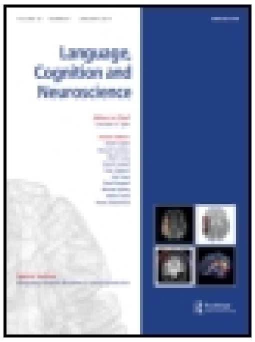

Routledge
Taylor & Francis Group

Language, Cognition and Neuroscience

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Concurrent listening affects speech planning and fluency: the roles of representational similarity and capacity limitation

Jieying He, Antje S. Meyer & Laurel Brehm

To cite this article: Jieying He, Antje S. Meyer & Laurel Brehm (2021): Concurrent listening affects speech planning and fluency: the roles of representational similarity and capacity limitation, Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2021.1925130

To link to this article: https://doi.org/10.1080/23273798.2021.1925130

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

View supplementary material

Published online: 13 May 2021.

Submit your article to this journal

Article views: 36

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=plcp21

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2021.1925130

Routledge
Taylor & Francis Group

REGULAR ARTICLE

OPEN ACCESS

Concurrent listening affects speech planning and fluency: the roles of representational similarity and capacity limitation

Jieying He[a,b], Antje S. Meyer[a,c] and Laurel Brehm[a]

[a]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; [b]International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands; [c]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

ABSTRACT

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

ARTICLE HISTORY
Received 22 August 2020
Accepted 23 April 2021

KEYWORDS
Dual-tasking; representational similarity; attention demand; spoken production

Introduction

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

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

Sources of interference in linguistic dual-tasking

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

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

CONTACT Jieying He Jieying.He@mpi.nl Max Planck Institute for Psycholinguistics, P.O. Box 310, Nijmegen, AH 6500, The Netherlands

Supplemental data for this article can be accessed https://doi.org/10.1080/23273798.2021.1925130

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

(Navon & Miller, 1987). In this paper, we use the term "representational similarity" to emphasize the role of shared representations between production and comprehension in eliciting interference.

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

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

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

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

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

Some aspects of understanding spoken language also require attention, especially for processes above the word level (Kristensen et al., 2013; Moisala et al., 2015). However, early word recognition processes may occur with little attentional engagement (Dupoux et al., 2003). For instance, dichotic listening studies, where participants are asked to attend to one source of information (e.g. a female voice) while ignoring another source (e.g. a male speaker), have shown that the unattended speech is nonetheless processed to some extent (Aydelott et al., 2015; Dupoux et al., 2003; Rivenez et al., 2006; Rivenez et al., 2008). This means speakers' goals (e.g. attend to or ignore speech input) matter to the comprehension of auditory information. If the speech input is irrelevant, attention (especially executive control; Posner & Rothbart, 2007) is needed to suppress its processing and focus on target task (Dupoux et al., 2003). By contrast, if the speech input is relevant to speaker's goals, attention needs to be divided between processing the speech input and the target task. Therefore, in Experiment 1 we explored how speech planning was influenced by the representational similarity between the irrelevant auditory input and planned speech, and in Experiment 2 we contrasted speech planning when the speech input was relevant versus irrelevant to the speakers' goals.

Flexible planning units in multi-word production

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

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

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

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

Current study

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

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

Whether because of code conflict or increased capacity demand for suppression, representational similarity is predicted to have a graded impact on interference in production. Auditory Dutch speech (*the high similarity condition*) overlaps with the representations used for production at multiple processing levels and should lead to increased capacity demand for their suppression and therefore to high levels of interference in production. In contrast, auditory Chinese speech (*the moderate similarity condition*) should only activate shared linguistic representations at the phonological/ phonetic level, requiring less capacity for suppression and leading to less interference. We contrasted these conditions with a language-like noise condition (eight-talker babble), which was Dutch-like in its acoustic properties only (*the low similarity condition*) and should lead to low capacity demand for suppression.

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

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

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

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

Experiment 1

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

Method

Participants

We recruited 21 native Dutch speakers (16 females) from the Max Planck Institute for Psycholinguistics' database. This sample size was selected because power simulations showed that 20 participants and 126 items would allow 99% power to measure a plausibly-sized interaction between name agreement and similarity on the onset latency measure. The interaction effect size used in the simulations was a name agreement effect of 50 ms or smaller (SD = 100 ms) in the eight-talker babble and Chinese conditions, but 100 ms or larger (SD = 100 ms) in the Dutch condition.[1] All participants were university students with a mean age of 22 years (range: 19–26) and reported normal or corrected-to-normal vision as well as no speech or hearing problems. They provided informed consent and received a payment of 6 € for their participation. The study was approved by the ethics board of the Faculty of Social Sciences of Radboud University.

