Representation of others' synchronous and asynchronous sentences interferes with sentence production.

---In press in Quarterly Journal of Experimental Psychology---

Chiara Gambi

University of Edinburgh and Cardiff University

Joris Van de Cavey

Ghent University and Karus vzw, Belgium

Martin J. Pickering

University of Edinburgh

Address for correspondence:

Chiara Gambi

School of Psychology

70, Park Place

Cardiff University

CF10 3AT Cardiff, U.K.

Email: GambiC@cardiff.ac.uk

Abstract

In dialogue, people represent each other's utterances in order to take turns and communicate successfully. In previous work [Gambi, C., Van de Cavey, J., & Pickering, M. J. (2015). Interference in joint picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41(1),* 1-21.], speakers who were naming single pictures or picture pairs represented whether another speaker was engaged in the same task (versus a different or no task) concurrently, but did not represent in detail the content of the other speaker's utterance. Here, we investigate co-representation of whole sentences. In three experiments, pairs of speakers imagined each other producing active or passive descriptions of transitive events. Speakers took longer to begin speaking when they believed their partner was also preparing to speak, compared to when they did not. Interference occurred when speakers believed their partners were preparing to speak at the same time as them (synchronous production and co-representation; Experiment 1), and also when speakers believed that their partner would speak only after them (asynchronous production and co-representation; Experiments 2a and 2b). However, interference was generally no greater when speakers believed their partner was preparing a different compared to a similar utterance, providing no consistent evidence that speakers represented what their partners were preparing to say. Taken together, these findings indicate that speakers can represent another's intention to speak even as they are themselves preparing to speak, but that such representation tends to lack detail.

Keywords: sentence production; picture description; imagination; joint task; co-representation

Representation of others' synchronous and asynchronous sentences interferes with sentence production

Introduction

The norm in conversation is for speakers to take turns (Sacks, Schegloff, & Jefferson, 1974) with remarkably precise timing (Stivers et al., 2009). This turn-taking system has deep evolutionary roots (Levinson, 2016), emerges early in childhood (Hilbrink, Gattis, & Levinson, 2015), and capitalizes on neural mechanisms of prediction (e.g., Magyari, Bastiaansen, De Ruiter & Levinson, 2014) to ensure smooth coordination between speakers (Levinson, 2016). But speakers can also produce language simultaneously, for example in special ritualised forms of choric speech (e.g., in singing, chanting, or praying; Cummins, 2009; Jasmin et al., 2016).

Whether they speak simultaneously or consecutively, speakers coordinate with one another: The timing and content of their utterances is dependent on the timing and content of their partner's utterances. For example, speakers typically slow down during choric speech (Cummins, 2003, 2009), perhaps to facilitate synchronization. During conversation, they predict what their partner will say so they can begin preparing an appropriate response ahead of time (Levinson, 2016). These observations illustrate how simultaneous and consecutive speaking involve different coordination demands, which are likely to be served by different coordination mechanisms. But importantly, they also show that, both when speaking simultaneously and when speaking consecutively, speakers represent their partner's use of language. In this paper, we ask how speakers represent others' utterances: Do they represent *whether* others are preparing to speak, *what* they will say, and *when* they will say it?

Co-representation of actions and utterances

Co-representation accounts of joint action propose that co-acting individuals build representations of their partner's tasks and actions which are in the same format as representations

of their own task and actions (Knoblich, Butterfill, & Sebanz, 2011; Sebanz, Knoblich, & Prinz, 2003, 2005). Such co-representation can support the coordination of actions in time and space between individuals because it means others' actions can directly affect their own: The shared representational format allows "cross-talk" between representations of their own and others' actions. To illustrate, participants in one study (Schmitz, Vesper, Sebanz, & Knoblich, 2017) had to synchronize the landing times of their reaching movements; when one of the participants encountered an obstacle that delayed her landing time, the other participant also tended to move as if there were an obstacle in his path, suggesting he represented his partner's action. Similarly, when participants were tasked with completing a sequence of two actions, they were slower if they knew their partner was simultaneously performing the same two actions in a different order; remarkably, this was the case even if partners were not asked to synchronize their landing times, suggesting that co-actors may co-represent each others' actions even without an explicit task goal to do so (Schmitz, Vesper, Sebanz, & Knoblich, 2018).

Importantly, there is evidence of cross-talk between self- and other-representations for utterances as well as for actions. Baus et al. (2014) had participants take turns naming pictures with a confederate, while EEG was recorded. There was a larger P200 component when the participant prepared to produce a low-frequency compared to a high-frequency word, which suggests this component indexes difficulty of lexical retrieval. Crucially, there was a larger P300 component when the participant stayed silent and the confederate prepared to produce a low-frequency compared to a high-frequency word, but not when the confederate stayed silent as well. Despite the delayed latency, the authors interpreted this other-related frequency effect on the P300 component as suggesting that participants were sensitive to the frequency manipulation even when their partner was preparing to name the picture (but they themselves were not), and that this was because they partly retrieved words their partner was about to produce.

Similarly, in a joint Stroop task in which one participant responded to color words by saying "red" when they were printed in red, and the other responded to them by saying "green" when they were printed in green, there was a larger P300 component when participants stayed silent and their partner prepared to respond (joint experiment), compared to when they both stayed silent (solo experiment; Demiral, Gambi, Nieuwland, & Pickering, 2016). While it is possible the enhanced P300 is due to greater attention to stimuli that the partner responded to, it may also be that speakers mapped their partner's stimuli to their partner's response even though they did not have to respond themselves. Accordingly, the study also found indication of reduced perceptual conflict (i.e., conflict between the ink color and the meaning of the color words, when they mismatched; e.g., the word *green* printed in red ink) in the joint compared to solo experiment, which suggests that speakers were less likely to treat the other color as conflicting when it was their partner's response than when it was not.

Finally, Kuhlen and Adbel Rahman (2017) and Hoedemaker, Enrst, Meyer, and Belke (2017) showed that listening to another speaker name pictures delays naming times for pictures from the same semantic category, similarly to naming those pictures oneself. In other words, there appears to be a between-speaker cumulative semantic inhibition effect (Brown, 1981). The within-speaker version of this effect is thought to result from changes in the strength of connections between conceptual and lexical representations, or between features and concepts, which are the result of previous retrieval episodes (Belke, 2013; Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Oppenheim, Dell, & Schwartz, 2010). Thus, the between-speaker cumulative semantic inhibition effect may suggest that other-produced utterances have a similar effect on production representations as self-produced utterances. Evidence that a between-speaker effect occurs even when speakers cannot hear their partner, but are merely told that they partner is naming the pictures (Experiments 2 and 3; Kuhlen & Abdel Rahman, 2017), supports this interpretation. But it should be noted that in a more recent attempt to replicate this finding (Kuhlen & Abdel Rahman, 2021), the

authors found no evidence for between-speaker cumulative semantic inhibition when the partner was not present in the same room as the participant.

In sum, a number of findings suggest that speakers represent another speaker's utterances before comprehending them and, sometimes at least, even when they believe they will not be able to hear them (Gambi & Pickering, 2016). These findings suggest that the representations speakers form of others' utterances are similar to the representations they form of their own utterances when they are preparing to speak. This evidence is therefore consistent with co-representation of utterances (Gambi & Pickering, 2011, 2013).

Representing another's utterances while simultaneously preparing to speak

In these studies, speakers took turns with their (present or imagined) co-speakers, which means they never had to represent their partner *while* preparing to speak themselves. In contrast, in previous work we manipulated speakers' beliefs about the task their partner was performing in a different room (so they could never hear their partner) while they prepared to name pictures themselves (Gambi, Van de Cavey, & Pickering, 2015). Across four experiments, speakers were told their partner would produce an utterance which was the same or different from their own utterance, stay silent, or respond *yes* or *no* to a semantic categorization question. Naming latencies were faster when participants believed their partner was not speaking, or was speaking but not engaging in lexical retrieval (categorization condition), than when they believed their partner was retrieving lexical items. This finding that merely imagining another speaker producing an utterance affects concurrent production is particularly striking, because these speakers represented that their partner's production system was simultaneously activated even though the task did not require them to pay attention to their partner's task.

Importantly, Brehm, Taschenberger, and Meyer (2019) found a similar result to Gambi et al. (2015) using a simpler design. In two experiments, they had participants name single pictures that

were presented either on the right- or the left-hand side of the screen, while their partner (who was sitting next to them) either named the picture presented on the other side of the screen or categorised it (participants wore noise-cancelling headphones so they could not hear their partner); the two pictures could either be the same or different, so that participants were either producing/categorizing the same utterance as their partner or a different one. Experiment 1 also included an individual condition in which the participant performed the same task alone (they were told their partner had failed to show up for the experiment). Crucially, participants took longer to name the pictures when they knew their partner was present and was performing a different task (categorization) than the same task (naming). Note that this study reverses the pattern found by Gambi et al. (2015), who instead showed *shorter* latencies when participants believed their partner was categorizing pictures rather than naming them (see Brehm et al., 2019 for a discussion of methodological differences that may have led to this reversal of the effect). Crucially, however, the study indicates that speakers represented their partner's task and that these representations affected their own production.

However, in both Gambi et al. (2015) and Brehm et al. (2019) there was no consistent evidence that speakers represented the content of their partner's utterances. In Gambi et al. (2015), naming latencies were no slower when speakers believed their partner was preparing to produce a different utterance compared to the same utterance. While there was a tendency in some of the experiments for naming errors to be more common when participants believed their partner was producing a different utterance, this tendency was not consistent across experiments. Similarly, in Brehm et al.'s (2019) Experiment 1, latencies were longer when the participant knew the partner was naming (but not categorising) a different picture, but this was the case even when no actual partner was present, and moreover the effect was not replicated in Experiment 2. Interestingly, participants tended to look more often at their partner's picture when they knew they were naming rather than categorizing, and also when the partner's picture was different to their own, but overall

there were very few looks to partner's pictures, suggesting that speakers did not regularly engage in detailed representation of their partner's utterances.

In sum, the findings of Gambi et al. (2015) and Brehm et al. (2019) are consistent with other work (Baus et al., 2014; Demiral et al., 2016; Kuhlen & Abdel Rahman, 2017; Hoedemaker et al., 2017) in showing that speakers represent their partners' act of production in a similar way to how they represent their own act of production. However, they also suggest that the content of others' utterances (i.e., what they are saying) may not be represented: there was only inconsistent evidence of increased interference when producing a different utterance from one's partner, compared to producing the same utterance. Specifically, this finding contrasts with Baus et al. (2014) and Kuhlen and Adbel Rahman (2017), who showed that people are affected by linguistic properties of words that their partners have produced or are about to produce, suggesting that people can represent detailed aspects of the content of others' utterances (frequency, semantic category). It also contrasts with evidence for co-representation of other's actions, such as the finding that people execute movement sequences less efficiently when they know their partner is executing the same sequence in reverse order versus the same order (Schmitz et al., 2018).

Does simultaneous production affect the degree of representation of another's utterances?

One explanation of these divergent findings is that language production is a demanding cognitive process (Roelofs & Piai, 2011). Thus, speakers may have had less resources to allocate to representing their partner's utterance when they were simultaneously preparing to speak themselves, resulting in less detailed co-representation – that is, limited to co-representation of *whether* another is speaking but not of *what* they are saying – in Gambi et al. (2015) and Brehm et al. (2019) compared to the other studies. Similarly, Hoedemaker and Meyer (2018) also found only limited evidence that speakers represented their partner's utterance when they had to speak themselves. They had speakers name one picture, then listen to their partner name a second picture, and finally name a third picture themselves. Speakers looked at the second picture more often and

earlier when their partner named it than when nobody named it but, importantly, looks to the second picture were much longer and earlier when participants named it themselves, indicating that co-representation stopped well short of planning the partner's utterances. Although in Hoedemaker and Meyer (2018) participants took turns speaking – similarly to studies that found evidence for detailed co-representation of utterance content (Baus et al., 2014; Kuhlen & Abdel Rahman, 2017) – picture triplets were presented on the screen simultaneously, which likely encouraged speakers to plan the third picture name while their partner was naming the second picture. Simultaneous planning may have taken resources away from speakers' representation of their partner's utterance, making them less likely to represent its content in detail. In fact, in other conditions speakers may have not represented their partner's act of production at all: When speakers were asked to name the first of two pictures, they were as fast to begin naming the first picture whether they knew their partner later named the second picture or they knew their partner would remain silent (whereas they were much slower when they themselves had to name the second picture as well).

In sum, speakers sometimes form detailed representations of the content of their partner's utterances, but whether they do so may depend on available resources. For example, if either representing another's utterance or producing their own utterance is particularly demanding, speakers may not represent their partner's utterances in any detail (or indeed, they may not represent their partner at all). Conversely, if either co-representation or production (or both) are made easier, speakers may be able to form more detailed representation of their partner's utterances even if they are speaking simultaneously. For example, Brehm et al. (2019) showed that co-representation can have a facilitatory effect on production when the task of representing another speaker is made easier: In their study, beliefs about the partner's task were manipulated across blocks and they found shorter latencies when participants believed their partner was naming pictures (like them) than when they believed their partner was categorizing pictures. In contrast, Gambi et al. (2015) manipulated beliefs on a trial-by-trial basis (and used a more difficult picture naming task) and found longer latencies when speakers believed their partner was naming (like

them) compared to categorizing. This may suggest that, as task difficulty increases, co-representation turns from a potentially facilitatory factor to a source of interference with concurrent language production, making speakers less likely to engage with it. However, even with a much simpler task, Brehm et al. still found little evidence of detailed representation of the content of the partner's utterances, so it remains unclear whether limited resources are the only reason why some studies show little or no evidence for detailed co-representation.

