Journal of Memory and Language 146 (2026) 104717

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Phonological input processing is reduced during speech planning in turn taking

Mathias Barthel

Leibniz Institute for the German Language, Mannheim, Germany

ARTICLE INFO

Dataset link: https://osf.io/fzha2/?view_only=b7cfc71964d84864b9c5c660fec8cc6f

Keywords:
Turn taking
Dialogue
Phoneme monitoring
Speech planning
Dual task

ABSTRACT

In conversational turn taking, speech planning and language comprehension are known to frequently overlap in time, which has previously been found to lead to less efficient speech production due to parallel processing of linguistic input. Four experiments were conducted to investigate whether comprehension processes are also impaired by concurrent speech planning, or whether language comprehension is prioritised during process overlap in turn taking. In Experiment 1a, participants verbally responded to quiz questions while monitoring the auditorily presented question for a target phoneme. Phoneme detection performance was found to be reduced when the target phoneme was located at a position in the question where participants were already planning their response as compared to when they were not yet planning their response. A comparison with participants' phoneme detection performance in Experiments 1b and 1c, featuring either no or a delayed speech planning task, shows that interlocutors prioritise speech planning in phases of limited processing capacity during turn taking situations. In Experiment 2, participants answered the same quiz questions used in Experiment 1a but had to detect tones instead of phonemes. A comparison with the phoneme detection performance in Experiment 1a suggests that speech planning interferes with phonological input processing in particular. Whether domain-general, non-linguistic auditory detection abilities are hindered by concurrent speech planning, as well as the social and processing-related causes and consequences of the common conversational strategy of planning during comprehension are discussed.

Introduction

One of the central traits of human communicative interaction is turn taking — the rapid exchange of turns at talk between interlocutors, who repeatedly switch roles of being listener and speaker (Sacks, Schegloff, & Jefferson, 1974). While both comprehending and producing speech are cognitively highly complex processes (e.g., Gambi & Pickering, 2017; Hagoort, 2019; Pickering & Garrod, 2004), they are frequently executed simultaneously in conversational contexts (Barthel, Meyer, & Levinson, 2017; Bögels, 2020; Pickering & Garrod, 2013).

Comprehenders are known to process incoming turns incrementally (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Tanenhaus, Magnuson, Dahan, & Chambers, 2000), building anticipations about the upcoming message and its relation to the unfolding discourse (Barthel, Tomasello, & Liu, 2024; Gisladottir, Bögels, & Levinson, 2018; Heilbron, Armeni, Schoffelen, Hagoort, & de Lange, 2022; Huettig, 2015; Ryskin & Nieuwland, 2023). These anticipations, in turn, are frequently used to start planning a relevant next turn already in overlap with the incoming turn (Barthel, 2020; Barthel & Rühlemann, 2025; Bögels, Magyari, & Levinson, 2015; Corps, Crossley, Gambi, & Pickering, 2018). Interlocutors commonly overlap these related processes of comprehension and production presumably for a combination of two reasons: First, planning complex turns takes a considerable amount of time. Even planning to produce a single noun takes about 600 ms (Indefrey, 2011; Strijkers & Costa, 2011), and planning a full sentence can take about one and a half seconds (Griffin & Bock, 2000; Sauppe, 2017; Schnur, Costa, & Caramazza, 2006). Second, interlocutors treat the timing of speaking turns as being meaningful. If, for instance, a request or an offer are not responded to within a short amount of time, the request might be down-scaled or the offer reformulated in a way to display the expectation of a non-preferred response, like a rejection or a denial (Davidson, 1984; Pomeranz, Atkinson, & Heritage, 1984; Pomeranz & Heritage, 2012). Even in less extreme cases, in which the first speaker (who put forth the request or offer) does not re-select for another turn after some time without getting a response, a markedly long delay before the response would still be interpreted to be meaningful. Even if the request is granted or the offer accepted, a long gap before the response can be interpreted to signal low willingness of the second speaker or to be relevant in some

E-mail address: barthel@ids-mannheim.de.

https://doi.org/10.1016/j.jml.2025.104717
Received 1 April 2025; Received in revised form 14 November 2025; Accepted 21 November 2025
Available online 26 November 2025

other meaningful way (Blohm & Barthel, 2025; Henetz, 2017; Roberts, Francis, & Morgan, 2006; Roberts, Margutti, & Takano, 2011). For these reasons, taking turns in conversation regularly proceeds fast and interlocutors are motivated to produce their turns quickly as soon as the floor is open, which puts their language production processes under considerable time pressure during conversation (Levinson & Torreira, 2015; Roberts & Levinson, 2017).

Speech planning has been shown to be cognitively demanding in unscripted conversational contexts (Barthel & Rühlemann, 2025; Rühlemann & Barthel, 2024). Particularly when executed in overlap with the incoming turn, speech planning has been found to both lead to increased processing load and to take longer than when executed in silence between turns (Barthel & Sauppe, 2019; Barthel, Sauppe, Levinson, & Meyer, 2016). The most probable cause for the reduced efficiency of speech planning processes during speech comprehension has been proposed to be interference of the two processes (Fargier & Laganaro, 2016; He, Meyer, & Brehm, 2021; Levelt, Roelofs, & Meyer, 1999; Roelofs, 2021). One main strand of evidence for interference between comprehension and production comes from picture-word-interference experiments (e.g., Bürki, Elbuy, Madec, & Vasishth, 2020; Schriefers, Meyer, & Levelt, 1990; Wilshire, Singh, & Tattersall, 2016), where the presentation of a word (visual or auditory) can decrease participants' speed and accuracy in a picture-naming task. The candidate explanation for the observed interference effects is that speech production and comprehension compete for the same cognitive resources and are run on partly overlapping neural infrastructure and mental representations (Hagoort & Indefrey, 2014; Menenti, Gierhan, Segaert, & Hagoort, 2011; Segaert, Menenti, Weber, Petersson, & Hagoort, 2012; Silbert, Honey, Simony, Poeppel, & Hasson, 2014). This overlap in processes is particularly relevant in turn taking situations, where both comprehension and production are under time pressure and might compete for processing capacities within the 'crunch zone' in the vicinity of turn transitions, where one speaker's turn comes to an end and the next speaker's turn should start within a short gap, often as short as 200 ms (Levinson & Torreira, 2015; Stivers et al., 2009).

Given that speech planning in overlap with the incoming turn is a common strategy in conversational situations, and given that the processes of planning and comprehension regularly interfere with each other, two hypotheses are conceivable about the processes of comprehension in overlap with speech planning: Either comprehension is prioritised over planning at times of resource competition, which we will call the *comprehension-prioritised hypothesis*, or planning is prioritised over comprehension, which we will call the *planning-prioritised hypothesis*.

Both hypotheses can be argued for a-priori from a goal-oriented processing point of view: Favouring the comprehension-prioritised hypothesis, the primary task in dialogue situations may reasonably be argued to be accurate comprehension, since perceived auditory information can only be held in working memory for a very limited amount of time (e.g., Baddeley, 2003; Christiansen & Chater, 2016). Because the rate and timing of information contained in the incoming turn is not under the comprehender's control, this information needs to be processed as soon as it is encountered and translated into more abstract representations that can be kept in memory more permanently and used to update the comprehender's discourse model of the ongoing conversation. As a consequence, comprehension of the incoming turn cannot be delayed and should thus take priority over speech planning, lest the input remains unprocessed until its phonological traces fade, which would jeopardise conversational success. Hence, in cases of limited processing capacity, speech planning processes would need to be deferred in favour of comprehension processes.