Apparatus

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

Materials

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

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

On each trial of the experiment, six pictures, all with high name agreement or all with low name agreement, were presented simultaneously in a 2 × 3 grid (size:

20 cm × 30 cm). The pictures per grid were neither semantically related (i.e. they were from different semantic categories) nor phonologically related (i.e. avoiding the overlap of their 1st phonemes), as judged by a native speaker of Dutch. There were 14 grids for each set of pictures resulting in 42 grids in total. In addition, 36 additional pictures (6 grids) were selected from the same database as practice stimuli.

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

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

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

Design

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

Procedure

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

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

Analysis

Five dependent measures were coded to index interference in naming. Production *accuracy* indexed the proportion of trials where all six items were named with the correct responses. Picture names were coded as correct if they matched the first or second most common names given to the picture in the MultiPic database (Duñabeitia et al., 2018)[2], were synonymous to one of the two most common names (e.g. *laboratorium / lab*), or contained a diminutive version of one of the two most common names (e.g. *munt / muntje*), as judged by trained research assistants.

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

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

Accuracy, log-transformed onset latency, and log-transformed speech duration were analysed with mixed-effect models implemented using the *lme4* package (Bates et al., 2015) in R version 3.6.1 (R Core Team, 2018). Predictors were name agreement (high NA / low NA) and representational similarity (Dutch / Chinese / Babble). Name agreement (high NA / low NA) was contrast coded with (0.5, −0.5). For similarity, the first contrast was coded with (0.25, 0.25, −0.5) and compared the two language conditions (Dutch and Chinese speech) to language-like noise condition (eight-talker babble), and the second contrast was coded with (0.5, −0.5, 0) and compared Dutch with Chinese speech. The random effect structure in all models included random intercepts for participants and items. No random slopes were included because of convergence issues or evidence of model overfitting (high correlations between random terms). For the dependent measure of accuracy, a logistic mixed-effect model was fitted because of the binary nature of the responses. For the timing measures, separate linear mixed-effect models were fitted.

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

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

Results

Naming accuracy

Participants produced the intended responses on 65% of the naming trials. As shown in Tables 1 and 2, accuracy for high name agreement pictures was considerably higher than for low name agreement pictures ($\beta = 2.12$, $SE = 0.22$, $p < 0.001$), but did not vary by representational similarity. Name agreement and representational similarity did not interact.

Onset latency

As shown in Figure 1 (left), log-transformed onset latency was affected by name agreement and representational similarity. As supported by a linear mixed-effect model (see Table 2), it took participants reliably longer to plan names for low name agreement pictures than high name agreement pictures ($\beta = -0.12$, $SE = 0.03$, $p < 0.001$). Log-transformed onset latencies in the two language conditions (Dutch and Chinese) were reliably slower than in the eight-talker babble condition (β =

Table 1. Dependent measures in Experiment 1 by name agreement and representational similarity. For accuracy, range follows in parentheses, for other measures, standard deviation follows in parentheses.

	High name agreement			Low name agreement		
	Dutch	Chinese	Babble	Dutch	Chinese	Babble
Accuracy (%)	81 (57-95)	84 (43-100)	86 (57-100)	44 (19-67)	46 (19-76)	45 (19-76)
Onset latencies (ms)	1231 (577)	1101 (495)	973 (378)	1332 (582)	1231 (546)	1184 (427)
Speech durations (ms)	5295 (1453)	4732 (1206)	4673 (1236)	5963 (1690)	5593 (1433)	5544 (1499)
Total chunk number	3.1 (1.4)	2.8 (1.4)	2.6 (1.4)	3.5 (1.5)	3.5 (1.5)	3.2 (1.5)
First chunk length	2.7 (1.7)	3.1 (2.0)	3.4 (1.9)	2.5 (1.5)	2.3 (1.7)	2.6 (1.8)

Note. All timing and chunking measures reflect fully correct trials only.