The current study

In the current study, instead of simplifying the task, we opted to make it harder. Naming pictures is cognitively demanding, but planning whole sentences is even more so (Ferreira, 1991). Thus, speakers may not be able to represent others' utterances at all when they are asked to concurrently plan a whole sentence. Previous work has used comparatively simple production tasks, where speakers produce either single words, or pairs of words (Baus et al., 2014; Brehm et al., 2019; Demiral et al., 2016; Gambi et al., 2015; Hoedemaker et al., 2017; Hoedemaker & Meyer, 2018; Kuhlen & Abdel Rahman, 2017). In contrast, we asked speakers to describe pictured events using full transitive sentences (e.g., *The nun follows the doctor*). Producing sentences involves additional planning stages (Bock & Levelt, 1994) that are not involved in producing single words (or even ordered pairs of words). Specifically, thematic roles (agent and patient) must be assigned to syntactic positions (subject and object) so that an appropriate syntactic structure can be chosen (Ferreira, 1994; Segaert, Wheeldon, & Hagoort, 2016). Importantly, transitive events can be described using an active structure (as in the example above), or a passive structure (as in *The doctor is followed by the nun*). To ensure that speakers focused on such additional planning stages, they produced active and passive sentences, and sentence voice was varied on on a trial-by-trial basis.

It is possible that the complex task of preparing to produce a whole utterance may entirely prevent speakers from representing their partner's act of production. If so, production latencies

should be unaffected by whether speakers believe their partner is also preparing to produce a sentence, or simply remaining silent. Such a finding would in turn suggest a limited role for other-representations in utterance coordination, because such representations would not be available to interlocutors who are either speaking (e.g., in choric speech) or preparing to speak (e.g., when they are about to take a turn in dialogue).

If, however, speakers represent their partner's act of production even while perfoming a demanding production task, it would suggest that they can routinely represent another speaker's act of production while concurrently preparing to speak. If so, we expected slower production latencies when speakers believed their partner was engaging in language production, compared to when they believed their partner was silent, similarly to our previous research using words and word pairs (Gambi et al., 2015). We also compared production latencies when speakers believed their partner was producing a sentence in a different voice, to when speakers believed their partner was producing an utterance in the same voice. Even if increasing task difficulty does not make representation of others' utterances less likely, previous work (Gambi et al., 2015; Brehem et al., 2019) suggests such representations are unlikely to be very detailed. If so, production latencies should not be affected by whether speakers believe their partner is producing an utterance in the same or a different voice, even if they are affected by whether their partner is preparing to produce an utterance at all. This is because the voice of the partner's utterance would matter only if speakers formed a representation of the partner's utterance that was detailed enough to distinguish between a passive and an active utterance.

In addition, planning a passive sentence may be more difficult than planning an active sentence – for example, Ferreira (1994) reported that passive sentences can take up to 1000-1400ms longer to formulate than corresponding active sentences in certain conditions. If this is the case, it is also possible that the degree to which participants represent their partner's utterances will vary depending on whether they themselves are planning a passive (i.e.. more difficult) or active (less

difficult) utterance. This would be reflected in an interaction between our manipulation of partner's task and sentence voice. However, other studies have found that passive sentences are planned as quickly as active sentences, for example in the no-syntactic repetition condition of Experiment 2 of Segaert et al. (2016) and the no-word repetition condition of Segaert, Menenti, Weber, and Hagoort (2011). So it is also possible that there will be no voice-related differences in planning latencies in our study.

Finally, we have assumed that when speakers believe their partner will speak at the same time as them, they experience interference because production and representation of others' utterances are cognitively demanding processes that compete for resources *and* take place synchronously. But what if these processes are asynchronous, such as when speakers believe their partner will speak after them? It is possible that no interference would be observed in this situation, as speakers may delay representing their partner's utterance until after they have prepared their own utterance. In accord with this possibility, Hoedemaker and Meyer (2018) showed that participants did not begin speaking later when they knew their partner would speak after them than when they knew nobody would do so.

But given the evidence that speakers' representations of others' utterances often lack detail, it is also possible that speakers represent only that their partner is intending to perform an act of production, but not the timing of this act. For example, speakers may represent their partner's intention to speak without representing *when* their partner will actually plan an utterance. This would be compatible with evidence that generating an intention to speak and planning an utterance are separate components of the production process (Levelt, 1989). For example, there is a dissociation between processes determining what to say and processes determining when to speak during turn-taking (Corps, Crossley, Gambi, & Pickering, 2018). Further, lexical access is faster when there is a conscious intention to speak compared to when this is absent (Strijkers, Holcomb, & Costa, 2011). In sum, if this account is correct, it will make no difference whether speakers believe

their partner intends to speak at the same time as them or after them – in both cases, they should experience interference compared to conditions in which they believe the partner does not intend to speak.

In conclusion, previous work has shown that speakers can represent others' utterances and that doing so affects their own production, but there is mixed evidence regarding the ease with which these representations are formed and the degree of detail involved. Relevant factors may be the ease of speakers' own production and the extent to which preparation of speakers' own utterances and representation of others' utterances overlap in time. In this study we explored how these factors may affect the degree of interference from representations of others' utterances on production.

In all experiments, participants produced full sentences to increase the difficulty of the production task compared to previous work. In Experiment 1 we had speakers believe they were speaking simultaneously (as in Gambi et al., 2015), whereas in Experiment 2 (2a and 2b) we had speakers believe their partner would speak after them. Importantly, participants sat in separate soundproof booths and were not able to hear one another at any point during either experiment. Thus, in Experiment 2a we led participants to believe their partner would speak after them, even though they in fact spoke simultaneously, which allowed us to collect comparable data from both speakers in a pair. In Experiment 2b, instead, we had one randomly-selected participant actually speak after their partner; although this meant we could use data from only half of the participants (those who spoke first), it provided a replication of Experiment 2a that did not rely on creating a false belief in the participants.

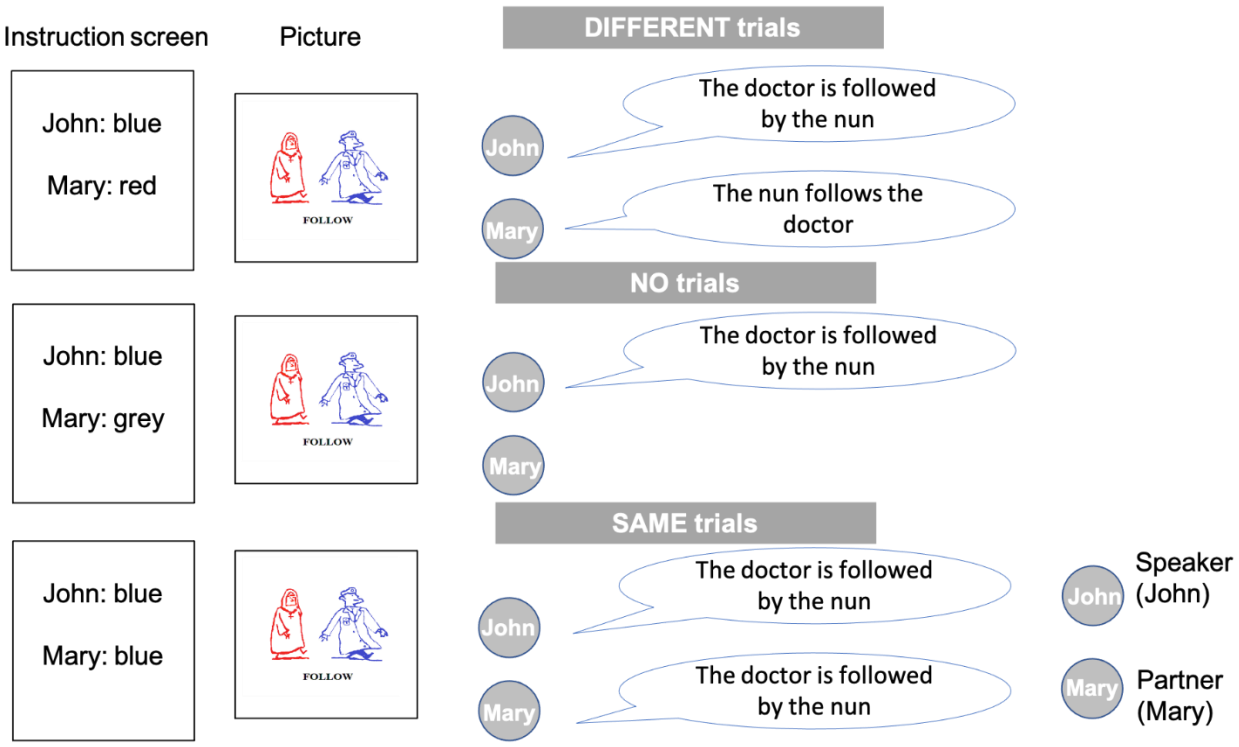
Experiment 1: Synchronous production and representation of another's utterance

In Experiment 1, either pairs of participants described pictures simultaneously or one speaker described a picture while the other remained silent (as in Gambi et al., 2015, Experiment 1).

Thus, production and representation (of other utterances) were synchronous. Speakers produced short descriptions of transitive events (e.g., of a nun following a doctor, in Figure 1) in either the active (e.g., *The nun follows the doctor*) or the passive voice (e.g., *The doctor is followed by the nun*). We controlled these descriptions using a variant of the so-called "traffic-light" paradigm (e.g., Menenti, Gierhan, Segaert, & Hagoort, 2011), in which participants were assigned a color on each trial, and were instructed to first describe the character who appeared in that color. For example, in Figure 1, top row, John has been assigned the color blue; since the picture displays the patient of the action in blue, John will begin his description with the patient, and will thus produce a passive sentence. In contrast, Mary has been assigned the color red, so she will start her description with the agent and produce an active sentence.

Crucially, on every trial, the instruction screen also alerted participants to what their partner was to do once the picture appeared, so we could manipulate what speakers believed their partner was doing (Partner's Task). There were three types of trials (Figure 1): On *Different* trials (top row), participants were assigned different colors (John: blue, Mary: red or John: red, Mary: blue), so that they described the picture starting with different characters, and produced sentences of different voice; on *No* trials (middle row), the speaker described the picture starting with either character while their partner remained silent (as indicated by the color grey; John: blue – Mary: grey, John: red – Mary: grey, or John: grey, Mary: blue, John: grey, Mary: red); finally, on *Same* trials (bottom row), participants were assigned the same color word (e.g., John: red – Mary: red or John: blue – Mary: blue) and thus described the picture starting with the same character, producing a sentence in the same voice as each other (both actives or both passives).

Figure 1. Schematic representation of the design used in Experiment 1, 2a, and 2b. Participants saw their actual names on the instruction screen (*John* and *Mary* are used here for illustrative purposes). In this example, John produces a passive description because he has been assigned the color blue and the blue character is the patient in the picture. John is the "speaker" and Mary is the "partner" because John is the one speaking on No trials in this example, but on other trials their roles were reversed (i.e., we collected data from both participants in all conditions).



If speakers who are preparing to produce a whole sentence do not concurrently represent their partner's act of production, speech onset latencies should be unaffected by our Partner's Task manipulation: Latencies should be similar across all three types of trials (*Different*, *No*, and *Same*). But if speakers represent their partner's act of production, we expect to find a joint interference effect (similar to Gambi et al., 2015): Speakers should take longer to begin their utterances when they believe their partner is also speaking (*Different* and *Same*), compared to when they believe their partner is silent (*No*).

Further, if speakers merely represent that their partner is producing, then latencies should be unaffected by whether speakers believe their partner is producing the same utterance as themselves, or a different utterance. But if speakers represent the content of their partner's act of production, then we would also expect a joint interference effect that depended on whether the speakers' utterance is the same as their partner's (presumed) utterance or not. Presumably, there would be greater interference when self- and other-representations conflicted, as speakers would find the incompatible representation of their partner's utterance interfering with their own act of production. Specifically, speakers should have more difficulty producing an active if they believed their partner was producing a passive than an active, and more difficulty producing a passive if they believed their partner was producing an active than a passive (i.e., latencies should be longer on Different than Same trials, irrespective of Voice).

Finally, we know that sentential planning is incremental; specifically, when speakers are preparing to produce simple sentences containing two nouns (e.g., *the arrow is next to the bag*), they appear to retrieve both lemmas before beginning to speak, but also that they retrieve the phonological representation for the second noun only *after* speech onset (Meyer, 1996; Smith & Wheeldon, 2004). If so, it may be that evidence for interference due to representing the other act of production can be measured *after* as well as (or instead of) before speech onset. Therefore, we examined not only speech onset latencies, but also utterance durations.

Method

Participants. Gambi et al. (2015) found reliable effects with sample sizes varying between 24 and 32 participants, but since the current experiment included fewer trials per participant (192 vs. 300), we increased sample size and recruited 40 native English-speaking participants from the student population of the University of Edinburgh, who were assigned to 20 pairs. Participants were paired with each other on the basis of availability and no other criteria were followed when pairing participants, except that members of a pair did not know each other before taking part in the

experiment. All procedures were approved by the Ethics Committee of the Department of Psychology, University of Edinburgh and participants gave informed written consent to participating. They were paid £6 or given course credit for taking part. Data from one participant were discarded prior to analysis because their average speech onset latency (1,681 ms) fell more than 2 standard deviations above the average onset latency across all participants (1,164 ms); data for their partner were retained since they could not have been aware of the other particiapant's unusually delayed responding.

Materials and Design. We created 32 pictures depicting simple scenes comprising two human characters and a printed verb. The pictures were created by combining 8 different transitive verbs (*follow, chase, hit, punch, shoot, kill, tickle, touch*), each repeated 4 times, and 14 characters (*nun, doctor, boxer, cowboy, robber, pirate, swimmer, waitress, ballerina, painter, monk, sailor, soldier, clown*), each repeated a variable number of times. Across repetitions of the same verb, we counterbalanced both the position of the agent (left or right) and the color of the agent (blue or red); the patient was always depicted in the other color.