Alternatively, favouring the planning-prioritised hypothesis, the primary task in dialogue situations may reasonably be argued to be speech planning, since early planning serves the goal of producing a turn in tight coordination with the point of completion of the incoming turn. Producing a turn at talk without a marked delay is frequently essential to communicate the intended message and to avoid being misinterpreted (Blohm & Barthel, 2025; Kendrick & Torreira, 2014; Roberts & Francis, 2013). In moments of high processing load during dialogue, speech planning would therefore have to be prioritised over comprehension. In such cases, the planning of a relevant next turn needs to be partly based on the anticipated message of the incoming turn (Levinson & Torreira, 2015). While prioritising planning, next speakers would have to rely mainly on previously generated anticipations of the input would and thus risk comprehension accuracy in moments when the input is monitored only shallowly for the sake of effective speech planning (e.g., Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007).

While the planning-prioritised hypothesis and the comprehension-prioritised hypothesis represent extreme positions on a continuum and interlocutors might adjust their processing strategies depending on the conversational context, these opposing hypotheses make distinct testable predictions about interlocutors' language comprehension performance in dialogue situations that feature parallel comprehension and planning: According to the comprehension-prioritised hypothesis, planning should only be pursued in overlap with the incoming turn if resources are permitting, so that comprehension should not suffer from parallel planning. According to the planning-prioritised hypothesis, in contrast, comprehension accuracy should be reduced in situations of resource scarcity, so that a timely production of the next turn is not compromised by input comprehension processes. In either case, one process would be momentarily prioritised at the cost of another in moments of high processing load, since the two processes are competing for processing resources and potentially interfere with one another.

The sources of these observed interferences can be located on (a combination of) any linguistic processing level(s), ranging from phonetic analysis/assembly to semantic and discourse representations. Hence, the two opposing hypotheses, prioritising either planning or comprehension, can be tested on different levels of linguistic processing. Testing a verbal question response task with quiz questions that contained semantic illusions, Barthel (2021) showed that comprehension accuracy is reduced on the semantic level during concurrent speech planning. Participants in that study accepted semantic illusions like "What animal ate Little Red Riding Hood | when she visited her aunt?"[1] (early planning condition) more often than illusions like "When Little Red Riding Hood visited her aunt, what animal ate her |?" (late planning condition).[2] In the early planning condition, participants already engaged in planning their response at the point in time when they encountered the illusion, making the illusion go undetected significantly more often than in the late-planning condition, where participants did not yet engage in response planning when they encountered the illusion. These results suggest that semantic input processing during parallel speech planning can be more shallow than in the absence of speech planning, supporting the planning-prioritised hypothesis.

Investigating related research questions using EEG in a go/no-go picture naming task with sentence primes, Hustá, Nieuwland, and Meyer (2023) tested N400 effects, which are related to semantic input processing, when participants listened to a sentence containing an expected versus unexpected final word and additionally overtly named a picture that was presented either while hearing the final word or two seconds later. Hustá et al. (2023) found the N400 to be more negative in unexpected words than in expected words. In line with the results by Barthel (2021), the authors found this N400 effect to be attenuated when participants were planning their response in overlap with being presented with the sentence final word, indicating that

[1] The '|' symbol marks the point in the question when planning the answer can begin.

[2] Participants were subsequently checked to know that Little Red Riding Hood actually visited her grandma.

semantic processing of the target word was reduced during concurrent planning. Interestingly, in trials in which the picture was presented in overlap with the target word, the authors also found evidence for an increased N400 component in unexpected as compared to expected words when subjects did not have to overtly name the picture. When the picture showed an object from the no-go category (either fruit or vegetable, depending on the experimental list), subjects did not have to go through with naming the picture but instead had to press a button to skip the naming. These no-go trials showed a comparable, while reduced, effect of predictability of the sentence final word, suggesting that not the processes of phonological or articulatory planning but rather processes of conceptual preparation are the main source location of the observable interference of speech planning with comprehension. Nonetheless, these results do not exclude the possibility of additional interference on lower processing levels, since the picture categorisation task will probably have led to phonological processing of the picture name (Bles & Jansma, 2008; Meyer & Damian, 2007; Navarrete & Costa, 2005). Moreover, the sentence comprehension and picture naming tasks were unrelated to each other, calling for an investigation using a more ecologically valid test that contains a speech planning task which is designed to be conditionally relevant to the just comprehended input.

While language comprehension has been found to be more shallow on the level of semantic processing during intervals of parallel speech planning, it is unknown whether these effects can already be observed on lower levels of abstraction. Since the speech planning process in overlap with comprehension of the incoming turn is pursued at least until the phonological level (Barthel & Levinson, 2020), it is conceivable that comprehension accuracy in overlap with speech planning suffers already at the rather early stage of phonological processing. To pursue this question, the present study tests participants' phoneme detection accuracy in question-response sequences, comparing participants' phoneme detection performance during intervals of speech comprehension in isolation on the one hand with intervals of comprehension in overlap with response planning on the other hand. If phonological input processing deteriorates during concurrent speech planning, phoneme monitoring performance should be observed to be worse than in situations without concurrent planning. If, however, input comprehension is prioritised during phases of processing resource scarcity, phoneme detection performance should not significantly differ between situations of parallel planning on the one hand and comprehension in absence of planning on the other hand.

To evaluate these hypotheses, participants' phoneme detection performance was tested in three Experiments 1a, 1b, and 1c. In Experiment 1a, participants performed a dual task in which they heard a question that they had to verbally respond to and at the same time monitor for a target phoneme. Firstly, if next speakers start planning their responses in overlap with the incoming turn, if possible, participants' verbal response latencies should be shorter in early planning questions than in late planning questions, as has been shown before (e.g. Barthel & Levinson, 2020; Bögels, 2020). Secondly, if planning is prioritised during phases of concurrent planning and comprehension, phoneme detection performance should be worse and probably also slower when the target phoneme is presented while participants are concurrently planning their response as compared to while they are not concurrently planning. If, however, comprehension is prioritised in these situations, phoneme detection performance should be unimpaired by concurrent planning. In Experiment 1b, participants performed the same phoneme monitoring task but without having to verbally respond to the questions. Since participants do not engage in verbal response planning in this Experiment, the type of question should not affect phoneme detection performance or speed (unless there is an inherent difference in difficulty to detect a target phoneme at different positions in the question). In Experiment 1c, participants were given a speech planning task that they had to do subsequent to the phoneme monitoring task. While participants are engaged in speech planning in this Experiment, phases of speech planning do never overlap with the presentation of the target phoneme, so that the type of question should also not affect phoneme detection performance or speed (again, unless phonemes are generally more easily detected in one sentence position than in another, e.g., because the probability of encountering the target phoneme continuously rises as the question unfolds).

To test whether not phonological processing in particular but rather auditory processing in general was affected by parallel planning, participants in Experiment 2 were given the same question-answering task as participants in Experiment 1a, but with the additional task to detect tones instead of phonemes while hearing the question. Firstly, participants' verbal response latencies should pattern like in Experiment 1a, indicating that early planning questions indeed led to planning in overlap with the incoming questions. Secondly, if this tone detection Experiment shows an effect of question type that matches the predicted effect of Experiment 1a (worse detection performance when planning during the presentation of the target phoneme), than this would be taken as evidence that auditory processing in general is affected by parallel speech planning. If, however, this Experiment shows no effects of question type, as predicted for Experiments 1b and 1c, the respective effects in Experiment 1a would be taken as evidence that phonological processing in particular is affected by parallel speech planning.