0.15, SE = 0.02, $p < 0.001$), and log-transformed onset latencies were reliably slower in the Dutch speech than Chinese speech conditions (β = 0.09, SE = 0.02, $p < 0.001$). Name agreement and representational similarity interacted on the first contrast (β = 0.13, SE = 0.05, $p < 0.01$), showing that log-transformed onset latencies in the two language conditions were slower than in the eight-talker babble condition for high name agreement pictures (β = 0.08, SE = 0.01, $p < 0.001$), but this difference was not observed for low name agreement pictures.

Speech duration

As shown in Figure 1 (right), log-transformed speech duration was affected by name agreement and representational similarity. As supported by a linear mixed-effect model (see Table 2), log-transformed speech duration was reliably longer for low name agreement pictures than high name agreement pictures (β = −0.13, SE = 0.02, $p < 0.001$). Log-transformed speech durations in the two language conditions (Dutch and Chinese) were reliably longer than in the eight-talker babble condition (β = 0.08, SE = 0.02, $p < 0.001$), and log-transformed speech duration was reliably longer in the Dutch speech than Chinese speech conditions (β = 0.08, SE = 0.01, $p < 0.001$). Name agreement and representational similarity did not interact.[3]

Total chunk number

As shown in Figure 2 (left) and Table 1, total chunk number was affected by name agreement and representational similarity. As supported by an ordinal mixed

Table 2. Mixed-effect models for log-transformed onset latencies (Log-Onset), log-transformed speech durations (Log-Duration), accuracy, and chunk measures in Experiment 1.

	Fixed effects	Estimate	SE	*t* value	*p*
Log-Onset	Intercept	7.00	0.04	197.934	< 0.001
	NA (High vs. Low)	−0.12	0.03	−4.525	< 0.001
	Similarity ((Dutch & Chinese) vs. Babble)	0.15	0.02	6.722	< 0.001
	Similarity (Dutch vs. Chinese)	0.09	0.02	4.626	< 0.001
	NA×Similarity ((Dutch & Chinese) vs. Babble)	0.13	0.05	2.826	< 0.01
	NA×Similarity (Dutch vs. Chinese)	0.03	0.04	0.711	0.477
Log-Duration	Intercept	8.54	0.03	302.136	< 0.001
	NA (High vs. Low)	−0.13	0.02	−6.599	< 0.001
	Similarity ((Dutch & Chinese) vs. Babble)	0.08	0.02	4.792	< 0.001
	Similarity (Dutch vs. Chinese)	0.08	0.01	5.827	< 0.001
	NA×Similarity ((Dutch & Chinese) vs. Babble)	0.03	0.03	1.041	0.298
	NA×Similarity (Dutch vs. Chinese)	0.04	0.03	1.586	0.113
	Fixed effects	Estimate	SE	*z* value	*p*
Accuracy	Intercept	0.83	0.15	5.411	< 0.001
	NA (High vs. Low)	2.12	0.22	9.771	< 0.001
	Similarity ((Dutch & Chinese) vs. Babble)	−0.18	0.14	−1.272	0.203
	Similarity (Dutch vs. Chinese)	−0.18	0.12	−1.530	0.126
	NA×Similarity ((Dutch & Chinese) vs. Babble)	−0.40	0.28	−1.415	0.157
	NA×Similarity (Dutch vs. Chinese)	−0.14	0.23	−0.581	0.561
Total chunk number	NA (High vs. Low)	−0.09	0.02	−5.904	< 0.001
	Similarity ((Dutch & Chinese) vs. Babble)	0.37	0.11	3.344	0.001
	Similarity (Dutch vs. Chinese)	0.02	0.10	0.198	0.843
	NA×Similarity ((Dutch & Chinese) vs. Babble)	−0.01	0.03	−0.285	0.775
	NA×Similarity (Dutch vs. Chinese)	0.04	0.02	1.888	0.059
First chunk length	NA (High vs. Low)	0.35	0.07	4.825	< 0.001
	Similarity ((Dutch & Chinese) vs. Babble)	−0.32	0.08	−4.106	< 0.001
	Similarity (Dutch vs. Chinese)	0.01	0.07	0.191	0.848
	NA×Similarity ((Dutch & Chinese) vs. Babble)	−0.21	0.16	−1.322	0.186
	NA×Similarity (Dutch vs. Chinese)	−0.45	0.14	−3.302	<0.001