The design included 6 conditions obtained by crossing Partner's Task (Same, Different, No) with Voice (Active vs. Passive); we manipulated both Partner's Task and Voice by assigning participants to color words on each trial using an instruction screen (see above and Figure 1). Note that in order to collect data from each participant in each condition, there were two additional "dummy" sets of trials ("dummy" because they provided no data for the participant in question), in which the speaker (John in Figure 1) remained silent but their partner (Mary in Figure 1) described the picture instead, in either the Passive or Active voice. Participants encountered each of the 32 pictures once in each of these 8 sets of trials (i.e., 6 conditions + 2 "dummy" sets of trials).

Procedure. Participants were tested in adjacent soundproof rooms. The experimenter controlled stimulus presentation using E-Prime (Version 2.0) from a separate room. Stimulus presentation was simultaneous on each participant's screen. Participants could not hear each other during the

experiment. There was a window between the participants' rooms (so that they were aware of their partner) but this window was perpendicular to the participants' line of vision when they faced their monitors.

Participants were first introduced to one another and the experimenter entered their first names in the program so they could be displayed on the instruction screens. They were then taken into their booths and practiced naming the characters they would later encounter during the experiment. They did this at the same time and without hearing one another. When a participant made a mistake, the experimenter provided corrective feedback to that participant (but both participants could hear the feedback). Participants were instructed how to describe pictures of transitive events (through examples) based on the color cue that was assigned to them. These instructions were delivered to both participants at the same time in the control room. The experimenter emphasized the "joint" nature of the task, noting that participants would now work together, and drawing attention to how they would be doing the same or different tasks as each other, depending on the instruction screen. The participants then went back to their respective booths to practice, before beginning the first of four experimental blocks (practice and experimental trials were similar, but practice trials used a different set of pictures). In total, participants completed 256 experimental trials (i.e., 32 pictures, each repeated 8 times); within each block there were two uses of each picture.

On each trial, a fixation cross was displayed for 1,000 ms, followed by the instruction screen for 2,000 ms. After a 500-ms blank screen, the picture appeared for 1,500 ms. Participants were instructed to produce their description as quickly and accurately as possible. Each trial ended with a 1,000-ms blank screen before the next trial began. Between blocks, participants were left free to rest until both were ready to continue, and the experiment resumed when both indicated they were ready to continue. The experiment lasted about 1 hour.

Recording and data analysis. Both participants' responses were recorded and coded for accuracy offline. A participant's description was marked as incorrect if the participant was disfluent, used a word for the character names other then the trained one, used a verb other than the one printed on the picture, or produced an active instead of a passive, or *vice versa*. Accuracy was not affected by our manipulations in Experiment 1, so we do not discuss it further (see Online Supplementary material, Tables S1.1-4, for statistical analyses). For the purpose of the analyses reported below, all incorrect responses were discarded (see Results for percentages). Recordings were then pre-processed to reduce background noise, with speech onsets being automatically tagged (using the Silence finder algorithm in Audacity, Version 1.2.5) and then manually checked for non-speech noises (e.g., lip smacks). In addition, speech offsets were manually coded, and utterance durations were defined as speech offset minus speech onset (i.e., including mid-utterance silences).

In our main analysis, we fitted linear mixed-effects models (Gaussian link function) separately to description latencies and durations. The fixed effect model structure was always ~ 1 + Voice + Partner Task + Voice: Partner Task. Voice compared passive to active utterances (to assess whether speakers took longer to produce passive utterances). For Partner Task, we defined two contrasts: The *Speaking* contrast compared the average of the Same and Different conditions to the No condition, to assess whether speakers slowed down when they believed their partner was speaking at the same time as them; the *Form* contrast compared the Same to the Different condition, to assess whether speakers were faster when they believed the form of their partner's utterance to be the same as the form of their own utterance (i.e., in the same voice).

We initially fit the full random structure justified by our design, for both participants and items (with correlations between random factors set to zero to aid convergence; Bates, Kliegl, Vasishth, & Baayen, 2015). However, models with full random structure were overparametrised, as indicated by many of the random factors being estimated to be zero or near-zero. We thus re-fit models with simplified random structures, after removing all slopes that were estimated to be less

than 10^{-3}. In text we report coefficients, standard deviations, *t* values, and 95% confidence intervals (from the *confint* function, method = "Wald") from these simplified models, but the models with full random structure always yielded comparable findings (see Online Supplementary Material, Tables S1.5-6, S2.5-6, S3.5-6 for latencies and Tables S1.14-15, S2.14-15, S3.14-15 for durations). We consider |t| values >= 2 to indicate significant findings.

As in Gambi et al. (2015), the models were fitted to the data after excluding extreme outliers (< 300 ms or > 3000 ms for description latencies; > 3500 ms for utterance durations). To assess the robustness of the findings, we performed additional analyses after data that were more than 3SD over or below each participant's mean were either discarded (trimmed dataset; see Online Supplementary Material, Tables S1.12-13, S2.12-13, S3.12-13 for latencies and S1.21-22, S2.21-22, S3.21-22 for durations) or replaced with the cut-off value (winsorised dataset; see Online Supplementary Material, Tables S1.10-11, S2.10-11, S3.10-11 for latencies and S1.19-20, S2.19-20, S3.19-20 for durations). Such additional analyses used simplified random structures to avoid overparametrization as described above, and typically yielded the same pattern as the main analyses reported below, which excluded only extreme outliers, but we note when they diverged. Full models outputs for all analyses are reported in the Online Supplementary Material. All data and analyses scripts are available at https://osf.io/v7t5z/

Results

After removing incorrect responses (Different: 13.26%, Same: 12.06%, No: 11.70%), we discarded 7 outliers for the description latency analyses, and 1 outlier for the utterance duration analyses.

Table 1. Description latencies (means and SDs, in ms) by Voice and Partner's Task.

	Experiment 1			Experiment 2a			Experiment 2b		
Partner Task	Diff.	No	Same	Diff.	No	Same	Diff.	No	Same
Voice									
Active	958 (245)	940 (241)	953 (240)	1025 (316)	1027 (313)	1045 (329)	955 (386)	953 (385)	970 (420)
Passive	948 (247)	936 (235)	960 (244)	1027 (324)	1021 (316)	1046 (327)	972 (381)	907 (345)	969 (375)
Total	953 (246)	938 (238)	956 (242)	1026 (320)	1024 (315)	1046 (328)	963 (384)	930 (366)	970 (398)

Description Latency. Speakers took longer to begin their description when they believed their partner was speaking at the same time as them (M = 955 ms, SD = 244 ms) than when they believed their partner was silent (M = 938 ms, SD = 238 ms; see Table 1 and Figure 2, left panel). Accordingly, the Speaking contrast was significant (B = 19.13, SE = 5.43, t = 3.52, CI = [8.48,29.77]). In contrast, description latencies did not differ when speakers believed their partners were preparing to say the same utterance as them and when they were preparing to say a different-voice utterance (Form: B = 3.61, SE = 6.89, t = 0.52, CI = [-9.89,17.12]; see Table 1).

Speakers did not take longer to initiate passive (M = 948 ms, SD = 242 ms) than active descriptions (M = 950 ms, SD = 242 ms) (Voice: B = -0.79, SE = 11.93, t = -0.07, CI = [-24.18,22.59]). There was also no evidence of an interaction between Voice and Partner (Speaking X Voice: B = 2.95, SE = 10.86, t = 0.27, CI = [-18.34,24.23]; Form X Voice: B = 14.33, SE = 13.70, t = 1.05, CI = [-12.53,41.18]; see Table 1).

Table 2. Utterance durations (means and SDs, in ms) by Voice and Partner's Task.

	Experiment 1			Experiment 2a			Experiment 2b		
Partner Task	Diff.	No	Same	Diff.	No	Same	Diff.	No	Same
Voice									
Active	1379	1355	1359	1436	1422	1423	1630	1607	1582
	(308)	(293)	(280)	(327)	(312)	(317)	(467)	(440)	(406)
Passive	1545	1529	1523	1646	1666	1648	1736	1754	1742
	(291)	(277)	(269)	(341)	(354)	(355)	(421)	(420)	(437)
Total	1461	1443	1441	1541	1544	1535	1684	1679	1662
	(311)	(298)	(287)	(350)	(355)	(355)	(447)	(437)	(429)

Utterance Duration. The analysis of utterance duration revealed a different and unexpected pattern compared to the analysis of description latencies. First, durations did not differ whether speakers believed their partner was speaking (M = 1,451 ms, SD = 299 ms) or was silent (M = 1,443 ms, SD = 298 ms; see Figure 3, left panel, and Table 2) (Speaking: B = 8.94, SE = 6.10, t = 1.46, CI = [-3.03,20.90]). In contrast, speakers produced longer utterances when they believed their partner was

producing a different utterance than when they were producing the same utterance (Form: B = -20.61, SE = 7.92, t = -2.60, CI = [-36.13,-5.08]; see Table 2).[1]

Unsurprisingly, passive utterances had longer durations than active ones (M = 1,532 ms, SD = 279 ms vs. M = 1,364 ms, SD = 294 ms; Voice: B = 167.73, SE = 12.21, t = 13.73, CI = [143.80,191.67]), but there were no interactions between Partner and Voice (Speaking X Voice: B = -7.40, SE = 12.91, t = -0.57, CI = [-32.70,17.91]; Form X Voice: B = 0.47, SE = 15.42, t = 0.03, CI = [-29.75,30.69]; see Table 2).

Discussion

Experiment 1 demonstrated a joint interference effect in the production of sentences. Thus, speakers are able to allocate resources to imagining their partner's act of production even when they are concurrently engaged in planning a whole sentence. As in previous research using simpler production tasks (Gambi et al., 2015), speakers took longer to begin their utterances when they believed their partner was preparing to produce an utterance at the same time as them, compared to when they believed their partner was remaining silent.

[1] Form was significant (t = -2.46) also when we replaced data points above and below 3SD from each by-participant average with the cut-off value. However, the effect was not significant (t = -1.12) when we instead discarded these data points (1.14%), suggesting that the effect may have been driven by a subset of very long durations. In support of this, when we ran a Bayesian ex-Gaussian distributional analysis (Bürkner, 2016), we found a small effect of Form on the mean of the Gaussian component (95% Credible Interval = [0.01, 0.03]), but a much larger effect on the rate of the exponential component (95% Credible Interval = [-0.31, -0.11]), which captures the thickness of the right tail of the distribution; only the latter effect remained significant after trimming.

The nature of the representations speakers formed of their partner's imagined utterances was also similar in this task to simpler production tasks (Gambi et al., 2015; Brehm et al., 2019): There was no evidence that speakers took longer to begin their utterances when they believed their partner was preparing to produce an utterance in a different, compared to the same, voice. However, when we analysed utterance durations, we unexpectedly found that speakers spent more time producing their utterances when the partner produced a different utterance than when they produced the same utterance. Since this finding appeared to be driven by a subset of very long durations (see Footnote 2), we address it further in Experiment 2.

Finally, we did not find any evidence that participants took longer to plan a passive than an active utterance. As mentioned in the introduction, differences in latencies between actives and passives are not consistently found in the literature (see e.g., Segaert et al., 2011, 2016). Thus, we could not test whether participants would be less likely to represent their partner's utterance when they themselves were planning a more difficult utterance. But importantly, utterance planning was in general more difficult in the current study than in previous work: Latencies in Experiment 1 were about 70 ms longer than in Gambi et al.'s (2015) Experiment 1. Also importantly, we could still test whether speakers formed detailed representations of the voice of their partners' utterances – such detailed representation should have caused interference in the Different compared to Same condition regardless of whether sentence structures differ in planning difficulty, because the interference is due to the mismatch between representations of one's own and one's partner's utterance rather than to the degree of difficulty associated with planning one's partner's utterance.

Figure 2. Speech onset latencies for picture descriptions across Different, No, and Same trials in Experiments 1, 2a, and 2b. Error bars represented 95% bootstrapped confidence intervals.