Methods

Participants

Sixty participants were tested in each of the four Experiments. Each participant took part in only one of the Experiments. Six participants have been excluded from analyses of Experiment 1a and two have been excluded from Experiment 1c (see Section 'Data Coding and Analysis'). All participants were healthy German native speakers between 18 and 49 years of age who were recruited online via Prolific, gave their written consent to participate in the data collection, and received monetary compensation for their participation.

Regarding one experiment in isolation (and ignoring any increase in measurement reliability due to repeated measures), to attest the effect of Planning (early/late, see Section 'Materials and Design') in a given experiment would call for 52 participants (Brysbaert, 2019). Further, sample size recommendations by Brysbaert (2019) to obtain sufficient power to detect an interaction effect of a 2-level between-subjects factor and a 2-level within subjects factor for repeated-measures tests containing a sufficiently large number of items (as we do here, with 60 items tested per participant, see Section 'Materials and Design'; see also Brysbaert and Stevens (2018)) range between 90 and 130 participants in total. This would call for 45 to 65 participants per experiment to reach a power of .8 to detect an interaction of Planning × Experiment at an alpha level of .05. Testing 60 participants per Experiment can thus be expected to detect an interaction effect with an effect size of $f = .26$ with a power of .8 at an alpha level of .05 and a simple effect of the within-subjects variable with an effect size of $d = .37$ with the same power at the same alpha level, as calculated using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007).

Materials and design

In total, 100 German questions were created in two versions that were used to make the answer to the question known either already during the question (early-planning condition; e.g.: "Wenn man Rot und Gelb mischt, | welche Farbe erhält man dann?" (When you mix red and yellow, | what colour do you get?); the "|" symbol representing the point in time when the answer to the question can be known) or only at the very end of the question (late-planning condition; e.g.: "Welche Farbe erhält man, wenn man Rot und Gelb mischt | ?" (What colour do you get when you mix red and yellow | ?)). Quiz questions required short responses, mainly single word answers (e.g., "Berlin") or proper nouns (e.g., "Neil Armstrong"). The different versions of each

Table 1
Example stimuli for one of the four tested target phonemes (Experiments 1a, 1b, and 1c) or target tones (Experiment 2) in the early and late planning condition.

Target phoneme or tone	Question type	Question
[v]/680 Hz	early planning	Diese Tiere, die auch Briefe ausliefern können, | sind ein **w**ichtiges Symbol in der Kunst. (*These animals, which can also deliver mail, are an important symbol in art.*)
[v]/680 Hz	late planning	Diese Tiere, die ein **w**ichtiges Symbol in der Kunst sind, können auch Briefe ausliefern |. (*These animals, which are an important symbol in art, can also deliver mail.*)

The bold, underlined characters represent the target phoneme. In Experiment 2, the target tone was played for 250 ms, starting at the onset of the phoneme that is the target in the phoneme detection task of Experiments 1a, b, and c. The | symbol shows the point at which the answer to the question can be known.

of the questions required the same answers and were worded almost identically, only differing in the order of their components (see Table 1).[3] Sixty critical questions contained one word starting with one of the target phonemes [f], [b], [g], or [v] (15 questions for each of the four target phonemes; each question only featured a single token of the target phoneme). In the early-planning condition, the answer to the question could be planned early on during the question. Questions in this condition contained the word beginning with the target phoneme towards the end, i.e., after the answer to the question could be known. Questions in the late-planning condition were designed so that the answer could only be known at the end of the question. Questions in this condition contained the word beginning with the target phoneme closer to the question beginning, where the answer to the question could not yet be known (cf. the underlined 'F' in the example above). Thereby, participants could start to plan their response earlier than when being presented with the target phoneme in the early-planning condition, but later than when being presented with the target phoneme in the late-planning condition. For Experiment 2, the critical questions were additionally overlaid with 250 ms long sine tones, generated with Audacity (Audacity Team, 2024), that were played at the position of the target phonemes. The fifteen questions with the target phoneme [b] contained a tone of 440 Hz at the position of the [b]; the questions with the target phoneme [f] contained tones of 520 Hz; the questions with the target phoneme [g] contained tones of 600 Hz; and the questions with the target phoneme [v] contained tones of 680 Hz. Forty questions were designed as filler questions and did not contain a target phoneme (and therefore also no tone in Experiment 2). All questions were recorded by a female native speaker of German and had a mean length of 5393 ms (SD = 1491 ms; $M_{critical/early}$ = 5248 ms, SD = 1222 ms; $M_{critical/late}$ = 5768 ms, SD = 1514 ms). Two experimental lists were created, with half of the questions appearing in the early-planning condition and half appearing in the late-planning condition in each list, so that each question was played to any participant in only one of the conditions. The order of items was randomised within lists and participants were randomly assigned to one of the two lists.

In Experiment 1c (phoneme detection with subsequent picture naming task), one-hundred pictures of easy-to-name concrete objects were selected from the MultiPic database (Duñabeitia, Crepaldi, Meyer, New, Pliatsikas, Smolka, & Brysbaert, 2018) and assigned to the items so as to be unrelated to the questions or their answers.

Procedure

All four Experiments were tested online via PCIbex (Zehr & Schwarz, 2018) using participants' own computers, which has been shown to lead to valid and reliable measures and effects given sufficiently large sample sizes (Anwyl-Irvine, Dalmaijer, Hodges, & Evershed, 2021; Fairs & Strijkers, 2021; Rodd, 2024; Vogt, Hauber, Kuhlen, & Abdel Rahman, 2021).

Experiment 1a: Dual task phoneme monitoring and response planning

In Experiment 1a, participants were auditorily presented with quiz questions and had to give their answers verbally as fast and accurately as possible (Fig. 1, panel A). Moreover, they were instructed to monitor the question for one out of four target phonemes at a time, which occurred in 60% of the questions (critical items), and to press the space bar as fast as possible when they heard the target phoneme. The four target phonemes ([f], [b], [g], or [v]) were monitored in four blocks of 25 trials each, one block per phoneme, with the order of blocks being randomised across participants. Each block contained 15 critical items and 10 filler items, which did not contain the respective target phoneme in the question.

The experiment began with a sound check to ensure that the voice of the participants could be recorded. Participants were then introduced to the verbal quiz task and answered four practice questions as fast as possible to familiarise themselves with the task. Next, participants received instructions about the manual phoneme detection task. They were then presented with a recording of the phoneme they had to focus on in the first block of 25 questions, followed by a recording containing a series of twenty single words, ten of which started with the target phoneme of the first block while the other ten did not contain the target phoneme. Participants were instructed to press the space bar each time they recognised the target phoneme in the presented list of words, with each button press receiving visual feedback in the shape of a space bar presented on the screen for 500 ms. This familiarisation with the target phoneme together with the short practice phase preceded each of the four blocks, instructing participants to monitor the questions of the respective block for one particular target phoneme ([f], [b], [g], or [v]).

Each trial began with a fixation cross presented for 1000 ms in the centre of the screen, followed by the auditorily presented question. In critical trials, which contained a target phoneme, participants had to press the space bar as soon as they detected the phoneme. In all trials, critical as well as filler, participants answered the question verbally as fast as possible, with their responses being recorded. After giving their answer, participants could proceed to the next trial by clicking a button on the screen with their mouse.

Experiment 1b: Single task phoneme monitoring (no verbal response)

The design, materials, and procedure of the single control task in Experiment 1b were identical to Experiment 1a, with the difference that participants in Experiment 1b were instructed that they did not have to give any verbal responses to the quiz questions and only had to perform the phoneme detection task (Fig. 1, panel B). Also, participants did not go through a practice phase for the verbal response task.

[3] Note that in Experiments 1b and 1c, participants were not required to answer the questions. For the sake of comparison, the two question types will still be referred to as 'early planning questions' and 'late planning questions' in these conditions, even though no response planning is required.