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

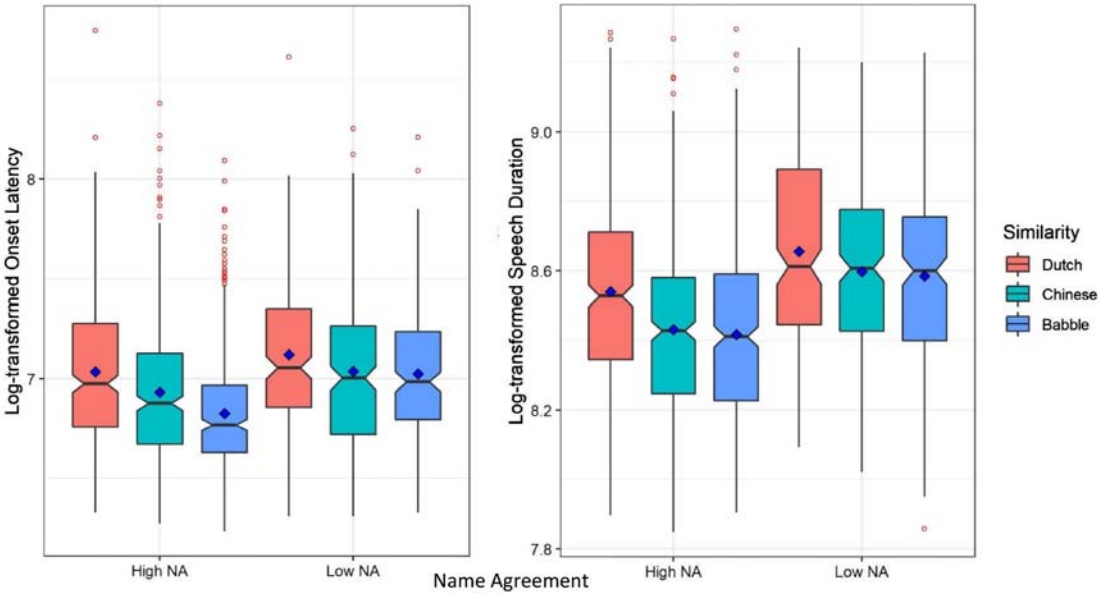

Figure 1. Log-transformed onset latencies (left) and log-transformed speech durations (right) in Experiment 1 split by representational similarity (Dutch speech, Chinese speech, eight-talker babble) and name agreement (NA; high, low). Blue squares represent condition means and red points reflect outliers.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

model (see Table 2), participants grouped their responses in more small chunks for low name agreement pictures than high name agreement pictures (β = −0.09, SE = 0.02, $p < 0.001$). Total chunk number was greater in the two language conditions (Dutch and Chinese) than in the eight-talker babble condition (β = 0.37, SE = 0.11, $p < 0.001$), but no difference between the Dutch and Chinese conditions was observed. Name agreement and representational similarity did not interact.

First chunk length

As shown in Figure 2 (right) and Table 1, first chunk length was affected by name agreement and representational similarity. As supported by an ordinal mixed model (see Table 2), participants planned, on average, fewer names in their first response chunk for low name agreement pictures than high name agreement pictures (β = 0.35, SE = 0.07, $p < 0.001$), as they made fewer responses with maximal first chunks (i.e. chunk length = 6) in the low name agreement than in the high