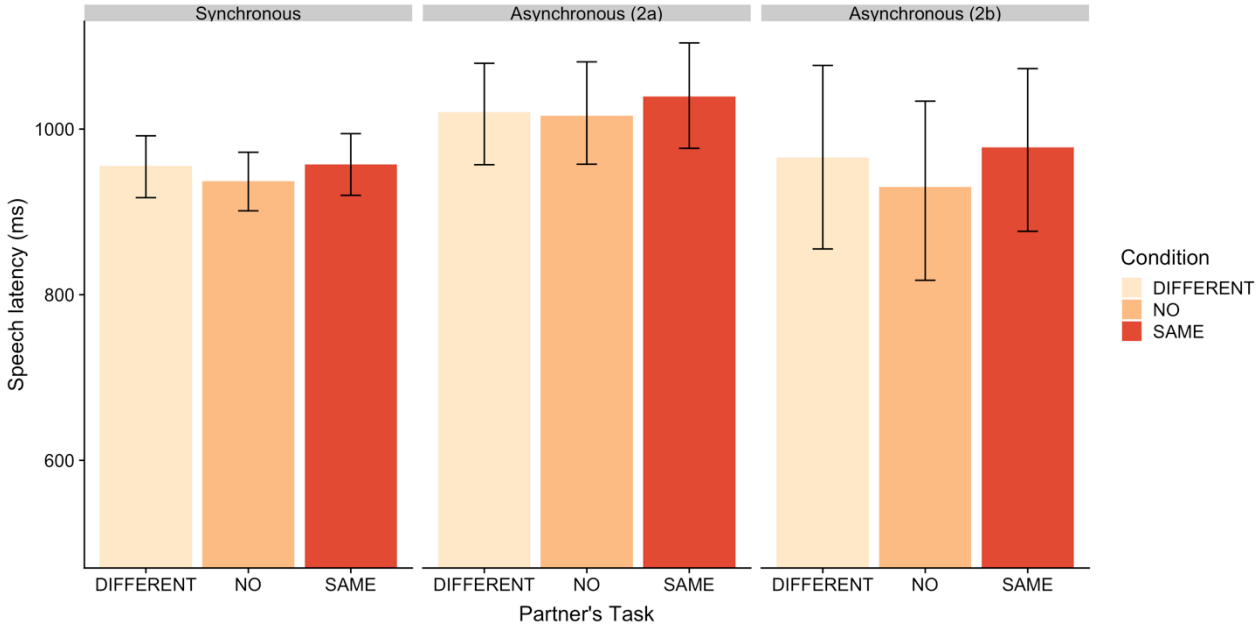
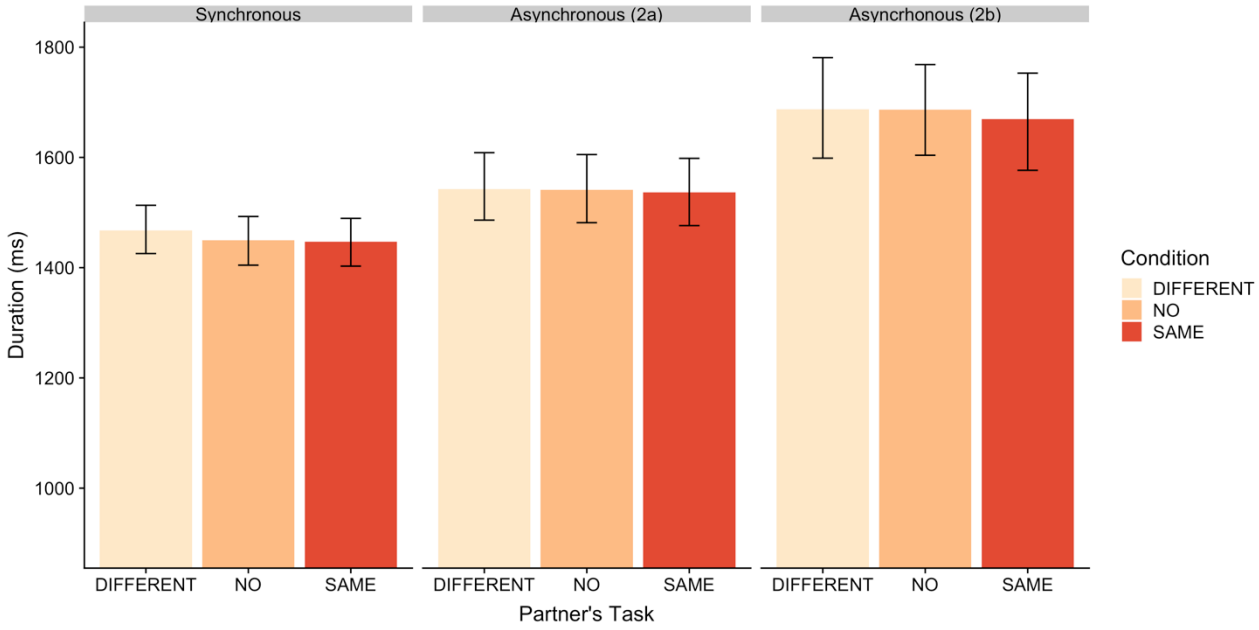


Figure 3. Utterance durations for picture descriptions across Different, No, and Same trials in Experiments 1, 2a, and 2b. Error bars represented 95% bootstrapped confidence intervals.



Experiment 2: Asynchronous production and representation of another's utterance

In Experiment 2, we asked speakers to plan a sentence and imagine their partner preparing to speak after them, so that planning and representation of their partner's act of production could be asynchronous. If speakers represent their partner's act of production at the time when they believe their partner will speak, we should find no evidence for joint interference effects. In contrast, if speakers represent their partner's intention to speak in the future in the same way as they represent their intention to speak now (or if they do not represent when their partner intends to speak), we should find longer latencies when speakers believe their partner is preparing to speak after them compared to when they believe their partner will remain silent. Since Experiment 1 showed some indication that Different trials may lead to more interference than Same trials, at least after speech onset (utterance duration analyses), we also retained the Voice manipulation, and analysed both latencies and durations.

In Experiment 2a, we had speakers believe their partner would speak after them on Same and Different trials (while remaining silent on No trials), but in fact partners spoke at the same time as the speakers. This method allowed us to collect (equivalent) data from both participants. The purpose of Experiment 2b was to replicate Experiment 2a's findings without deception (i.e., when speakers held a true, rather than a false belief that their partner would speak after them), to rule out the possibility that participants may have been aware that in fact their partner always spoke at the same time as them; e.g., by picking up movement cues in peripheral vision (see Method below).

Method

Participants. An additional 42 participants from the same population as in Experiment 1 took part in Experiment 2a, and were assigned to pairs; one pair was excluded due to experimenter error, leaving 20 pairs for analysis. A further 40 participants, divided into 20 pairs, took part in Experiment 2b. In Experiment 2b, half of the pairs were made up of two friends, and half were

made up of strangers. This factor did not interact with any within-participant manipulation, and will not be discussed further. Data from two pairs were excluded due to experimenter error, leaving 18 pairs for analysis.

Materials and Design. These were the same as in Experiment 1.

Procedure. The procedure was the same as in Experiment 1, except as follows. In Experiment 2a, both participants in a pair described the picture as soon as it appeared on the screen, as in Experiment 1. However, participants were told that their partner would speak after them, and would do so when a green square (i.e., a go signal) appeared on the screen. The go signal was displayed (for 400 ms) 2,500 ms after the picture appeared on the screen (i.e., the picture was displayed for 1,500 ms, as in Experiment 1, and then a blank screen was displayed for 1,000 ms). The timing of this go signal was set based on the sum of average latencies and durations in Experiment 1, which were just less than 2,500 ms. Finally, the go signal was followed by a blank screen for 2,500 ms. At the start of the session, participants were told that the computer would randomly assign one of them to the role of first speaker, and the other one to the role of second speaker; in fact, they were both informed that they had been assigned the role of first speaker, so they both believed their partner was speaking after them (while in reality they spoke at the same time). In Experiment 2b, participants saw the same sequence of events on each trial, but now one participant in each pair was randomly assigned the role of second speaker at the start of the experiment, and they actually produced their descriptions only after the go signal (the green square) appeared on the screen.

Recording and data analysis. These were also the same as in Experiment 1. In Experiment 2a, we analysed data from both participants (as in Experiment 1). However, in Experiment 2b we analysed data only from first speakers since the aim was to investigate effects of representing another's act of production on speech production processes, and we could not be sure at what point in time second speakers planned their utterance (i.e., immediately as the picture appeared, or closer to the go signal, as they got ready to speak).

Results

We first discarded incorrect responses. In Experiment 2a, they accounted for 18.01% of the data in the Different condition, 17.85% in the Same condition, and 17.19% in the No condition; in Experiment 2b they accounted for 28.39% of the data in the Different condition, 29.95% in the Same condition, and 29.51% in the No condition. As in Experiment 1, the error rate was not affected by any of our manipulations in either Experiment 2a (see Tables S2.1-4, Online supplementary material) nor in Experiment 2b (see Tables S3.1-4, Online supplementary materials), so we do not discuss it further.[2] In Experiment 2a, we then discarded 14 outliers from the description latency analysis, and 4 outliers from the utterance duration analysis; the corresponding numbers for Experiment 2b were 9 and 7, respectively.

Description Latency. In Experiment 2a (Figure 2, middle panel and Table 1), speakers took longer to begin speaking (M = 1,036 ms, SD = 324 ms) when they believed their partner would produce either the same or a different sentence after them, compared to when they believed their partner would be silent (M = 1,024 ms, SD = 315 ms; Speaking: B = 14.76, SE = 7.10, t = 2.08, CI

[2] The error rates were higher in Experiment 2b than in Experiment 2a. It is unclear what accounted for this difference; participants performed very similar tasks and underwent the same amount of practice at the start of the experiment. Perhaps more significantly, error rates were fairly high (>10%) across all experiments (cf. Segaert et al., 2011, 2016 – who used a similar paradigm to elicit production of actives and paasives – and reported error rates around 6-7%). This high error rate was probably down to some of our pictures being highly confusable (e.g., *nun* and *monk, sailor* and *soldier*). Please visit the stimuli&eprime_files folder in the OSF repository https://osf.io/v7t5z/ to view all pictures.

=[0.83,28.68])[3], thus indicating that they experienced interference from their partner's imagined act of production even though production and imagined act appeared asynchronous. Unexpectedly, speakers also took longer to begin speaking when they believed their partner would be producing the same utterance as themselves, compared to when they believed their partner would be producing a different utterance (see Table 1; Form: B = 18.71, SE = 7.88, t = 2.37, CI = [3.26,34.17]). As can be seen in Figure 1, middle panel, interference appeared to occur primarily in the Same condition (and numerically the average latency in the Different condition is very similar to the average for the No condition).

Importantly, Experiment 2b suggests that these differences between Experiment 1 and 2a are unlikely to be related to the fact that participants believed their partner would be speaking after them. This is because Experiment 2b revealed a pattern similar to the one in Experiment 1 (see Table 1 and Figure 2, right panel): longer latencies when the partner produced a description after the participant (M = 966 ms, SD = 391 ms) compared to when the partner did not produce a description at all (M = 930 ms, SD = 366 ms; Speaking: B = 39.99, SE = 13.92, t = 2.87, CI = [12.72,67.26]), but similar latencies regardless of whether the partner produced the same or a different utterance (Form: B = 11.68, SE = 14.79, t = 0.79, CI = [-17.31,40.67]; see Table 1 and Figure 2, right panel).

As in Experiment 1, description latencies were not affected by whether speakers were about to produce an active or a passive (Exp. 2a: active M = 1,033 ms, SD = 320 ms vs. passive M =

[3] Speaking was significant also when we replaced data points above and below 3SD from each by-participant average with the cut-off value (t = 2.06), but not (t = 1.95) when these data points were discarded (1.35%). It is possible that the effect of Speaking was driven by a few extreme data points, but Bayesian ex-Gaussian analyses did not reveal any effect of Speaking on the rate of the exponential component, nor on the mean of the Gaussian component.

1,031 ms, SD = 322 ms; Exp. 2b: active M = 959 ms, SD = 397 ms vs. passive M = 950 ms, SD = 369 ms), as confirmed by the lack of any significant effects involving Voice. Experiment 2a: Voice, B = 3.81, SE = 17.38, t = 0.22, CI = [-30.26,37.88]; Voice X Speaking, B = 7.42, SE = 13.03, t = 0.57, CI = [-18.12,32.97]; Voice X Form, B = 1.79, SE = 15.09, t = 0.12, CI = [-27.79,31.37]. Experiment 2b: Voice, B = -15.44, SE = 22.24, t = -0.69, CI = [-59.03,28.15]; Voice X Speaking, B = 55.93, SE = 39.25, t = 1.42, CI = [-21.00,132.86]; Voice X Form, B =-12.87, SE = 37.93, t = -0.34, CI = [-87.20,61.47]; see Table 1.

Utterance Duration. Unlike Experiment 1, there was no indication that representing the partner's task affected how long participants spent describing the pictures, in either Experiment 2a (see Figure 2, middle panel) or 2b (see Figure 2, right panel). Utterance durations were highly similar in all three conditions (Table 2). Accordingly, neither Speaking (Exp 2a: B = -2.35, SE = 7.28, t = -0.32, CI = [-16.62,11.91]; Exp. 2b: B = -14.41, SE = 16.91, t = -0.85, CI = [-47.56,18.73]) nor Form (Exp 2a: B = -3.65, SE = 8.15, t = -0.45, CI = [-19.62,12.33]; Exp 2b: B = -16.27, SE = 19.31, t = -0.84, CI = [-54.12,21.58]) was significant.

Unsurprisingly, passive utterances had longer durations than active utterances (Exp 2a: Passive, M = 1,653 ms, SD = 350 ms; Active, M = 1,427 ms, SD = 319 ms; Exp 2b: Passive, M = 1,744 ms, SD = 426 ms; Active, M = 1,606 ms, SD = 439 ms), as confirmed by a significant main effect of Voice in both experiments (Exp 2a: B = 225.72, SE = 12.73, t= 17.73, CI = [200.78,250.67]; Exp 2b: B = 142.89, SE = 19.48, t= 7.33, CI = [104.71,181.08]), but there were no further interactions with Voice. Exp 2a: Speaking X Voice, B = -26.83, SE = 13.75, t = -1.95, CI = [-53.77,0.11]; Form X Voice: B = 10.18, SE = 15.42, t = 0.66, CI = [-20.04,40.41]). Exp 2b: Speaking X Voice, B = -30.11, SE = 32.14, t = -0.94, CI = [-93.10,32.88]; Form X Voice: B = 51.95, SE = 37.00, t = 1.40, CI = [-20.58,124.48]; see Table 2.

Discussion

In Experiment 2, we tested whether speakers experience interference from representing their partner's utterance when they believe that their partner will speak after them, rather than at the same time as them. We were motivated to do so to explore the nature of other-representation. If speakers represent others' utterances at the time when they believe such utterances are being produced, then we should have found no evidence for interference when speakers believe their partner will speak after them. However, in Experiment 2a we found that speakers took longer to begin speaking when they believed their partner would speak after them, compared to when they believed their partner would remain silent, a finding which was further replicated in Experiment 2b. Thus, our findings suggest that speakers represent whether their partner intends to speak, and they represent their partner's intention to speak in the future much in the same way as they represent their intention to speak at the same time as them. In other words, speakers' representations of others' utterances do not distinguish between synchronous or asynchronous speaking – that is, speakers represent that their partners are or are about to speak but not precisely when they will speak. Accordingly, evidence for joint interference was present in the asynchronous task just as in the synchronous task used in Experiment 1.