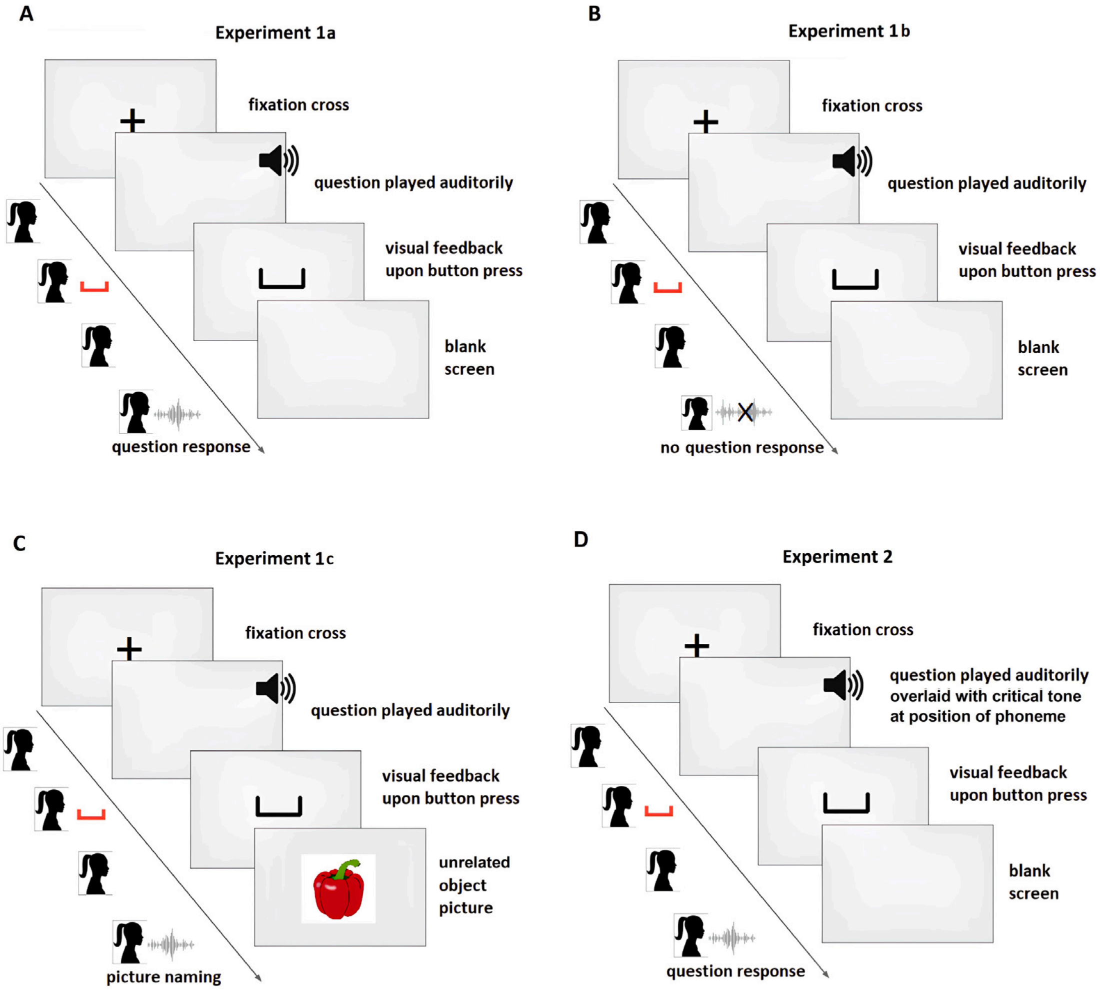
A
Experiment 1a
fixation cross
question played auditorily
visual feedback upon button press
blank screen
question response
B
Experiment 1b
fixation cross
question played auditorily
visual feedback upon button press
blank screen
no question response
C
Experiment 1c
fixation cross
question played auditorily
visual feedback upon button press
unrelated object picture
picture naming
D
Experiment 2
fixation cross
question played auditorily overlaid with critical tone at position of phoneme
visual feedback upon button press
blank screen
question response

Fig. 1. Trial structures of the four Experiments. Experiment 1a (panel A): Phoneme monitoring plus verbal question answering; Experiment 1b (panel B): Phoneme monitoring without speech planning; Experiment 1c (panel C): Phoneme monitoring with subsequent unrelated picture naming.; Experiment 2 (panel D): Tone monitoring plus question answering.

Experiment 1c: Dual-task phoneme monitoring and subsequent picture naming

The design, materials, and procedure of Experiment 1c were identical to Experiment 1b, with the difference that participants in Experiment 1c were instructed that, in addition to the phoneme detection task, they had to name a picture that would appear after each question as fast as possible (Fig. 1, panel C). The picture was presented 500 ms after question offset and disappeared when subjects clicked the proceed button with their mouse to continue with the next trial. The objects presented in the pictures were unrelated to the respective questions or their answers, so that the speech planning task would not be related to the listening task.

Experiment 2: Dual task tone detection and response planning

The design, materials, and procedure of Experiment 2 were identical to Experiment 1a, with the difference that participants in Experiment 2 were instructed that they had to do a tone detection task instead of a phoneme detection task (Fig. 1, panel D). Also, the practice and familiarisation parts of the experiment used tones in place of phonemes.

Data coding and analysis

Participants' verbal response latencies in Exp. 1a and 2 (answers to questions) and Exp. 1c (picture naming) were measured manually in Audacity (Audacity Team, 2024, v.3.6.1) from question offset (Exp. 1a and 2) or picture onset (Exp. 1c) to response onset. Participants' button press accuracy and latencies (from target phoneme/tone onset) were logged during the experiment. All data were analysed in R (R. Core Team, 2023) using the packages lme4 (Bates, Mächler, Bolker, & Walker, 2015) to build (generalised) linear mixed-effects regression models, lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) to assess the statistical significance of single predictors, and emmeans (Lenth, 2022) to conduct post-hoc tests to assess the significance of simple effects by comparing estimated marginal means of factor levels based on F-tests with Kenward-Roger approximations of degrees of freedom (Fox & Weisberg, 2011; Halekoh & Hojsgaard, 2014; Kenward & Roger, 1997; Searle, Speed, & Milliken, 1980), correcting significance levels for multiple testing using multivariate t adjustment. Three subjects in Exp. 1a were discarded from the analysis, since their verbal responses were either not given or not recorded. Three more subjects in Exp. 1a were excluded from analyses, because they did not perform the phoneme detection task. Two subjects in Exp. 1c were discarded since their audio was not recorded during the testing session. Two items were discarded from the analyses of Exp. 1a and 1b due to technical errors. 719 trials in Exp. 1a, 166 trials in Exp. 1c, and 762 trials in Exp. 2 containing wrong answers or ‘I don't know’ answers to the quiz questions or the picture naming task were discarded from the analyses. 70 trials from Exp. 1a, 193 trials from Exp. 1b, 239 trials

Table 2
Numbers and percentages of undetected phonemes/tones by Experiment for early vs. late planning trials, and overall.

	n missed/total$_{early}$	n missed/total$_{late}$	n missed/total$_{overall}$
Experiment 1a	378/1150 (32.8%)	153/1193 (12.8%)	531/2343 (22.6%)
Experiment 1b	201/1623 (12.3%)	185/1694 (10.9%)	386/3317 (11.6%)
Experiment 1c	231/1487 (15.5%)	216/1588 (13.6%)	447/3075 (14.5%)
Experiment 2	23/1434 (0.8%)	15/1398 (0.5%)	38/2832 (1.3%)

from Exp. 1c, and 5 trials from Exp. 2 with erroneous button presses that were given before the target phoneme/tone was presented were discarded. In total, 72.3% of the critical trials originally presented to the 54 valid participants of Exp. 1a, 92.1% of the 60 participants of Exp. 1b, 88.3% of the 58 participants of Exp. 1c, and 78.6% of the 60 participants of Exp. 2 were retained for analyses. For the results presented here, filler trials were not analysed.