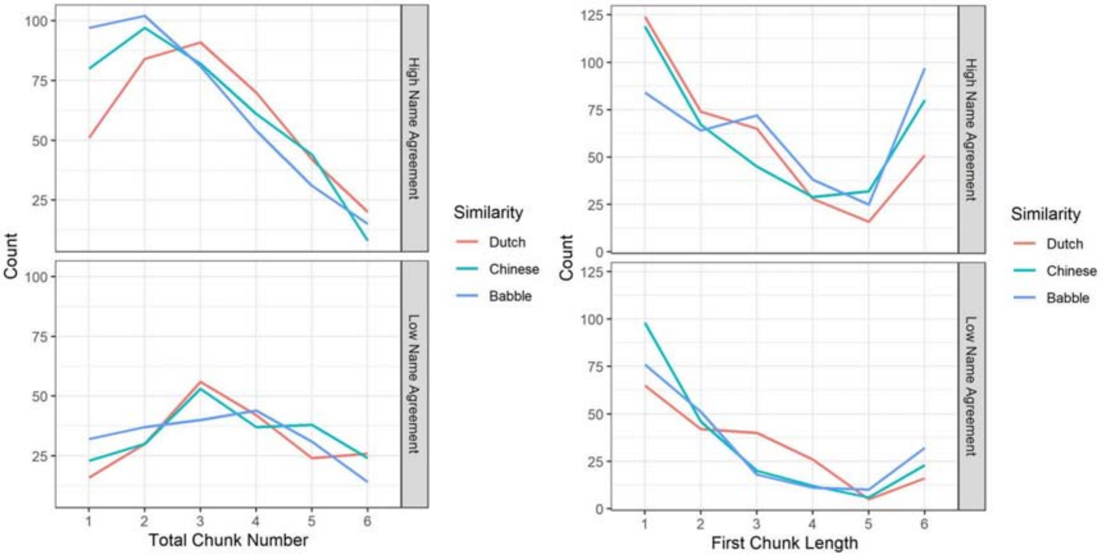

Figure 2. Total chunk number (left) and first chunk length (right) in Experiment 1 split by representational similarity (Dutch speech, Chinese speech, eight-talker babble) and name agreement (high, low). All measures reflect fully correct naming trials only.

name agreement conditions (see Figure 2 (right)). The first chunk length for pictures in the two language conditions (Dutch and Chinese) was shorter, on average, than in the eight-talker babble condition (β = −0.32, SE = 0.08, $p < 0.001$). Collapsed across name agreement, participants made more responses with minimal first chunks (i.e. chunk length = 1) and fewer responses with maximal first chunks (chunk length = 6) in the language conditions than in the babble condition (see Figure 2 (right)). There was no difference in first chunk length for the Dutch and Chinese speech conditions. However, name agreement and representational similarity did interact on the second contrast (β = −0.45, SE = 0.14, $p < 0.001$), which showed that while there was no main effect of Dutch versus Chinese speech, this main effect was qualified by name agreement such that participants produced more names in their first response chunk in the Dutch speech than in the Chinese speech conditions for high name agreement pictures (β = −0.21, SE = 0.08, $p < 0.05$) but not for low name agreement pictures.

Trials with correct first responses

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

Discussion

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

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

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

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

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

Experiment 2

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

Method

Participants

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

Apparatus

The same apparatus was used as in Experiment 1.

Materials

Visual stimuli. A subset of the pictures (40 of the original 42 picture grids) from Experiment 1 was selected to yield 120 high name agreement items (100%) and 120 low name agreement items (50%–87%). Independent *t*-tests revealed that the two sets of items differed significantly in name agreement, but not in any of the 10 psycholinguistic attributes described in Experiment 1 (i.e. visual complexity, AoA, WF, number of phonemes, number of syllables, word prevalence, PNF, PNS, ONF, and ONS). These pictures were divided into two subsets for the two blocks; both subsets were matched on all above-mentioned 10 properties including name agreement.

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

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

All of the 48 word lists were recorded by a female native Dutch speaker in neutral prosody.[4] As in Experiment 1, each list was then edited to make an audio file lasting 12 s by deleting initial and final silences and compressing the trial duration by a small amount if necessary (up to 9.5%). All auditory files were also matched on intensity (80 dB) using Praat.

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

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

Design

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

Procedure

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

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

In the divided-attention condition (Block 2), trials began with a fixation cross that was presented for 500 ms, followed by a blank screen for 300 ms. Then a 2 × 3 grid appeared on the screen in which six pictures were presented while a 12 s long sound file played. Participants were again asked to name the pictures one by one in order (first row, followed by second row) while listening to the Dutch speech. Next a blank screen was presented for 700 ms, followed by an auditory word. Participants needed to decide whether this word appeared in the Dutch speech stream they just heard by pressing the left or right button on a button box; assignment of the buttons to yes/no responses was counterbalanced across participants. Then a blank screen was presented for 1500 ms before the start of the next trial.

Analysis

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

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

Results

Naming accuracy

Participants produced the intended names of all six pictures on 63% of naming trials. As shown in Table 3, naming accuracy was affected by both name agreement and attention demand. As supported by a logistic mixed-effect model (see Table 4), accuracy for high name

Table 3. Dependent measures in Experiment 2 by name agreement and attention demand. For accuracy, range follows in parentheses, for other measures, standard deviation follows in parentheses.