In Experiment 2a, latencies were also longer when speakers believed their partner was producing an utterance in the same, compared to in a different voice, but this effect was not replicated in Experiment 2b, where latencies did not differ as a function of what the partner was instructed to say. Taken together, the findings from Experiment 2 thus confirm that, when participants are preparing to speak while they represent their partner's act of production (which was the case in both our synchronous and our asynchronous task), concurrent production limits the resources available for other-representation. As a result, we do not have clear evidence that speakers typically represent any detailed aspect of their partner's utterance planning (e.g., retrieval of specific syntactic representations). This conclusion is clearly supported by Experiment 2b, where neither latencies nor utterance durations were affected by whether the partner would produce a different or the same utterance as the participant. Utterance durations were also unaffected by this

manipulation in Experiment 2a, and although description latencies were affected in this experiment, speakers were actually *slower* when they believed their partner would produce the same utterance, rather than a different utterance. In sum, there was no indication that speakers experienced more conflict when representing an utterance that was different from the one they were preparing to produce.

Note that there were other differences between Experiment 2a and 2b. For example, speech latencies were somewhat shorter in Experiment 2b, but utterance duration was longer, possibly suggesting that speakers in Experiment 2b planned less before starting to speak and more while speaking compared to speakers in Experiment 2a. But overall, the pattern of findings was remarkably similar in Experiment 2a and 2b. In both experiments, the partner's task did not affect utterance durations and speech latencies were no longer when partners prepared to produce a different compared to the same utterance; however, speech latencies were shorter when partners remained silent.

General Discussion

In three experiments, we showed that speakers take longer to begin producing a sentence when they believe their partner intends to speak, compared to when they believe their partner does not intend to speak. Importantly, in all experiments we observed this interference effect between production and representation of another's act of production, even though speakers were asked to plan a whole sentence in either the active or the passive form – a task which is more demanding than tasks used in previous research. Moreover, we observed interference whether speakers believed their partner would speak at the same time as them (Experiment 1) or after them (Experiments 2a and 2b), suggesting that co-representations of other speakers' utterances do not include timing (see Pickering and Garrod, 2021, chapter 9).

In contrast, across the three experiments there was little indication that speakers represented the content of others' utterances. In Experiment 1, utterance durations were longer when participants believed their partner was producing a different (rather than the same) utterance at the same time. This finding may reflect interference due to conflict between active and passive syntactic representations, thus suggesting that, given more time, speakers can represent *what* their partner is saying. Importantly, however, this finding was not replicated in either Experiment 2a or 2b. While it is possible that this discrepancy between Experiment 1 and Experiment 2 is related to the change from a synchronous to an asynchronous task, we propose this is highly unlikely, because our findings are otherwise consistent across the synchronous and the asynchronous tasks, with both showing evidence of co-representation of the partner's intention to speak, but not of the content of their utterance. Overall, we have now found this same pattern of results consistently across six different experiments (three reported here, three in Gambi et al., 2015). Similarly, we think it is unlikely that utterance durations are affected differently from speech onset latencies by co-representation. Although the clearest evidence for detailed co-representation came from utterance durations in Experiment 1, the other two experiments did not replicate this finding and the most likely conclusion based on these data and the rest of the literature (Gambi et al., 2015; Brehm et al., 2019) is that speakers do not engage in detailed co-representation of others' utterances while they prepare to speak themselves.

In this study, our first aim was to clarify to what extent speakers are able to imagine others' utterances while they are concurrently speaking. In Gambi et al. (2015), we showed that speakers represent that another person is engaging in lexical retrieval while they concurrently prepare to produce a simple one- or two-word utterance. This is one of several studies that support utterance co-representation – that is, that people represent others' utterances in a format that is similar to that in which they represent their own utterances. In the Introduction, we noted that some studies support a stronger version of co-representation, in which speakers represent detailed aspects of others' utterances (e.g., frequency, semantic category), and appear to be engaging in at least partial

planning of others' utterances (Baus et al., 2014; Kuhlen & Abdel Rahman, 2017). However, other studies support a weaker version of co-representation, where speakers' production processes are affected by the fact that another person is engaging in language production but not by the content of what they are producing: Rather than planning others' utterances, speakers appear to merely represent the fact that others are speaking or about to speak (Gambi et al., 2015); in some cases, speakers may not represent their partners' utterance at all (Hoedemaker & Meyer, 2018).

We suggest that these apparent discrepancies can be reconciled if task differences are carefully considered. Specifically, we argue that content co-representation (i.e., detailed representation of what the partner is preparing to say) occurs only when speakers are not concurrently tasked with preparing to speak. In contrast, a weaker form of co-representation, co-representation of the partner's intention to speak, takes place when concurrent speaking is required, because production is cognitively demanding and limits the resources available for co-representation.

In our experiments, we aimed to probe the limits of co-representation by asking speakers to produce more complex utterances than used in previous research. While a direct comparison to previous work is not possible, the average latencies across all three experiments in this study were longer than in any of the picture naming studies reported by Gambi et al. (2015). For example, latencies in Experiment 1 of this study were about 70 ms longer than in Gambi et al.'s Experiment 1. Despite this increase in complexity, speakers in all three experiments still clearly engaged in the weaker form of co-representation, indicating that they are able to at least represent whether another is speaking in a wider range of situations than had previously been considered. Since in most instances of joint language production, some degree of overlap between planning of one's own utterances and representation of another's utterances is necessary (not only in choric speech, but also when taking turns; Levinson, 2016), this finding is important because it shows that a weak form of co-representation is possible even while concurrently planning a fairly complex utterance.

In addition, we showed that, while speakers represent whether their partner is planning to speak or not, this representation mechanism does not include information about timing. This suggests that this form of weak co-representation could affect coordination both in situations when people speak simultaneously (e.g., choric speech) and in those where people take turns (e.g., in conversation). There is already evidence that people slow down during choric speech (Cummins, 2009), and we suggest that interference between representation of the other's utterance and production of one's own utterance may contribute to this slowing down.

In conversation, it may be that this weak co-representation mechanism helps listeners who are preparing to initiate their turn to delay their entry when they detect the current speaker's intention to carry on speaking, or indeed another interlocutor's intention to take the floor. Thus, it may help prevent simultaneous starts during conversation. Future research could test these proposals using versions of our task that more closely resemble natural conversational exchanges, for example by incorporating a turn-taking component.

Finally, the finding that people represent future utterances in much the same way as current utterances is compatible with the idea that co-representation uses similar mechanisms as prediction, and that both rely on taking the production system "offline", and using it to simulate another speaker (Gambi & Pickering, 2016). Our findings are consistent with the proposal that representation of others' utterances uses the production system, because we have shown interference between other-representation and language production. However we have not conclusively demonstrated that the production system is directly involved in other-representation. This is because it is possible that other-representation instead relied on a different, non-language specific mechanism (e.g., rehearsal of a working memory representation of the other's task instruction), which also happened to (indirectly) affect the speaker's act of production (i.e., because participants were also trying to rehearse their own task instruction). But importantly, this alternative explanation is rendered much less likely by the fact that in previous work we found that interference

was greater when speakers imagined the partner was engaged in lexical retrieval, compared to a semantic categorization task (Gambi et al., 2015; see also Brehm et al., 2019). It is therefore more likely that language production mechanisms are involved in other-representation (Gambi & Pickering, 2016).

Note that one finding is potentially inconsistent with the account we have presented. Speakers in Hoedemaker and Meyer (2018) were no slower at naming a picture when they believed their partner would name a second picture after them, as opposed to believing nobody would name that picture (they were instead much slower when they had to name the second picture themselves). This finding is in contrast to the delayed latencies in our Experiment 2. While there are several potential reasons for this discrepancy (as the studies used quite different tasks), one possibility is that speakers in Hoedemaker and Meyer's study were under the influence of two competing effects. On the one hand, they experienced interference due to representing that their partner was about to start speaking (as in our Experiment 2). On the other, they were under pressure to speed up production to avoid overlapping with their partner, which would have led them to invest extra resources to begin production more quickly. If this is correct, Hoedemaker and Meyer may have found no evidence for interference due to co-representation in consecutive production because this effect was masked by a competing pressure to speak quickly. In contrast, participants in our Experiment 2 were not required to coordinate their utterances (and in fact, could not hear each other at all), so that no pressure to speak quickly should have been present.

In sum, we have shown that speakers experience interference when they represent that another speaker also intends to speak while, or soon after, they speak. We propose that this joint interference effect occurs because speakers represent their partner's intention to speak and that the locus of the effect occurs when a decision is made to activate the language production system.

Acknowledgments

We thank Anna Catherine Mackenzie and Kristen Nelissen for assistance with data collection. CG was supported by a scholarship from the School of Philosophy, Psychology and Language Sciences, University of Edinburgh. JVdC was supported by an EU Erasmus scholarship to visit the University of the Edinburgh.

References

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. Retrieved from http://arxiv.org/pdf/1506.04967.pdf

Baus, C., Sebanz, N., de la Fuente, V., Branzi, F. M., Martin, C., & Costa, A. (2014). On predicting others' words: Electrophysiological evidence of prediction in speech production. *Cognition, 133*(2), 395-407.

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

Bock, K., & Levelt, W. J. M. (1994). Language Production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 945-984). San Diego: Academic Press.

Brehm, L., Taschenberger, L., & Meyer, A. (2019). Mental representations of partner task cause interference in picture naming. *Acta Psychologica, 199,* 102888.

Brown, A. S. (1981). Inhibition in cued retrieval. *Journal of Experimental Psychology: Human Learning and Memory, 7*(3), 204-215.

Bürkner, P.-C. (2016). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1-28.

Corps, R., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: listeners use content predictions to determine what to say but not when to say it. *Cognition, 175*, 77-95.

Cummins, F. (2003). Practice and performance in speech produced synchronously. *Journal of Phonetics, 31*(2), 139-148.

Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics, 37*(1), 16-28.

Demiral, Ş. B., Gambi, C., Nieuwland, M. S., & Pickering, M. J. (2016). Neural correlates of verbal joint action: ERPs reveal common perception and action systems in a shared-Stroop task. *Brain Research, 1649*, 79-89.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30*(2), 210-233.

Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language, 33*, 715-715.

Gambi, C., & Pickering, M. J. (2011). A cognitive archietcture for the coordination of utterances. *Frontiers in Psychology, 2*. doi:10.3389/fpsyg.2011.00275

Gambi, C., & Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in Psychology, 4*, doi:10.3389/fpsyg.2013.00340.

Gambi, C., & Pickering, M. J. (2016). Predicting and imagining language. *Language, Cognition and Neuroscience, 31*(1), 60-72.

Gambi, C., Van de Cavey, J., & Pickering, M. J. (2015). Interference in joint picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(1), 1-21.

Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Frontiers in psychology, 6*, 1492. https://doi.org/10.3389/fpsyg.2015.01492

Hoedemaker, R. S., Ernst, J., Meyer, A. S., & Belke, E. (2017). Language production in a shared task: Cumulative semantic interference from self-and other-produced context words. *Acta Psychologica, 172*, 55-63.

Hoedemaker, R. S., & Meyer, A. S. (2018). Planning and coordination of utterances in a joint naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition. Advance online publication*. doi:10.1037/xlm0000603.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition, 100*(3), 464-482.

Jasmin, K. M., McGettigan, C., Agnew, Z. K., Lavan, N., Josephs, O., Cummins, F., & Scott, S. K. (2016). Cohesion and joint speech: right hemisphere contributions to synchronized vocal production. *Journal of Neuroscience, 36*(17), 4669-4680.

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 54, pp. 59-101). Burlington: Academic Press.

Kuhlen, A. K., & Abdel Rahman, R. (2017). Having a task partner affects lexical retrieval: Spoken word production in shared task settings. *Cognition, 166*, 94-106.

Kuhlen, A. K., & Abdel Rahman, R. (2021). Joint language production: An electrophysiological investigation of simulated lexical access on behalf of a task partner.. *Journal of Experimental Psychology: Language, Memory, and Cognition.* https://doi.org/10.1037/xlm0001025

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

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

Magyari, L., Bastiaansen, M. C., De Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience, 26(11),* 2530-2539.

Menenti, L., Gierhan, S., Segaert, K., & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by fMRI. *Psychological Science, 22*(9), 1173-1182.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language, 35*(4), 477-496.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition, 114*(2), 227-252.

Pickering, M. J., & Garrod, S. (2021). *Understanding dialogue: Language use and social interaction.* Cambridge, UK: Cambridge University Press.

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

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696-735.

Schmitz, L., Vesper, C., Sebanz, N., & Knoblich, G. (2017). Co-representation of others' task constraints in joint action. *Journal of Experimental Psychology: Human Perception and Performance, 43*(8), 1480-1493.

Schmitz, L., Vesper, C., Sebanz, N., & Knoblich, G. (2018). Co-actors represent the order of each other's actions. *Cognition, 181*, 65-79.

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition, 88*(3), B11-B21.

Sebanz, N., Knoblich, G., & Prinz, W. (2005). How two share a task: Corepresenting stimulus-response mappings. *Journal of Experimental Psychology: Human Perception and Performance, 31*(6), 1234-1246.

Segaert K., Menenti L., Weber K., & Hagoort P. (2011) A paradox of syntactic priming: Why response tendencies show priming for passives, and response latencies show priming for actives. *PLoS ONE 6(10): e24209.* https://doi.org/10.1371/journal.pone.0024209

Segaert, K., Wheeldon, L., & Hagoort, P. (2016). Unifying structural priming effects on syntactic choices and timing of sentence generation. *Journal of Memory and Language, 91*, 59-80.