Results

The main question of the study is whether language comprehension on the phonological level is impaired during concurrent speech planning. To target this question, phoneme and tone detection performance in the early planning condition versus the late planning condition will be analysed and compared between the present experiments. To test whether participants indeed planned their verbal responses early in the early planning conditions, verbal response latencies are analysed at the end of this section.

Detection tasks

Phoneme and tone detection accuracy

To test whether phoneme and/or tone detection rates were worse during concurrent speech planning, detection rates in the early versus late planning condition were compared between the present experiments. In the four experiments, different proportions of target phonemes/tones went undetected (see Table 2 and Fig. 2). A generalised linear mixed-effects model was built to estimate the proportion of trials with successful phoneme/tone detection[4] using Planning and Experiment as well as their interaction as fixed predictors and Phoneme/Tone as an additional control predictor. The model contained random intercepts by subject and by item and random slopes of Planning by subject and by item as well as random slopes of Experiment by item.[5] Planning and Experiment were dummy coded, with early planning and the single-task Exp. 1b as reference levels. Phoneme/Tone was deviation coded.

No significant difference was found in subjects' phoneme detection performance in early vs. late planning questions when participants did not have to give a verbal response (Exp. 1b) (Table 3), indicating that the position of the target phoneme in the question did not affect participants' detection accuracy in itself. When participants had to verbally answer the question (Exp. 1a), significantly more target phonemes were not detected compared to Exp. 1b in the early planning condition ($p < .001$), indicating that phoneme detection accuracy is reduced during concurrent speech planning. In contrast, when participants had to name a picture after the question had been presented (Exp. 1c), detection rates in early planning trials were not found to be significantly different from Exp. 1b, indicating that the difference in detection rates between Exp. 1a and 1b was not merely due to the fact that Exp. 1a contained a speech planning task on top of the phoneme detection task. A significant interaction of Planning and Experiment was attested comparing detection rates when the questions had to be answered (Exp. 1a) vs. not answered (Exp. 1b) ($p < .001$). This interaction was not significant in the comparison of the picture naming task (Exp. 1c) vs. no verbal response task (Exp. 1b), nor in the comparison of the no verbal response task (Exp. 1b) vs. tone detection plus question answering task (Exp. 2). Post-hoc tests on single effects revealed that the significant interaction effect was due to a significant effect of Planning in Exp. 1a, with significantly more undetected target phonemes in early planning trials than in late planning trials ($\beta = -1.399$; $SE = 0.169$; $z = -8.274$; $p < .001$), indicating that phoneme detection accuracy was lower during concurrent speech planning than without concurrent speech planning. This effect was not significant Exp. 1b or 1c, which did not feature speech planning during the target phoneme (Exp. 1b: $\beta = -0.191$; $SE = 0.182$; $z = -1.047$; $p = .295$; Exp. 1c: $\beta = -0.251$; $SE = 0.177$; $z = -1.413$; $p = .157$), supporting that concurrent speech planning is indeed the driving factor of the effect of Planning in Exp. 1a. The effect of Planning was also not significant in Exp. 2, which required participants to detect tones instead of phonemes while answering the question ($\beta = -0.640$; $SE = 0.419$; $z = -1.528$; $p = .126$), suggesting that speech planning affects phoneme detection accuracy but not tone detection accuracy.

Phoneme and tone detection latencies

To test whether phoneme/tone detection latencies match the results pattern of detection accuracies and/or whether speech planning generally affects manual response latencies in the present experiments, detection latencies in the early versus late planning condition were compared between experiments. Analysing the trials in which the target phoneme/tone was correctly detected (see caption of Fig. 3), a linear mixed-effects regression model was built to estimate detection latencies in these trials, with Planning and Experiment as well as their interaction as fixed predictors and Phoneme/Tone as an additional control predictor.[6] The model contained random intercepts by subject and by item and random slopes of Planning by subject and by item (Fig. 3).[7] Planning and Experiment were dummy coded, with early planning and the single-task Experiment 1b as reference levels. Phoneme/Tone was deviation coded.

Detection latencies in Exp. 1b were found to be significantly slower in late planning trials as compared to early planning trials (Table 4; $\beta = 61$ ms; $SE = 25$ ms; $p = .016$), indicating that target phonemes are generally detected faster when they appear later in the question. Similarly, detection latencies in Exp. 1c (picture naming task) were also significantly slower in late planning trials compared to early planning trials ($\beta = 52$ ms; $SE = 26$ ms; $p = .044$, as attested by post-hoc tests), corroborating the result of Exp. 1b. Contrarily, in Exp. 1a, where participants verbally responded to the question, detection latencies were found to be significantly faster in late planning trials as compared to early planning trials ($\beta = -102$ ms; $SE = 30$ ms; $p < .001$, as attested by post-hoc tests), indicating that button press latencies were greater during concurrent planning than while not concurrently planning. This led to a significant interaction of Planning and Experiment when comparing Exp. 1a and 1b ($p < .001$). Like in Exp. 1a, detection latencies in Exp. 2 were found to be significantly faster in late planning trials as compared to early planning trials ($\beta = -114$ ms; $SE = 25$ ms; $p < .001$, as attested by post-hoc tests), also indicating that button presses were slower while planning than while not planning, and also leading to a significant interaction of Planning and Experiment when comparing Exp. 2 and 1b ($p < .001$).

[4] Since in Exp. 2 the critical tones were played at the moments of the phonemes that were critical in Exp. 1a, 1b, and 1c, the critical tones and the respective critical phonemes were treated as the same for the sake of this analysis, making up a single factor with four levels ([b]/440 Hz, [f]/520 Hz, [g]/600 Hz, [v]/680 Hz).

[5] No interaction of random slopes of Experiment with Planning by item was modelled, to avoid singularity of model fit.

[6] Phoneme/Tone was treated as a single factor with four levels, see footnote a.

[7] No random effects of Experiment by item were modelled to avoid singularity of model fit.

Table 3
Model output of mixed-effects model on phoneme/tone detection rates.

	β	*SE*	*z*	*p*	
Intercept	2.497	0.204	12.231		
Planning_late	0.190	0.182	1.047	0.295	
Experiment_1a	−1.685	0.237	−7.091	<0.001	***
Experiment_1c	−0.281	0.237	−1.186	0.236	
Experiment_2	2.428	0.379	6.408	<0.001	***
Phoneme_1	0.102	0.145	0.700	0.484	
Phoneme_2	0.059	0.138	0.430	0.667	
Phoneme_3	−0.167	0.139	−1.202	0.229	
Planning_late:Experiment_1a	1.208	0.221	5.462	<0.001	***
Planning_late:Experiment_1c	0.059	0.220	0.273	0.785	
Planning_late:Experiment_2	0.449	0.430	1.044	0.296	

Formula = hit ~ planning * experiment + phoneme + (1 + planning | subjectID) + (1 + planning + experiment | itemID); Family = binomial(link=logit). Experiment 1a = dual task of phoneme detection plus question answering; Experiment 1c = dual task of phoneme detection plus picture naming; Experiment 2 = dual task of tone detection plus question answering.