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

Note. All timing and chunking measures reflect fully correct naming trials only.

agreement pictures was reliably higher than for low name agreement pictures (β = 2.23, SE = 0.23, $p <$ 0.001), and accuracy in the focused-attention condition was reliably higher than in the divided-attention condition (β = 0.33, SE = 0.13, $p < 0.01$). Name agreement and attention demand also interacted (β = 0.51, SE = 0.26, $p < 0.05$), showing that accuracy for high name agreement pictures was higher in the focused-attention than in the divided-attention condition (β = 0.59, SE = 0.20, $p < 0.01$), with no such difference for low name agreement pictures.

Memory task accuracy

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

Onset latency

As shown in Figure 3 (left), log-transformed onset latency was affected by name agreement only. As supported by a linear mixed-effect model (see Table 4), it took reliably longer for participants to plan names for low name agreement pictures than high name agreement pictures (β = −0.18, SE = 0.04, $p < 0.001$). No attention demand effect was observed, and name agreement and attention demand did not interact.

Speech duration

As shown in Figure 3 (right), log-transformed speech duration was affected by name agreement and attention demand. As supported by a linear mixed-effect model (see Table 4), it took reliably longer for participants to plan names for low name agreement pictures than high name agreement pictures (β = −0.26, SE = 0.02, p < 0.001). Log-transformed speech duration in the divided-attention condition was reliably longer than in the focused-attention condition (β = −0.03, SE = 0.01, p < 0.05). Name agreement and attention demand did not interact.[5]

Total chunk number

As shown in Figure 4 (left) and Table 3, total chunk number was affected by name agreement and attention

Table 4. Mixed-effect models for log-transformed onset latencies (Log-Onset), log-transformed speech durations (Log-Duration), accuracy, and chunk measures in Experiment 2.

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

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

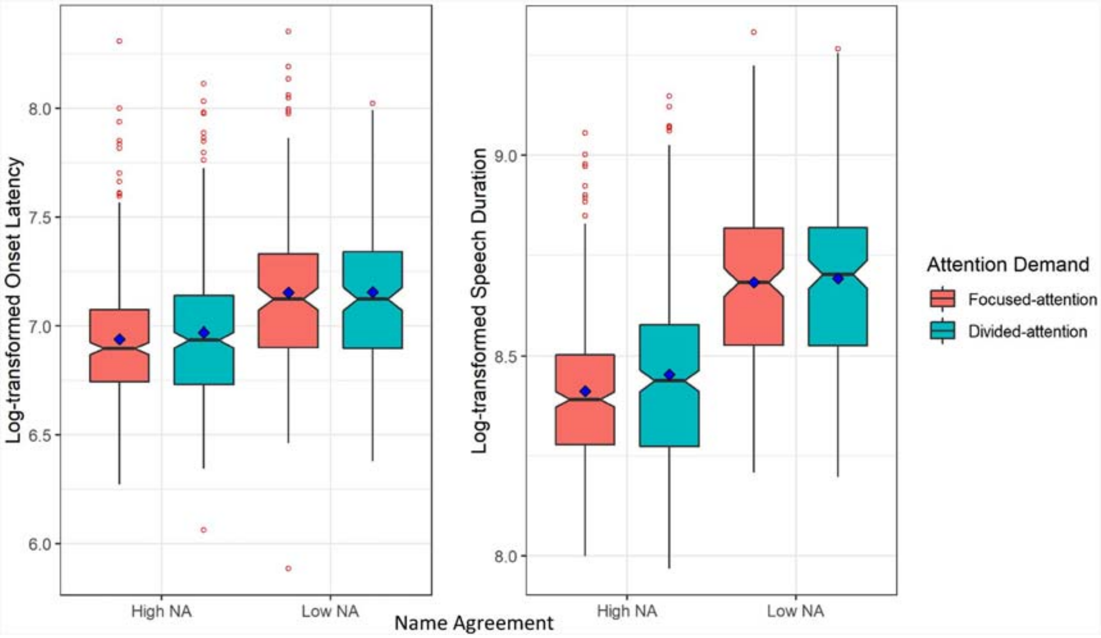

Figure 3. Log-transformed onset latencies (left) and log-transformed speech durations (right) in Experiment 2 split by attention demand (focused-attention, divided-attention) and name agreement (NA; high, low). Per condition, blue squares represent means and red points reflect outliers. All measures reflect fully correct naming trials only.