Smith, M., & Wheeldon, L. (2004). Horizontal information flow in spoken language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 675-686.

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.

Strijkers, K., Holcomb, P. J., & Costa, A. (2011). Conscious intention to speak proactively facilitates lexical access during overt object naming. *Journal of Memory and Language, 65*(4), 345-362.

Supplementary Material

1. Experiment 1 Models

1.1 Accuracy

1.1.1 Full model

```
## correct ~ (SPEAKING + FORM) * VOICE + (1 + (SPEAKING + FORM) * VOICE || Part
icipant) +
##     (1 + (SPEAKING + FORM) * VOICE || Item)
```

Table S1.1. Accuracy analysis - Full model fixed effects

	B	SE	z	P value
(Intercept)	2.2814	0.1614	14.1382	0.0000
SPEAKING	-0.0933	0.0832	-1.1214	0.2621
FORM	0.1123	0.0991	1.1328	0.2573
VOICE	-0.0253	0.0928	-0.2721	0.7855
SPEAKING:VOICE	-0.2861	0.1576	-1.8161	0.0694
FORM:VOICE	0.2316	0.1872	1.2372	0.2160

Table S1.2. Accuracy analysis - Full model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	0.6868	0.8288
Participant.1	SPEAKING	0.0000	0.0000
Participant.2	FORM	0.0592	0.2432
Participant.3	VOICE	0.0216	0.1471
Participant.4	SPEAKING:VOICE	0.0000	0.0000
Participant.5	FORM:VOICE	0.0000	0.0000
Item	(Intercept)	0.2008	0.4481
Item.1	SPEAKING	0.0191	0.1381
Item.2	FORM	0.0000	0.0000
Item.3	VOICE	0.0760	0.2756
Item.4	SPEAKING:VOICE	0.0000	0.0000
Item.5	FORM:VOICE	0.0904	0.3007

1.1.2 Reduced model

```
## correct ~ (SPEAKING + FORM) * VOICE + (1 + FORM + VOICE || Participant) +
##     (1 + SPEAKING + VOICE + FORM:VOICE || Item)
```

Table S1.3. Accuracy analysis - Reduced model fixed effects

	B	SE	z	P value
(Intercept)	2.2814	0.1614	14.1377	0.0000
SPEAKING	-0.0933	0.0832	-1.1214	0.2621
FORM	0.1123	0.0991	1.1328	0.2573
VOICE	-0.0253	0.0928	-0.2721	0.7855
SPEAKING:VOICE	-0.2861	0.1576	-1.8161	0.0694
FORM:VOICE	0.2316	0.1872	1.2372	0.2160

Table S1.4. Accuracy analysis - Reduced model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	0.6868	0.8288
Participant.1	FORM	0.0592	0.2433
Participant.2	VOICE	0.0216	0.1471
Item	(Intercept)	0.2008	0.4481
Item.1	SPEAKING	0.0191	0.1381
Item.2	VOICE	0.0760	0.2756
Item.3	VOICE:FORM	0.0904	0.3007

1.2 Onset latencies

1.2.1 Full model

```
## Onset * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     SPEAKING | Participant) + (0 + FORM | Participant) + (0 + VOICE |
##     Participant) + (0 + SPEAKING:VOICE | Participant) + (0 + FORM:VOICE |
##     Participant)) + ((1 | Item) + (0 + SPEAKING | Item) +
##     (0 + FORM | Item) + (0 + VOICE | Item) +
##     (0 + SPEAKING:VOICE | Item) + (0 + FORM:VOICE | Item))
```

Table S1.5. Latency analysis - Full model fixed effects

	B	SE	t
(Intercept)	951.3161	20.1267	47.2665
SPEAKING	19.1246	5.5767	3.4294
FORM	3.6136	6.8891	0.5245
VOICE	-0.7959	11.9297	-0.0667
SPEAKING:VOICE	2.9311	10.8582	0.2699
FORM:VOICE	14.3175	13.7002	1.0451

Table S1.6. Latency analysis - Full model random effects

grp	var1	vcov	sdcor

Participant	(Intercept)	14059.0786	118.5710
Participant.1	SPEAKING	0.0000	0.0000
Participant.2	FORM	303.6092	17.4244
Participant.3	VOICE	420.2581	20.5002
Participant.4	SPEAKING:VOICE	0.0000	0.0000
Participant.5	FORM:VOICE	0.0000	0.0000
Item	(Intercept)	1212.6812	34.8236
Item.1	SPEAKING	51.6555	7.1872
Item.2	FORM	0.0000	0.0000
Item.3	VOICE	3361.7043	57.9802
Item.4	SPEAKING:VOICE	0.0000	0.0000
Item.5	FORM:VOICE	934.9292	30.5766
Residual	NA	43030.7887	207.4386

1.2.2 Reduced model

```
## Onset * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     FORM | Participant) + (0 + VOICE | Participant)) + ((1 | Item) +
##     (0 + VOICE | Item) + (0 + VOICE:FORM | Item))
```

Table S1.7. Latency analysis - Reduced model fixed effects

	B	SE	t
(Intercept)	951.3148	20.1267	47.2664
SPEAKING	19.1272	5.4302	3.5223
FORM	3.6117	6.8899	0.5242
VOICE	-0.7932	11.9300	-0.0665
SPEAKING:VOICE	2.9456	10.8593	0.2713
FORM:VOICE	14.3258	13.7021	1.0455

Table S1.8. Latency analysis - Reduced model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	14059.2664	118.5718
Participant.1	FORM	303.6078	17.4243
Participant.2	VOICE	420.5844	20.5082
Item	(Intercept)	1212.4748	34.8206
Item.1	VOICE	3361.4553	57.9781
Item.2	VOICE:FORM	935.3207	30.5830
Residual	NA	43042.0946	207.4659

Table S1.9. Latency analysis - Confidence intervals for fixed effects from the reduced model

	2.5 %	97.5 %
SPEAKING	8.4841	29.7702

FORM	-9.8922	17.1156
VOICE	-24.1755	22.5891
SPEAKING:VOICE	-18.3382	24.2295
FORM:VOICE	-12.5298	41.1813

1.2.3 Reduced model, Windsorised dataset

```
## Onset * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     FORM | Participant) + (0 + VOICE | Participant)) + ((1 | Item) +
##     (0 + VOICE | Item) + (0 + VOICE:FORM | Item))
```

Table S1.10. Latency analysis on Windsorised dataset - Reduced model fixed effects

	B	SE	t
(Intercept)	948.9404	20.0195	47.4007
SPEAKING	18.6368	5.1608	3.6112
FORM	4.1080	6.4481	0.6371
VOICE	0.3217	11.7894	0.0273
SPEAKING:VOICE	1.3063	10.3206	0.1266
FORM:VOICE	11.5800	12.9410	0.8948

Table S1.11. Latency analysis on Windsorised dataset - Reduced model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	13962.3944	118.1626
Participant.1	FORM	224.0161	14.9672
Participant.2	VOICE	456.8547	21.3742
Item	(Intercept)	1175.0775	34.2794
Item.1	VOICE	3306.6196	57.5032
Item.2	VOICE:FORM	777.5469	27.8845
Residual	NA	38876.1696	197.1704

1.2.4 Reduced model, trimmed dataset

```
## Onset * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     VOICE | Participant)) + ((1 | Item) + (0 + VOICE | Item) +
##     (0 + VOICE:FORM | Item))
```

Table S1.12. Latency analysis on trimmed dataset - Reduced model fixed effects

	B	SE	t
(Intercept)	940.3876	19.3514	48.5954
SPEAKING	17.8343	4.8057	3.7111
FORM	5.7859	5.5781	1.0372
VOICE	1.7859	11.1681	0.1599
SPEAKING:VOICE	-2.9053	9.6105	-0.3023
FORM:VOICE	1.4959	12.0145	0.1245

Table S1.13. Latency analysis on trimmed dataset - Reduced model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	13084.4138	114.3871
Participant.1	VOICE	392.6732	19.8160
Item	(Intercept)	1079.1512	32.8504
Item.1	VOICE	3003.5463	54.8046
Item.2	VOICE:FORM	634.7572	25.1944
Residual	NA	33272.8473	182.4085

1.3 Duration

1.3.1 Full model

```
## Length * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     SPEAKING | Participant) + (0 + FORM | Participant) + (0 + VOICE |
##     Participant) + (0 + SPEAKING:VOICE | Participant) + (0 + FORM:VOICE |
##     Participant)) + ((1 | Item) + (0 + SPEAKING | Item) +
##     (0 + FORM | Item) + (0 + VOICE | Item) +
##     (0 + SPEAKING:VOICE | Item) + (0 + FORM:VOICE | Item))
```

Table S1.14. Duration analysis - Full model fixed effects

	B	SE	t
(Intercept)	1456.3810	27.9212	52.1603
SPEAKING	8.9355	6.1025	1.4642
FORM	-20.6079	7.9218	-2.6014
VOICE	167.7331	12.2134	13.7336
SPEAKING:VOICE	-7.3953	12.9114	-0.5728
FORM:VOICE	0.4677	15.4185	0.0303

Table S1.15. Duration analysis - Full model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	20420.1177	142.8990
Participant.1	SPEAKING	0.0000	0.0000
Participant.2	FORM	0.0000	0.0000
Participant.3	VOICE	1587.5262	39.8438
Participant.4	SPEAKING:VOICE	0.0000	0.0000
Participant.5	FORM:VOICE	0.0000	0.0000
Item	(Intercept)	7921.2024	89.0011
Item.1	SPEAKING	0.0000	0.0000
Item.2	FORM	405.4080	20.1347
Item.3	VOICE	2396.5417	48.9545

Item.4	SPEAKING:VOICE	567.5807	23.8240
Item.5	FORM:VOICE	1198.0885	34.6134
Residual	NA	54405.5534	233.2500

1.3.2 Reduced model

```
## Length * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     VOICE | Participant)) + ((1 | Item) + (0 + FORM |
##     Item) + (0 + VOICE | Item) + (0 + FORM:VOICE |
##     Item) + (0 + VOICE:SPEAKING | Item))
```

Table S1.16. Duration analysis - Reduced model fixed effects

	B	SE	t
(Intercept)	1456.3810	27.9213	52.1603
SPEAKING	8.9355	6.1025	1.4642
FORM	-20.6079	7.9218	-2.6014
VOICE	167.7331	12.2134	13.7336
SPEAKING:VOICE	-7.3953	12.9114	-0.5728
FORM:VOICE	0.4677	15.4185	0.0303

Table S1.17. Duration analysis - Reduced model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	20420.1673	142.8992
Participant.1	VOICE	1587.5276	39.8438
Item	(Intercept)	7921.2127	89.0012
Item.1	FORM	405.4102	20.1348
Item.2	VOICE	2396.5415	48.9545
Item.3	FORM:VOICE	1198.0446	34.6128
Item.4	VOICE:SPEAKING	567.5713	23.8238
Residual	NA	54405.5538	233.2500

Table S1.18. Duration analysis - Confidence intervals for fixed effects from the reduced model

	2.5 %	97.5 %
SPEAKING	-3.0252	20.8962
FORM	-36.1344	-5.0815
VOICE	143.7953	191.6709
SPEAKING:VOICE	-32.7012	17.9107
FORM:VOICE	-29.7520	30.6874

1.3.3 Reduced model, Windsorised dataset

```
## Length * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     VOICE | Participant)) + ((1 | Item) + (0 + FORM |
```

```
##     Item) + (0 + VOICE | Item) + (0 + FORM:VOICE |
##     Item) + (0 + VOICE:SPEAKING | Item))
```

Table S1.19. Duration analysis on Windsorised dataset - Reduced model fixed effects

	B	SE	t
(Intercept)	1453.9077	27.9111	52.0906
SPEAKING	9.0536	5.8401	1.5503
FORM	-18.2293	7.4118	-2.4595
VOICE	166.6002	11.9031	13.9963
SPEAKING:VOICE	-7.3064	12.1567	-0.6010
FORM:VOICE	-1.2151	14.5324	-0.0836

Table S1.20. Duration analysis on Windsorised dataset - Reduced model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	20620.1907	143.5973
Participant.1	VOICE	1589.6601	39.8705
Item	(Intercept)	7761.8113	88.1011
Item.1	FORM	290.2678	17.0372
Item.2	VOICE	2245.4322	47.3860
Item.3	FORM:VOICE	888.8269	29.8132
Item.4	VOICE:SPEAKING	363.6038	19.0684
Residual	NA	49828.8702	223.2238

1.3.4 Reduced model, trimmed dataset

```
## Length * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     VOICE | Participant)) + ((1 | Item) + (0 + FORM |
##     Item) + (0 + VOICE | Item))
```

Table S1.21. Duration analysis on trimmed dataset - Reduced model fixed effects

	B	SE	t
(Intercept)	1445.4406	27.7728	52.0451
SPEAKING	7.8967	5.5182	1.4310
FORM	-7.8875	7.0733	-1.1151
VOICE	167.8832	11.0789	15.1535
SPEAKING:VOICE	-4.4769	11.0354	-0.4057
FORM:VOICE	-9.2190	12.8049	-0.7200

Table S1.22. Duration analysis on trimmed dataset - Reduced model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	20881.4716	144.5042

Participant.1	VOICE	1574.5565	39.6807
Item	(Intercept)	7327.2468	85.5993
Item.1	FORM	288.1681	16.9755
Item.2	VOICE	1755.5859	41.8997
Residual	NA	44010.4539	209.7867

2. Experiment 2a Models

2.1 Accuracy

2.1.1 Full model

```
## correct ~ (SPEAKING + FORM) * VOICE + (1 + (SPEAKING + FORM) * VOICE || Part
icipant) +
##     (1 + (SPEAKING + FORM) * VOICE || Item)
```