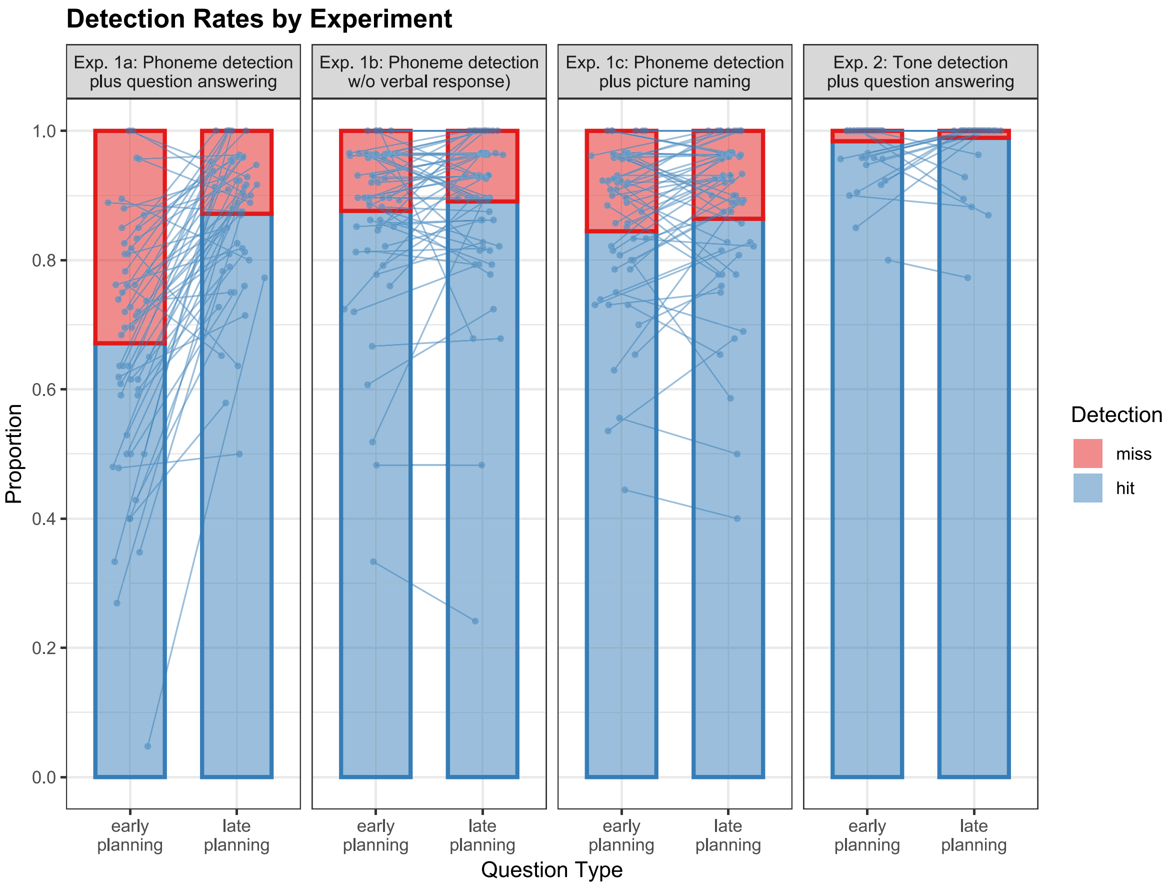

Fig. 2. Phoneme detection accuracy by Experiment. Dots indicate mean detection rates by subject. Lines between dots connect subjects' detection rates across conditions. $n_{Exp._{1a}/early} = 1150$; $n_{Exp._{1a}/late} = 1193$; $n_{Exp._{1b}/early} = 1623$; $n_{Exp._{1b}/late} = 1694$; $n_{Exp._{1c}/early} = 1487$; $n_{Exp._{1c}/late} = 1558$; $n_{Exp._{2}/early} = 1434$; $n_{Exp._{2}/late} = 1398$.

To additionally test for a relation of phoneme/tone detection latencies with verbal response latencies in the Experiments that required a verbal response to the question (Exp. 1a and 2), an additional model was built, estimating detection latencies with verbal response latencies as an additional predictor next to Planning and Experiment and their interaction (see Table A.3 in the Appendix). Phoneme/Tone was added as a control variable.[8] The model contained random intercepts by subject and by item and random slopes of Planning by subject and by item.[9] Planning and Experiment were dummy coded, with early planning and Exp. 1a (phoneme detection) as reference levels. Phoneme/tone was deviation coded. This model's output shows that verbal response latencies and phoneme detection latencies are linearly correlated in Exp. 1a, with early planning trials with longer verbal response latencies showing later button presses ($p < .001$). While going in the same direction, this effect is smaller in late planning trials, as shown by a significant interaction of verbal RT and Planning (p =

[8] Phoneme vs. tone was treated as a single factor with four levels, see footnote a.

[9] No random effects of Experiment by item were modelled to avoid singularity of model fit.

Table 4
Model output of mixed-effects model on phoneme detection latencies.

	β	*SE*	*t*	*p*	
Intercept	935.542	40.044	23.363		
Planning_late	61.132	25.128	2.433	0.016	*
Experiment_1a	177.130	45.361	3.905	<0.001	***
Experiment_1c	−165.338	41.248	−4.008	<0.001	***
Experiment_2	−196.975	41.336	−4.765	<0.001	***
Phoneme_1	−65.077	41.742	−1.559	0.124	
Phoneme_2	71.417	41.739	1.711	0.092	
Phoneme_3	−31.627	41.915	−0.755	0.453	
Planning_late:Experiment_1a	−163.948	38.429	−4.266	<0.001	***
Planning_late:Experiment_1c	−8.713	34.848	−0.250	0.802	
Planning_late:Experiment_2	−175.876	34.384	−5.115	<0.001	***

Formula = detectionLatency ~ planning * experiment + phoneme + (1 + planning | subjectID) + (1 + planning | itemID); Family = gaussian(link=identity). Experiment 1a = dual task of phoneme detection plus question answering; Experiment 1c = dual task of phoneme detection plus picture naming; Experiment 2 = dual task of tone detection plus question answering.