demand. As supported by an ordinal mixed model (see Table 4), participants grouped their responses in more small chunks for low name agreement pictures than high name agreement pictures (β = −1.27, SE = 0.11, $p < 0.001$). Participants also grouped their responses in more small chunks in the divided-attention than in the focused-attention conditions (β = −0.15, SE = 0.07, $p < 0.05$). Name agreement and attention demand interacted (β = −0.32, SE = 0.14, $p < 0.05$) such that participants grouped the high name agreement pictures into more small chunks in the divided-attention condition than in the focused-attention condition (β = −0.31, SE

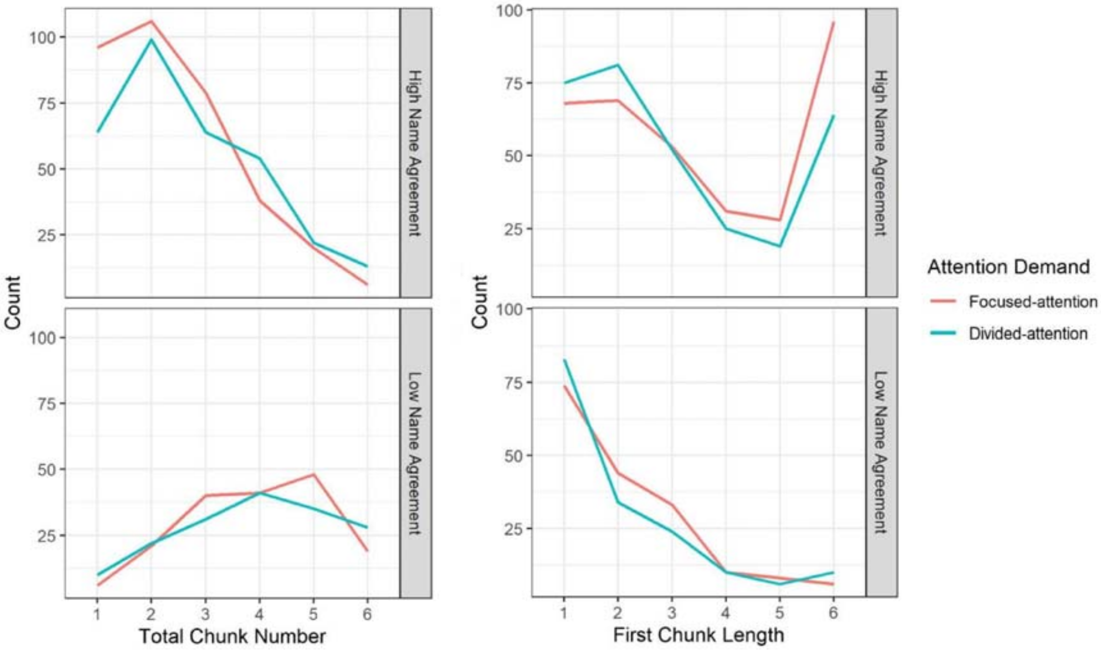

Figure 4. Total chunk number (left) and first chunk length (right) in Experiment 2 split by attention demand (focused-attention, divided-attention) and name agreement (NA; high, low). All measures reflect fully correct naming trials only.

= 0.08, $p < 0.001$), with no difference for low name agreement pictures.

First chunk length

As shown in Figure 4 (right) and Table 3, first chunk length was also affected by name agreement and attention demand. As supported by an ordinal mixed model (see Table 4), participants planned, on average, fewer names in their first response chunk for low name agreement than high name agreement pictures (β = 0.87, SE = 0.13, $p < 0.001$), as they made fewer responses with maximal first chunks (i.e. chunk length = 6) in the low name agreement than in the high name agreement conditions (see Figure 4 (right)). The first chunk length was also shorter, on average, in the divided-attention condition than in the focused-attention condition (β = 0.17, SE = 0.08, $p < 0.05$), as participants made more responses with maximal first chunks (i.e. chunk length = 6) in the focused-attention than in the divided-attention conditions (see Figure 4 (right)). Name agreement and attention demand did not interact.

Trials with correct first responses

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

Correct memory trials

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

Discussion

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

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

While name agreement and attention demand did not interact on the timing measures, we did observe an interaction between the two factors on accuracy and total chunk number. This suggests that when the attention demand for the comprehension task was high, individuals grouped high name agreement pictures into more chunks – coordinating the planning and articulation of the picture names more sequentially – and were reliably less accurate than when attention demand was low, but the effect was not found for low name agreement pictures. Similar to what we observed in Experiment 1, this pattern is opposite of what we predicted. We discuss this further in the General Discussion.