Table S2.1 Accuracy analysis - Full model fixed effects

	B	SE	z	P value
(Intercept)	1.8396	0.1745	10.5395	0.0000
SPEAKING	-0.0601	0.0685	-0.8776	0.3801
FORM	0.0125	0.0783	0.1591	0.8736
VOICE	-0.0215	0.0748	-0.2869	0.7742
SPEAKING:VOICE	0.0386	0.1371	0.2820	0.7780
FORM:VOICE	-0.0125	0.1566	-0.0797	0.9365

Table S2.2 Accuracy analysis - Full model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	0.5833	0.7638
Participant.1	SPEAKING	0.0000	0.0000
Participant.2	FORM	0.0000	0.0000
Participant.3	VOICE	0.0119	0.1089
Participant.4	SPEAKING:VOICE	0.0000	0.0000
Participant.5	FORM:VOICE	0.0000	0.0000
Item	(Intercept)	0.4626	0.6801
Item.1	SPEAKING	0.0000	0.0000
Item.2	FORM	0.0000	0.0000
Item.3	VOICE	0.0324	0.1799
Item.4	SPEAKING:VOICE	0.0000	0.0000

Item.5	FORM:VOICE	0.0000	0.0000

2.1.2 Reduced model

```
## correct ~ (SPEAKING + FORM) * VOICE + (1 + VOICE || Participant) + (1 +
##     VOICE || Item)
```

Table S2.3 Accuracy analysis - Reduced modelfixed effects

	B	SE	z	P value
(Intercept)	1.8395	0.1745	10.5391	0.0000
SPEAKING	-0.0601	0.0685	-0.8776	0.3801
FORM	0.0125	0.0783	0.1590	0.8736
VOICE	-0.0215	0.0748	-0.2878	0.7735
SPEAKING:VOICE	0.0386	0.1370	0.2820	0.7780
FORM:VOICE	-0.0125	0.1566	-0.0797	0.9365

Table S2.4 Accuracy analysis - Reduced model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	0.5833	0.7638
Participant.1	VOICE	0.0116	0.1076
Item	(Intercept)	0.4626	0.6801
Item.1	VOICE	0.0323	0.1798

2.2 Onset latencies

2.2.1 Full model

```
## onset * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     SPEAKING | Participant) + (0 + FORM | Participant) + (0 + VOICE |
##     Participant) + (0 + SPEAKING:VOICE | Participant) + (0 + FORM:VOICE |
##     Participant)) + ((1 | Item) + (0 + SPEAKING | Item) +
##     (0 + FORM | Item) + (0 + VOICE | Item) +
##     (0 + SPEAKING:VOICE | Item) + (0 + FORM:VOICE | Item))
```

Table S2.5 Latency analysis - Full model fixed effects

	B	SE	t
(Intercept)	1031.4259	33.3826	30.8971
SPEAKING	14.7571	7.1043	2.0772
FORM	18.7147	7.8831	2.3740
VOICE	3.8072	17.3836	0.2190
SPEAKING:VOICE	7.4249	13.0320	0.5697
FORM:VOICE	1.7879	15.0925	0.1185

Table S2.6 Latency analysis - Full model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	40950.7059	202.3628
Participant.1	SPEAKING	0.0000	0.0000
Participant.2	FORM	204.5201	14.3011
Participant.3	VOICE	1605.8693	40.0733
Participant.4	SPEAKING:VOICE	0.0000	0.0000
Participant.5	FORM:VOICE	0.0000	0.0000
Item	(Intercept)	2584.4921	50.8379
Item.1	SPEAKING	253.8304	15.9321
Item.2	FORM	0.0000	0.0000
Item.3	VOICE	7142.9565	84.5160
Item.4	SPEAKING:VOICE	0.0000	0.0000
Item.5	FORM:VOICE	0.0000	0.0000
Residual	NA	59564.8255	244.0591

2.2.2 Reduced model

```
## onset * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     FORM | Participant) + (0 + VOICE | Participant)) + ((1 | Item) +
##     (0 + SPEAKING | Item) + (0 + VOICE | Item))
```

Table S2. 7 Latency analysis - Reduced model fixed effects

	B	SE	t
(Intercept)	1031.4259	33.3826	30.8971
SPEAKING	14.7571	7.1043	2.0772
FORM	18.7147	7.8831	2.3740
VOICE	3.8072	17.3836	0.2190
SPEAKING:VOICE	7.4249	13.0320	0.5697
FORM:VOICE	1.7879	15.0925	0.1185

Table S2.8 Latency analysis - Reduced model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	40950.6509	202.3627
Participant.1	FORM	204.5195	14.3010
Participant.2	VOICE	1605.8743	40.0734
Item	(Intercept)	2584.4897	50.8379
Item.1	SPEAKING	253.8302	15.9321
Item.2	VOICE	7142.9506	84.5160
Residual	NA	59564.8260	244.0591

Tabke S2.9 Latency analysis - Confidence intervals for fixed effects from the reduced model

	2.5 %	97.5 %
SPEAKING	0.8328	28.6813
FORM	3.2642	34.1652
VOICE	-30.2642	37.8785
SPEAKING:VOICE	-18.1173	32.9670
FORM:VOICE	-27.7929	31.3687

2.2.3 Reduced model, Windsorised dataset

```
## onset * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     FORM | Participant) + (0 + VOICE | Participant)) + ((1 | Item) +
##     (0 + SPEAKING | Item) + (0 + VOICE | Item))
```

Table S2.10 Latency analysis on Windsorised dataset - Reduced model fixed effects

	B	SE	t
(Intercept)	1028.3416	33.3930	30.7951
SPEAKING	13.9442	6.7706	2.0595
FORM	19.4448	7.2878	2.6681
VOICE	4.6364	17.1068	0.2710
SPEAKING:VOICE	8.6420	12.4513	0.6941
FORM:VOICE	-1.5591	14.4195	-0.1081

Table S2.11 Latency analysis on Windsorised dataset - Reduced model fixed effects

grp	var1	vcov	sdcor
Participant	(Intercept)	41224.4374	203.0380
Participant.1	FORM	44.4326	6.6658
Participant.2	VOICE	1552.2779	39.3990
Item	(Intercept)	2415.1678	49.1444
Item.1	SPEAKING	224.3135	14.9771
Item.2	VOICE	6987.9555	83.5940
Residual	NA	54375.3208	233.1852

2.2.4 Reduced model, trimmed dataset

```
## onset * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     FORM | Participant) + (0 + VOICE | Participant)) + ((1 | Item) +
##     (0 + SPEAKING | Item) + (0 + VOICE | Item))
```

Table S2.12 Latency analysis on trimmed dataset - Reduced model fixed effects

	B	SE	t
(Intercept)	1018.5577	33.3512	30.5404
SPEAKING	12.2348	6.2793	1.9484

FORM	18.0477	7.1234	2.5336
VOICE	3.9656	15.9651	0.2484
SPEAKING:VOICE	8.3684	11.7902	0.7098
FORM:VOICE	-4.4997	13.6665	-0.3292

Table S2.13 Latency analysis on trimmed dataset - Reduced model fixed effects

grp	var1	vcov	sdcor
Participant	(Intercept)	41485.8796	203.6808
Participant.1	FORM	159.3377	12.6229
Participant.2	VOICE	1473.4379	38.3854
Item	(Intercept)	2145.8463	46.3233
Item.1	SPEAKING	148.0801	12.1688
Item.2	VOICE	5957.4730	77.1847
Residual	NA	48126.1971	219.3768

2.3 Duration

2.3.1 Full model

```
## length * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     SPEAKING | Participant) + (0 + FORM | Participant) + (0 + VOICE |
##     Participant) + (0 + SPEAKING:VOICE | Participant) + (0 + FORM:VOICE |
##     Participant)) + ((1 | Item) + (0 + SPEAKING | Item) +
##     (0 + FORM | Item) + (0 + VOICE | Item) +
##     (0 + SPEAKING:VOICE | Item) + (0 + FORM:VOICE | Item))
```

Table S2.14 Duration analysis - Full model fixed effects

	B	SE	t
(Intercept)	1544.0397	35.5839	43.3915
SPEAKING	-2.3519	7.2781	-0.3231
FORM	-3.6471	8.1515	-0.4474
VOICE	225.7240	12.7282	17.7342
SPEAKING:VOICE	-26.8282	13.7463	-1.9517
FORM:VOICE	10.1813	15.4207	0.6602

Table S2.15 Duration analysis - Full model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	39304.0746	198.2526
Participant.1	SPEAKING	115.1509	10.7308
Participant.2	FORM	0.0000	0.0000
Participant.3	VOICE	3158.9848	56.2048
Participant.4	SPEAKING:VOICE	457.5001	21.3893
Participant.5	FORM:VOICE	0.0000	0.0000
Item	(Intercept)	8745.6826	93.5184
Item.1	SPEAKING	180.4875	13.4346

Item.2	FORM	221.6112	14.8866
Item.3	VOICE	1361.8279	36.9030
Item.4	SPEAKING:VOICE	0.0000	0.0000
Item.5	FORM:VOICE	0.0000	0.0000
Residual	NA	62282.1746	249.5640

2.2.2 Reduced model

```
## length * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     VOICE | Participant) + (0 + SPEAKING | Participant) + (0 + VOICE:SPEAKIN
G |
##     Participant)) + ((1 | Item) + (0 + FORM | Item) +
##     (0 + SPEAKING | Item) + (0 + VOICE | Item))
```

Table S2.16 Duration analysis - Reduced model fixed effects

	B	SE	t
(Intercept)	1544.0397	35.5839	43.3915
SPEAKING	-2.3519	7.2781	-0.3231
FORM	-3.6471	8.1515	-0.4474
VOICE	225.7240	12.7282	17.7342
SPEAKING:VOICE	-26.8282	13.7463	-1.9517
FORM:VOICE	10.1813	15.4207	0.6602

Table S2.17 Duration analysis - Reduced model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	39304.0238	198.2524
Participant.1	VOICE	3158.9834	56.2048
Participant.2	SPEAKING	115.1479	10.7307
Participant.3	VOICE:SPEAKING	457.4976	21.3892
Item	(Intercept)	8745.6852	93.5184
Item.1	FORM	221.6110	14.8866
Item.2	SPEAKING	180.4911	13.4347
Item.3	VOICE	1361.8337	36.9030
Residual	NA	62282.1745	249.5640

Table S2.18 Duration analysis - Confidence intervals for fixed effects from the reduced model

	2.5 %	97.5 %
SPEAKING	-16.6167	11.9129
FORM	-19.6239	12.3296
VOICE	200.7772	250.6707
SPEAKING:VOICE	-53.7704	0.1140
FORM:VOICE	-20.0427	40.4054

2.3.3 Reduced model, Windsorised dataset

```
## length * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     VOICE | Participant) + (0 + SPEAKING | Participant) + (0 + VOICE:SPEAKIN
G |
##     Participant)) + ((1 | Item) + (0 + FORM | Item) +
##     (0 + SPEAKING | Item) + (0 + VOICE | Item))
```

Table S2.19 Duration analysis on Windsorised dataset - Reduced model fixed effects

	B	SE	t
(Intercept)	1541.4789	35.5196	43.3980
SPEAKING	-2.1434	7.1448	-0.3000
FORM	-3.8599	7.7419	-0.4986
VOICE	223.9191	12.5855	17.7919
SPEAKING:VOICE	-26.4393	13.1944	-2.0038
FORM:VOICE	8.8128	14.8262	0.5944

Table S2.20 Duration analysis on Windsorised dataset - Reduced model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	39406.6709	198.5111
Participant.1	VOICE	3156.6094	56.1837
Participant.2	SPEAKING	63.0480	7.9403
Participant.3	VOICE:SPEAKING	399.7578	19.9939
Item	(Intercept)	8542.1224	92.4236
Item.1	FORM	157.7357	12.5593
Item.2	SPEAKING	267.8450	16.3660
Item.3	VOICE	1345.1369	36.6761
Residual	NA	57572.7015	239.9431

2.3.4 Reduced model, trimmed dataset

```
## length * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     VOICE | Participant)) + ((1 | Item) + (0 + SPEAKING |
##     Item) + (0 + VOICE | Item))
```

Table S2.21 Duration analysis on trimmed dataset - Reduced model fixed effects

	B	SE	t
(Intercept)	1531.4364	35.0321	43.7152
SPEAKING	-1.3612	6.6139	-0.2058
FORM	-4.6766	6.9521	-0.6727
VOICE	220.9798	12.4798	17.7070
SPEAKING:VOICE	-23.2309	12.0114	-1.9341
FORM:VOICE	0.1353	13.9037	0.0097

Table S2.22 Duration analysis on trimmed dataset - Reduced model random effects

grp	var1	vcov	sdcor
Participant	(Intercept)	39006.7608	197.5013
Participant.1	VOICE	3247.7541	56.9891

Item	(Intercept)	7797.8361	88.3054
Item.1	SPEAKING	242.9604	15.5872
Item.2	VOICE	1329.3686	36.4605
Residual	NA	50072.6402	223.7692

3. Experiment 2b Models

3.1 Accuracy

3.1.1 Full model

```
## correct ~ (SPEAKING + FORM) * VOICE + (1 + (SPEAKING + FORM) * VOICE || Part
icipant) +
##     (1 + (SPEAKING + FORM) * VOICE || Item)
```

Table S3.1 Accuracy analysis - Full model fixed effects

	B	SE	z	P value
(Intercept)	0.9970	0.1662	6.0000	0.0000
SPEAKING	0.0132	0.0937	0.1411	0.8878
FORM	-0.0857	0.0971	-0.8832	0.3771
VOICE	-0.0333	0.1120	-0.2975	0.7660
SPEAKING:VOICE	0.1534	0.1848	0.8303	0.4064
FORM:VOICE	-0.0766	0.1941	-0.3947	0.6930

