	0.023	0.014	0.001	0.053	22589
	Item_sd (Intercept)	0.041	0.022	0.002	0.074	1562
	sd(Name Agreement)	0.083	0.044	0.004	0.147	1599
	sd(Speech vs. Quiet)	0.139	0.054	0.023	0.225	2527
	sd(Word List vs. Sentence)	0.112	0.044	0.018	0.182	2874

	sd(Experiment)	0.018	0.009	0.001	0.035	7237
	sd(NA × (S vs. Q))	0.273	0.108	0.041	0.447	2380
	sd(NA × (WL vs. S))	0.226	0.087	0.039	0.365	2790
	sd(NA × Experiment)	0.035	0.019	0.002	0.07	7414
	sd((S vs. Q) × Experiment)	0.041	0.023	0.002	0.084	6087
	sd((WL vs. S) × Experiment)	0.04	0.021	0.002	0.08	5466
	sd(NA × (S vs. Q) × Experiment)	0.08	0.046	0.004	0.169	5992
	sd(NA × (WL vs. S) × Experiment)	0.081	0.043	0.005	0.16	5395
Log-transformed total pause time						
	Intercept	6.284	0.062	6.163	6.405	4174
	Name Agreement	**-0.589**	**0.055**	**-0.697**	**-0.481**	**26776**
	Speech vs. Quiet	0.031	0.072	-0.111	0.174	37500
	Word List vs. Sentence	0.037	0.06	-0.083	0.155	37909
	Experiment	-0.03	0.113	-0.252	0.19	3829
Population-level effects	NA × (S vs. Q)	0.163	0.142	-0.119	0.443	35595
	NA × (WL vs. S)	0.017	0.121	-0.219	0.255	37295
	NA × Experiment	0.026	0.064	-0.099	0.152	18480
	(S vs. Q) × Experiment	-0.045	0.059	-0.162	0.071	51571
	(WL vs. S) × Experiment	-0.05	0.052	-0.152	0.052	62542
	NA × (S vs. Q) × Experiment	**-0.234**	**0.112**	**-0.455**	**-0.012**	**63364**
	NA × (WL vs. S) × Experiment	0.037	0.106	-0.17	0.246	59726
	Participant_sd (Intercept)	0.514	0.041	0.441	0.603	7707
	sd(Name Agreement)	0.227	0.026	0.18	0.281	29906
	sd(Speech vs. Quiet)	0.101	0.041	0.016	0.177	13912
Group-level effects	sd(Word List vs. Sentence)	0.031	0.023	0.001	0.085	28697
	sd(NA × (S vs. Q))	0.112	0.073	0.005	0.27	16436
	sd(NA × (WL vs. S))	0.11	0.062	0.007	0.239	18382
	Item_sd (Intercept)	0.118	0.06	0.006	0.205	1575
	sd(Name Agreement)	0.217	0.123	0.01	0.406	1524

	sd(Speech vs. Quiet)	0.348	0.141	0.052	0.576	2218
	sd(Word List vs. Sentence)	0.289	0.124	0.031	0.487	2346
	sd(Experiment)	0.058	0.034	0.003	0.125	8725
	sd(NA × (S vs. Q))	0.678	0.283	0.09	1.14	2238
	sd(NA × (WL vs. S))	0.575	0.248	0.067	0.97	2335
	sd(NA × Experiment)	0.117	0.069	0.006	0.253	8968
	sd((S vs. Q) × Experiment)	0.153	0.085	0.009	0.318	6683
	sd((WL vs. S) × Experiment)	0.16	0.089	0.009	0.328	6183
	sd(NA × (S vs. Q) × Experiment)	0.292	0.17	0.015	0.628	6590
	sd(NA × (WL vs. S) × Experiment)	0.322	0.178	0.018	0.656	6527
Log-transformed articulation time						
Population-level effects	Intercept	7.757	0.015	7.727	7.786	4999
	Name Agreement	**-0.089**	**0.019**	**-0.127**	**-0.052**	**37001**
	Speech vs. Quiet	**0.046**	**0.014**	**0.018**	**0.074**	**49698**
	Word List vs. Sentence	-0.005	0.012	-0.029	0.019	45323
	Experiment	0.025	0.024	-0.021	0.073	3748
	NA × (S vs. Q)	0.01	0.028	-0.045	0.064	48524
	NA × (WL vs. S)	-0.002	0.024	-0.049	0.046	47017
	NA × Experiment	0.008	0.013	-0.017	0.033	18403
	(S vs. Q) × Experiment	-0.016	0.009	-0.034	0.003	52214
	(WL vs. S) × Experiment	-0.004	0.006	-0.016	0.008	72990
	NA × (S vs. Q) × Experiment	-0.002	0.014	-0.03	0.026	89838
	NA × (WL vs. S) × Experiment	-0.004	0.013	-0.028	0.021	88482
Group-level effects	*Participant_sd (Intercept)*	0.11	0.009	0.095	0.13	8141
	sd(Name Agreement)	0.053	0.005	0.044	0.063	21399
	sd(Speech vs. Quiet)	0.03	0.005	0.021	0.041	29762
	sd(Word List vs. Sentence)	0.007	0.004	0	0.015	26055
	sd(NA × (S vs. Q))	0.018	0.011	0.001	0.042	16427
	sd(NA × (WL vs. S))	0.014	0.01	0.001	0.036	16253

Item_sd (Intercept)	0.043	0.024	0.002	0.077	1422
sd(Name Agreement)	0.086	0.048	0.004	0.154	1456
sd(Speech vs. Quiet)	0.064	0.036	0.003	0.117	1607
sd(Word List vs. Sentence)	0.056	0.03	0.003	0.102	1895
sd(Experiment)	0.008	0.005	0	0.017	12710
sd(NA × (S vs. Q))	0.13	0.073	0.006	0.235	1537
sd(NA × (WL vs. S))	0.116	0.061	0.006	0.205	1857
sd(NA × Experiment)	0.016	0.009	0.001	0.034	14920
sd((S vs. Q) × Experiment)	0.01	0.007	0	0.028	33328
sd((WL vs. S) × Experiment)	0.01	0.007	0	0.027	25544
sd(NA × (S vs. Q) × Experiment)	0.02	0.015	0.001	0.056	30810
sd(NA × (WL vs. S) × Experiment)	0.02	0.014	0.001	0.054	26730

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