General discussion

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

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

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

Lexical selection of planning in continuous speaking and listening

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

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

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

Representational similarity in concurrent production and comprehension

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

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

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

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

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

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

Attention demand of comprehension influences concurrent production

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

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

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

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

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

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

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

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

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

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

Outlook

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

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

Conclusion

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

Notes

1. After conducting the experiment, we had a smaller effective sample size than originally anticipated due to many excluded incorrect trials. Further simulations using 21 participants and 84 items (2/3 of the original item number) suggested that 21 participants should still lead to 98% power to observe an interaction where the name agreement effect was 40 ms (100 ms sd) in the eight-talker babble and Chinese listening conditions, and 80 ms (100 ms sd) in the Dutch listening condition.
2. We also coded naming responses strictly such that only the first common names for the pictures given by MultiPic database (Duñabeitia et al., 2018) were correct. We found there were too few fully correct trials in the most difficult conditions in both experiments (51 trials in the low NA & Dutch condition for Exp1, 51 trials in the low NA & divided-attention condition for Exp2). Thus, we did not conduct the same version of analyses for these data.
3. To explore planning done between producing chunks of words, a linear mixed-effect model was also fitted on the measure of log-transformed total pause time. Total pause time was defined as the sum of all within-utterance pauses with minimal durations of 200 ms. The results for this variable patterned in the same way as speech duration. Log-transformed total pause time was affected by name agreement ($\beta = -0.67$, $SE = 0.18$, $p < 0.001$) and representational similarity (language conditions vs. language-like noise condition ($\beta = 0.75$, $SE = 0.19$, $p < 0.001$); Dutch speech vs. Chinese speech ($\beta = 0.34$, $SE = 0.16$, $p < 0.05$)), with no reliable interactions.
4. This was a different speaker than in Experiment 1.
5. As in Experiment 1, we also performed analyses on log-transformed total pause time. Log-transformed total pause time was only affected by name agreement ($\beta = -1.40$, $SE = 0.22$, $p < 0.001$), suggesting that it took longer pause when planning names for low name agreement than high name agreement pictures. Attention demand did not affect log-transformed total pause time, and it also did not interact with name agreement.

Acknowledgments

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

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the Max Planck Society. We also would like to express our gratitude to the China Scholarship Council for the support of the first author's study at the Max Planck Institute for Psycholinguistics.

ORCID

Jieying He http://orcid.org/0000-0002-2937-5100

References

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

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

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

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

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

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

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

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

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

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. In B. Bel, & I. Marlien (Eds.), *Proceedings of the speech prosody 2002 conference* (pp. 199–202). Laboratoire Parole et Langage.

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

Cook, A. E., & Meyer, A. S. (2008). Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 886. https://doi.org/10.1037/0278-7393.34.4.886

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

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

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

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

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

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

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

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

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

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

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1187–1199. https://doi.org/10.1037/0278-7393.28.6.1187

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

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

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

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

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

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

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

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

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

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

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain*

Sciences, *22*(1), 1–38. https://doi.org/10.1017/S0140525X99001776

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

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

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

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

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

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168–173. https://doi.org/10.1016/j.cognition.2008.08.002

Moisala, M., Salmela, V., Salo, E., Carlson, S., Vuontela, V., Salonen, O., & Alho, K. (2015). Brain activity during divided and selective attention to auditory and visual sentence comprehension tasks. *Frontiers in Human Neuroscience*, *9*(86), https://doi.org/10.3389/fnhum.2015.00086

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

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

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

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

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

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

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

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

Rivenez, M., Guillaume, A., Bourgeon, L., & Darwin, C. J. (2008). Effect of voice characteristics on the attended and unattended processing of two concurrent messages. *European Journal of Cognitive Psychology*, *20*(6), 967–993. https://doi.org/10.1080/09541440701686201

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

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

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

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

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

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

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

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

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

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

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

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

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

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science*, *12*(6), 462–466. https://doi.org/10.1111/1467-9280.00386

Tombu, M., & Jolicœur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 3–18. https://doi.org/10.1037/0096-1523.29.1.3

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

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

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

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

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