Table S3.2 Accuracy analysis - Full model random effects

grp	var1	vcov	sdcor
Item	FORM:VOICE	0.0000	0.0000
Item.1	SPEAKING:VOICE	0.1801	0.4244
Item.2	VOICE	0.0507	0.2251
Item.3	FORM	0.0000	0.0000
Item.4	SPEAKING	0.0348	0.1867
Item.5	(Intercept)	0.3231	0.5684
Participant	FORM:VOICE	0.0000	0.0000
Participant.1	SPEAKING:VOICE	0.0000	0.0000
Participant.2	VOICE	0.0819	0.2861
Participant.3	FORM	0.0000	0.0000
Participant.4	SPEAKING	0.0098	0.0992
Participant.5	(Intercept)	0.2840	0.5330

3.1.2 Reduced model

```
## correct ~ (SPEAKING + FORM) * VOICE + (1 + VOICE + SPEAKING || Participant)
+
##     (1 + VOICE * SPEAKING || Item)
```

Table S3.3 Accuracy analysis - Reduced model fixed effects

	B	SE	z	P value
(Intercept)	0.9970	0.1662	6.0000	0.0000
SPEAKING	0.0132	0.0937	0.1411	0.8878
FORM	-0.0857	0.0971	-0.8832	0.3771
VOICE	-0.0333	0.1120	-0.2976	0.7660
SPEAKING:VOICE	0.1534	0.1848	0.8303	0.4064
FORM:VOICE	-0.0766	0.1941	-0.3947	0.6930

Table S3.4 Accuracy analysis - Reduced model random effects

grp	var1	vcov	sdcor
Item	VOICE:SPEAKING	0.1801	0.4244
Item.1	SPEAKING	0.0348	0.1867
Item.2	VOICE	0.0507	0.2251
Item.3	(Intercept)	0.3231	0.5684
Participant	SPEAKING	0.0098	0.0992
Participant.1	VOICE	0.0819	0.2861
Participant.2	(Intercept)	0.2840	0.5330

3.2 Onset latencies

3.2.1 Full model

```
## Onset * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     SPEAKING | Participant) + (0 + FORM | Participant) + (0 + VOICE |
##     Participant) + (0 + SPEAKING:VOICE | Participant) + (0 + FORM:VOICE |
##     Participant)) + ((1 | Item) + (0 + SPEAKING | Item) +
##     (0 + FORM | Item) + (0 + VOICE | Item) +
##     (0 + SPEAKING:VOICE | Item) + (0 + FORM:VOICE | Item))
```

Table S3.5 Latency analysis - Full model fixed effects

	B	SE	t
(Intercept)	965.5214	56.5406	17.0766
SPEAKING	39.9885	13.9153	2.8737
FORM	11.6819	14.7906	0.7898
VOICE	-15.4407	22.2422	-0.6942
SPEAKING:VOICE	55.9276	39.2511	1.4249
FORM:VOICE	-12.8666	37.9264	-0.3393

Table S3.6 Latency analysis - Full model random effects

grp	var1	vcov	sdcor
Item	FORM:VOICE	11054.6654	105.1412
Item.1	SPEAKING:VOICE	12143.1161	110.1958

Item.2	VOICE	8612.5713	92.8039
Item.3	FORM	0.0000	0.0000
Item.4	SPEAKING	891.7201	29.8617
Item.5	(Intercept)	4209.5430	64.8810
Participant	FORM:VOICE	3763.3724	61.3463
Participant.1	SPEAKING:VOICE	8769.5099	93.6457
Participant.2	VOICE	1350.0560	36.7431
Participant.3	FORM	0.0000	0.0000
Participant.4	SPEAKING	0.0000	0.0000
Participant.5	(Intercept)	54480.4530	233.4105
Residual	NA	88076.3500	296.7766

3.2.2 Reduced model

```
## Onset * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     VOICE | Participant) + (0 + VOICE:SPEAKING | Participant) + (0 +
##     VOICE:FORM | Participant)) + ((1 | Item) + (0 + SPEAKING |
##     Item) + (0 + VOICE | Item) + (0 + SPEAKING:VOICE |
##     Item) + (0 + VOICE:FORM | Item))
```

Table S3.7 Latency analysis - Reduced model fixed effects

	B	SE	t
(Intercept)	965.5214	56.5406	17.0766
SPEAKING	39.9885	13.9153	2.8737
FORM	11.6819	14.7906	0.7898
VOICE	-15.4407	22.2422	-0.6942
SPEAKING:VOICE	55.9276	39.2511	1.4249
FORM:VOICE	-12.8666	37.9264	-0.3393

Table S3.8 Latency analysis - Reduced model random effects

grp	var1	vcov	sdcor
Item	VOICE:FORM	11054.6824	105.1412
Item.1	SPEAKING:VOICE	12143.2008	110.1962
Item.2	VOICE	8612.5752	92.8040
Item.3	SPEAKING	891.7239	29.8617
Item.4	(Intercept)	4209.5435	64.8810
Participant	VOICE:FORM	3763.3869	61.3464
Participant.1	VOICE:SPEAKING	8769.5587	93.6459
Participant.2	VOICE	1350.0528	36.7431
Participant.3	(Intercept)	54480.4574	233.4105
Residual	NA	88076.3438	296.7766

Table S3.9 Latency analysis - Confidence intervals for fixed effects from the reduced model

	2.5 %	97.5 %
SPEAKING	12.7150	67.2621

FORM	-17.3072	40.6710
VOICE	-59.0345	28.1532
SPEAKING:VOICE	-21.0032	132.8584
FORM:VOICE	-87.2010	61.4679

3.2.3 Reduced model, Windsorised dataset

```
## Onset * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     VOICE | Participant) + (0 + VOICE:SPEAKING | Participant) + (0 +
##     VOICE:FORM | Participant)) + ((1 | Item) + (0 + SPEAKING |
##     Item) + (0 + VOICE | Item) + (0 + SPEAKING:VOICE |
##     Item) + (0 + VOICE:FORM | Item))
```

Table S3.10 Latency analysis on Windsorised dataset - Reduced model fixed effects

	B	SE	t
(Intercept)	962.6201	56.5404	17.0253
SPEAKING	39.3030	13.1003	3.0001
FORM	11.5647	14.3473	0.8061
VOICE	-12.5477	21.8287	-0.5748
SPEAKING:VOICE	54.5202	37.9099	1.4382
FORM:VOICE	-9.3580	36.8983	-0.2536

Table S3.11 Latency analysis on Windsorised dataset - Reduced model random effects

grp	var1	vcov	sdcor
Item	VOICE:FORM	11426.6524	106.8955
Item.1	SPEAKING:VOICE	11846.1111	108.8398
Item.2	VOICE	8379.7915	91.5412
Item.3	SPEAKING	507.7565	22.5335
Item.4	(Intercept)	3918.4327	62.5974
Participant	VOICE:FORM	3107.6663	55.7464
Participant.1	VOICE:SPEAKING	7795.9686	88.2948
Participant.2	VOICE	1312.3875	36.2269
Participant.3	(Intercept)	54685.0079	233.8483
Residual	NA	82875.9203	287.8818

3.2.4 Reduced model, trimmed dataset

```
## Onset * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     VOICE | Participant) + (0 + VOICE:SPEAKING | Participant)) + ((1 |
##     Item) + (0 + VOICE | Item) + (0 + VOICE:FORM |
##     Item) + (0 + VOICE:SPEAKING | Item))
```

Table S3.12 Latency analysis on trimmed dataset - Reduced model fixed effects

	B	SE	t
(Intercept)	950.0976	55.8343	17.0164
SPEAKING	37.6832	11.9278	3.1593
FORM	7.9996	13.7437	0.5821

VOICE	-5.9458	18.9438	-0.3139
SPEAKING:VOICE	43.0710	31.6225	1.3620
FORM:VOICE	3.3525	32.2026	0.1041

Table S3.13 Latency analysis on trimmed dataset - Reduced model random effects

grp	var1	vcov	sdcor
Item	VOICE:SPEAKING	10591.0084	102.9126
Item.1	VOICE:FORM	8862.9185	94.1431
Item.2	VOICE	6207.4454	78.7873
Item.3	(Intercept)	3080.9522	55.5063
Participant	VOICE:SPEAKING	1667.7374	40.8379
Participant.1	VOICE	635.9224	25.2175
Participant.2	(Intercept)	53780.9318	231.9072
Residual	NA	75083.3963	274.0135

3.3 Duration

3.3.1 Full model

```
## Dur * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     SPEAKING | Participant) + (0 + FORM | Participant) + (0 + VOICE |
##     Participant) + (0 + SPEAKING:VOICE | Participant) + (0 + FORM:VOICE |
##     Participant)) + ((1 | Item) + (0 + SPEAKING | Item) +
##     (0 + FORM | Item) + (0 + VOICE | Item) +
##     (0 + SPEAKING:VOICE | Item) + (0 + FORM:VOICE | Item))
```

Table S3.14 Duration analysis - Full model fixed effects

	B	SE	t
(Intercept)	1696.6912	50.7070	33.4607
SPEAKING	-14.4118	16.9115	-0.8522
FORM	-16.2686	19.3111	-0.8424
VOICE	142.8933	19.4812	7.3349
SPEAKING:VOICE	-30.1127	32.1377	-0.9370
FORM:VOICE	51.9500	37.0045	1.4039

Table S3. 15 Duration analysis - Full model random effects

grp	var1	vcov	sdcor
Item	FORM:VOICE	0.0000	0.0004
Item.1	SPEAKING:VOICE	0.0000	0.0001
Item.2	VOICE	2460.5777	49.6042
Item.3	FORM	0.0000	0.0001
Item.4	SPEAKING	866.8321	29.4420
Item.5	(Intercept)	16856.2623	129.8317
Participant	FORM:VOICE	0.0000	0.0000
Participant.1	SPEAKING:VOICE	0.0000	0.0000

Participant.2	VOICE	1260.1171	35.4981
Participant.3	FORM	530.4893	23.0324
Participant.4	SPEAKING	0.0000	0.0000
Participant.5	(Intercept)	35706.8761	188.9626
Residual	NA	138656.8320	372.3665

3.3.2 Reduced model

```
## Dur * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     VOICE | Participant) + (0 + FORM | Participant)) + ((1 | Item) +
##     (0 + SPEAKING | Item) + (0 + VOICE | Item))
```

Table S3.16 Duration analysis - Reduced model fixed effects

	B	SE	t
(Intercept)	1696.6912	50.7070	33.4607
SPEAKING	-14.4118	16.9115	-0.8522
FORM	-16.2686	19.3112	-0.8424
VOICE	142.8933	19.4812	7.3349
SPEAKING:VOICE	-30.1127	32.1377	-0.9370
FORM:VOICE	51.9500	37.0045	1.4039

Table S3.17 Duration analysis - Reduced model random effects

grp	var1	vcov	sdcor
Item	VOICE	2460.5768	49.6042
Item.1	SPEAKING	866.8210	29.4418
Item.2	(Intercept)	16856.2679	129.8317
Participant	FORM	530.4914	23.0324
Participant.1	VOICE	1260.1161	35.4981
Participant.2	(Intercept)	35706.8981	188.9627
Residual	NA	138656.8331	372.3665

Table S3.18 Duration analysis - Confidence intervals for fixed effects from the reduced model

	2.5 %	97.5 %
SPEAKING	-47.5577	18.7341
FORM	-54.1178	21.5805
VOICE	104.7108	181.0759
SPEAKING:VOICE	-93.1014	32.8761
FORM:VOICE	-20.5775	124.4776

3.3.3 Reduced model, Windsorised dataset

```
## Dur * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     VOICE | Participant) + (0 + FORM | Participant)) + ((1 | Item) +
##     (0 + SPEAKING | Item) + (0 + VOICE | Item))
```

Table S3.19 Duration analysis on Windsorised dataset - Reduced model fixed effects

	B	SE	t

(Intercept)	1693.1282	50.7045	33.3921
SPEAKING	-13.6916	16.1836	-0.8460
FORM	-16.5743	18.4018	-0.9007
VOICE	141.2661	19.0245	7.4255
SPEAKING:VOICE	-24.8992	31.2721	-0.7962
FORM:VOICE	48.0084	36.0075	1.3333

Table S3.20 Duration analysis on Windsorised dataset - Reduced model random effects

grp	var1	vcov	sdcor
Item	VOICE	2317.7682	48.1432
Item.1	SPEAKING	541.9674	23.2802
Item.2	(Intercept)	16017.3378	126.5596
Participant	FORM	246.8973	15.7130
Participant.1	VOICE	1245.5023	35.2917
Participant.2	(Intercept)	36232.3006	190.3478
Residual	NA	131311.7766	362.3697

3.3.4 Reduced model, trimmed dataset

```
## Dur * 1000 ~ (SPEAKING + FORM) * VOICE + ((1 | Participant) + (0 +
##     VOICE | Participant)) + ((1 | Item) + (0 + VOICE | Item))
```

Table S3.21 Duration analysis on trimmed dataset - Reduced model fixed effects

	B	SE	t
(Intercept)	1674.7923	50.6118	33.0910
SPEAKING	-12.1819	14.7340	-0.8268
FORM	-11.0043	16.9675	-0.6486
VOICE	136.0306	17.6692	7.6987
SPEAKING:VOICE	-11.1298	29.4591	-0.3778
FORM:VOICE	29.4639	33.9076	0.8689

Table S3.22 Duration analysis on trimmed dataset - Reduced model random effects

grp	var1	vcov	sdcor
Item	VOICE	1707.711	41.3245
Item.1	(Intercept)	13059.028	114.2761
Participant	VOICE	1142.516	33.8011
Participant.1	(Intercept)	37841.982	194.5302
Residual	NA	114699.551	338.6732