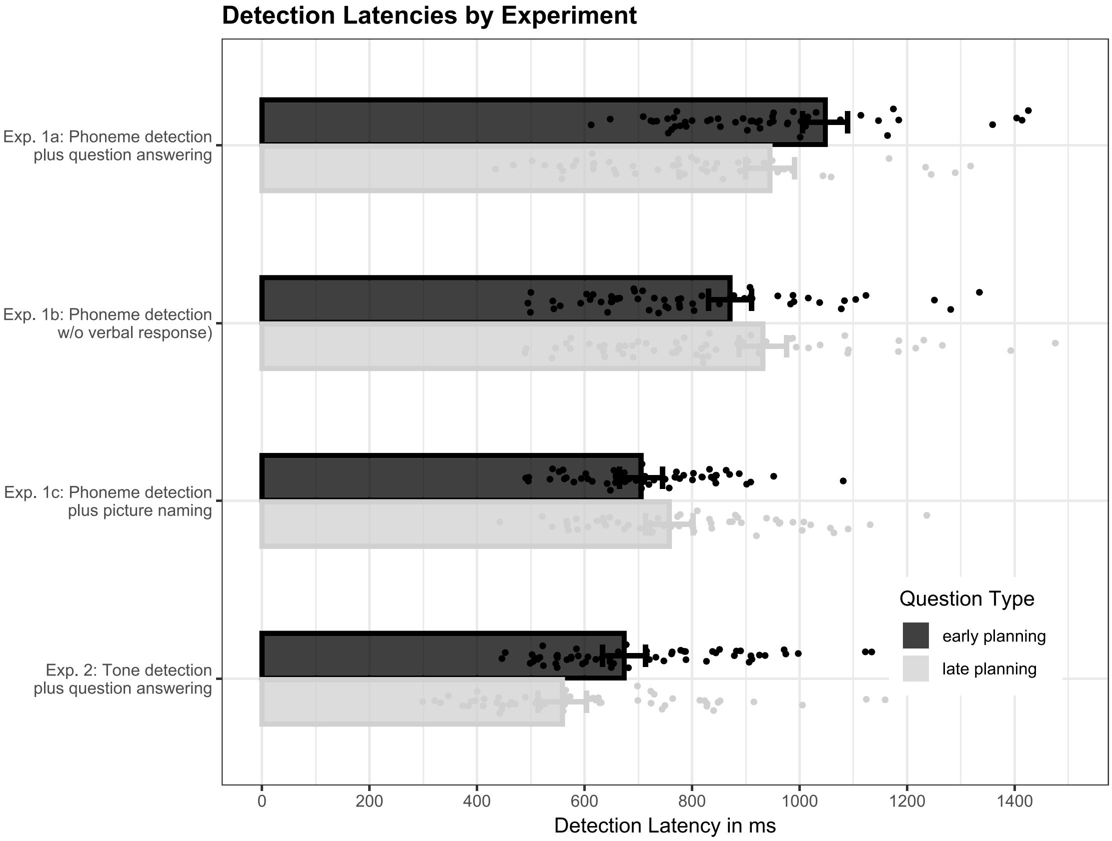

Fig. 3. Model predictions of phoneme detection latencies (see Table 4 for model details). Error bars represent one standard deviation from the mean. $n_{Exp._{1a}/early}$ = 772; $n_{Exp._{1a}/late}$ = 1040.; $n_{Exp._{1b}/early}$ = 1422; $n_{Exp._{1b}/late}$ = 1509; $n_{Exp._{1c}/early}$ = 1256; $n_{Exp._{1c}/late}$ = 1372; $n_{Exp._{2}/early}$ = 1411; $n_{Exp._{2}/late}$ = 1383. Overlaid dots represent predicted means by participant and condition.

.014). In Exp. 2 (tone detection), verbal response latencies and tone detection latencies are not found to be significantly correlated, neither in the early planning condition, nor in the late planning condition, as attested by post-tests, leading to a significant interaction of verbal RT and Experiment (p < .001).

Verbal response tasks

Verbal response latencies in Experiment 1a (phoneme detection plus question answering)

In order to test whether participants indeed planned their responses in overlap with the incoming question in the early planning condition, matching the logic of the study's manipulation, latencies of verbal responses to critical questions in Exp. 1a (N = 2832) were modelled in a linear mixed-effects regression with early versus late Planning as a fixed predictor and as random slopes by subject and by item in addition to random intercepts by subject and by item (Fig. 4, left panels). Detection of the target phoneme (detected vs. missed) was included as a control predictor, together with its interaction with Planning. Both Planning and Detection were dummy coded, with early planning and phoneme detected as reference levels. Subjects were found to respond significantly slower to late planning questions than to early planning questions (Table 5), indicating that participants indeed started to plan their answer to the question during the incoming question in the early planning condition. As shown by a significant interaction of Planning with Detection, this effect was even stronger in trials in which the target phoneme was not detected as compared to trials with successful phoneme detection (p = .027). As attested by post-hoc-tests,

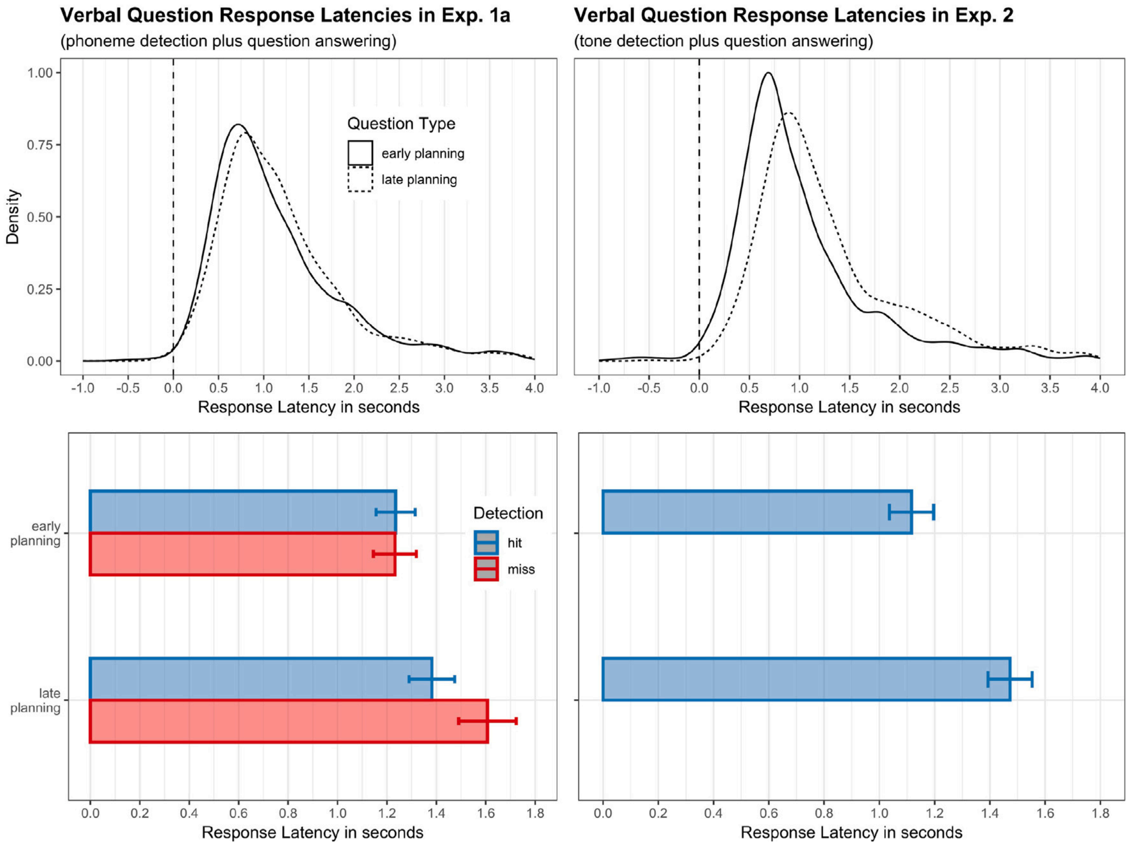

Fig. 4. Top panels: Density distributions of verbal response latencies in Experiments 1a (left) and 2 (right). The vertical dashed lines indicate question offset. Bottom panels: Model predictions of verbal response latencies (see Tables 5 and 6 for model details). Error bars represent one standard deviation from the mean. $n_{Exp.\ 1a\_early/hit}$ = 772; $n_{Exp.\ 1a\_early/miss}$ = 378; $n_{Exp.\ 1a\_late/hit}$ = 1040; $n_{Exp.\ 1a\_late/miss}$ = 153; $n_{Exp.\ 2\_early/hit}$ = 1411; $n_{Exp.\ 2\_late/hit}$ = 1383. Modelled estimates for trials with undetected tones in Experiment 2 are not displayed here since they are unreliable due to the low number of cases (n = 38).

Detection was no significant predictor of verbal response times in early planning trials (β = −3 ms; SE = 62 ms; t = −0.052; p = .958), but was a significant predictor in late planning trials (β = 224 ms; SE = 83 ms; t = 2.692; p = .007).[10] An additional model predicting verbal response latencies in trials in which the target phoneme was detected (n = 1812), with detection latency (in seconds) as a predictor and random intercepts by subject and by item showed that detection latency and verbal response latency were positively correlated, with trials in which the phoneme was detected more slowly showing slower verbal responses to the question (β = 116 ms; SE = 25 ms; t = 4.593; p < .001).

Verbal response latencies in Experiment 2 (tone detection plus question answering)

To test whether participants also started to plan their verbal response in overlap with the question in the early planning condition in Exp. 2, latencies of verbal responses to critical questions (N = 2832) were modelled in a linear mixed-effects regression model with early versus late Planning as a fixed predictor and random intercepts by subject and by item (Fig. 4, right panels).[11] Planning was dummy coded, with early planning as the reference level. Subjects were found to respond significantly slower to late planning questions than to early planning questions (Table 6). An additional model predicting verbal response latencies in trials in which the target phoneme was detected (n = 2794), with detection latency (in seconds) as a predictor and random intercepts by subject and by item showed that detection latency and verbal response latency were positively correlated, with trials in which the phoneme was detected more slowly showing slower verbal responses to questions (β = 147 ms; SE = 56 ms; t = 2.624; p < .01).

Picture naming latencies in Experiment 1c (phoneme detection plus picture naming)

To test whether picture naming latencies in Exp. 1c were affected by Planning (early vs. late)[12] or not, verbal response times (N = 3075) were modelled in a linear mixed-effects regression model with Planning as a fixed predictor and random intercepts by subject and by item[13] (see Fig. A.1 in the Appendix). Detection of the target phoneme (detected vs. missed) was included as a control predictor, together with its interaction with Planning. Both Planning and Detection were dummy coded, with early planning and phoneme detected as reference levels. No significant difference in picture naming latencies was found between early and late planning questions (Table 7), indicating that performance

[10] The respective mirror relation of verbal response times and detection performance was attested in a separate model on detection rates, with verbal response times as a predictor, finding that late planning trials with longer verbal response latencies show reduced phoneme detection rates (β = −0.160; SE = 0.071; z = −2.233; p = .025; see Tables A.1 and A.2 in a.

[11] Detection of the target phoneme (detected vs. missed) was not included as a control predictor, as results on the effects of Detection on verbal reaction times cannot be expected to be reliable in Exp. 2 due to a small number of trials in which the target tone was not detected, making prediction very difficult for regression models.

[12] Participants did not have to plan answers to the questions in this Experiment; Planning only indicates the type of question here, parallel to the other experiments.

[13] Planning was not included as a random slope in order to avoid singularity of model fit.

Table 5
Model output of mixed-effects model on verbal response latencies in Experiment 1a.

	β	SE	t	p	
Intercept	1.235	0.007	15.514		
Planning_late	0.146	0.066	2.192	0.032	*
Detection_missed	−0.003	0.062	−0.052	0.958	
Planning_late:Detection_missed	0.227	0.103	2.212	0.027	*

Formula = verbalRT ~ planning * detection + (1 + planning | subjectID) + (1 + planning | itemID). Family = gaussian(link=identity).

Table 6
Model output of mixed-effects model on verbal response latencies in Experiment 2.

	β	SE	t	p	
Intercept	1.125	0.084	13.327		
Planning_late	0.357	0.060	5.859	<0.001	***

Formula = verbalRT ~ planning + (1 + planning | subjectID) + (1 + planning | itemID). Family = gaussian(link=identity).

Table 7
Model output of mixed-effects model on picture naming latencies in Experiment 1c.

	β	SE	t	p
Intercept	1.305	0.051	25.353	
planning_late	0.023	0.019	1.174	0.240
Detection_missed	0.049	0.039	1.259	0.208
Planning_late:Detection_missed	−0.034	0.052	−0.661	0.509

Formula = verbalRT ~ planning * detection + (1 | subjectID) + (1 | itemID). Family = gaussian(link=identity).

in the picture naming task was not affected by the format of the preceding question. Furthermore, no significant effect of Detection was found, and no significant interaction of Detection and Planning.

Discussion

This study investigated whether phonological input processing during turn taking is less accurate while concurrently planning speech than while not concurrently planning speech, as predicted by the *planning-prioritised hypothesis* (in contrast to the *comprehension-prioritised hypothesis*). The reported experiments required participants to perform a manual phoneme detection task either in isolation (Exp. 1b), in combination with a verbal question response task (Exp. 1a), or in combination with a delayed picture naming task (Exp. 1c). Questions were presented in two versions, either allowing participants to start to plan their answer before the target phoneme appeared in the question (early planning condition) or allowing them to plan the answer only after the target phoneme had been presented (late planning condition).

Participants were found to show significantly reduced phoneme detection performance when they were concurrently planning their response to the quiz question (i.e., in the early planning condition in Exp. 1a) as compared to when they were not (yet) planning their verbal response (i.e., in Exp. 1b and 1c and in the late planning condition in Exp. 1a). The finding that question format (early planning question versus late planning question) affected phoneme detection accuracy only in Exp. 1a shows that the effect on detection performance was not due to the position of the target phoneme in the question but to concurrent planning of the answer in the early planning condition in Exp. 1a.

In an additional experiment (Exp. 2), participants were required to verbally answer the questions and perform a tone detection task, with the target tones presented at the positions of the phonemes that were the target phonemes in Exp. 1a. In contrast to the phoneme detection accuracy in Exp. 1a, participants' tone detection accuracy in Exp. 2 was not found to be impaired by concurrent planning, suggesting that phonological, but not general auditory input processing is reduced during concurrent speech planning. However, tone detection performance in Exp. 2 was at ceiling in both planning conditions, which might have covered up a possible interference effect of concurrent planning on tone detection. We will return to this point again in the discussion below.

In line with the results on participants' phoneme detection accuracy, phoneme detections in the main dual-task experiment (Exp. 1a) were found to be significantly slower in the early planning condition, in which speech planning happened in overlap with the target phoneme, than in the late planning condition, in which phoneme detection was performed without concurrent speech planning. The fact that this effect was only obtained when the question had to be answered verbally (Exp. 1a) but not when no response planning was required (Exp. 1b and 1c) shows that it is indeed due to speech planning in overlap rather than to the position of the respective target phonemes in the two sentence formats. Just as phoneme detection latencies in Exp. 1a, tone detection latencies in Exp. 2 were also found to be significantly slower in the early planning condition than in the late planning condition. This finding was to be expected, since speech planning has previously been shown to impair processes of (minute) motor control and hence affects the manual component of the phoneme/tone detection tasks in the present button press experiments. Earlier studies found speakers' motor activity to be impaired during phases of taxing speech planning, e.g., during mouse-tracking of an object on the screen, finger tapping, or driving (Boiteau, Malone, Peters, & Almor, 2014; Drews, Pasupathi, & Strayer, 2008; Sjerps & Meyer, 2015, see also below), possibly because at least some components of the manual detection task (like planning and executing the finger movement) and the speech preparation task (like planning and preparing the movement of the involved articulators) are carried out by processing mechanisms that are shared by these two tasks.

The verbal response latencies of the answers given to the quiz questions in the phoneme detection dual-task experiment (Exp. 1a) and the tone detection dual-task experiment (Exp. 2) were significantly faster in the early planning condition, where speech planning and phoneme/tone monitoring are executed concurrently, than in the late planning condition, where phoneme/tone monitoring and speech planning are executed consecutively, showing that subjects did indeed plan their answers in overlap with the incoming questions, if possible. This result is in line with previous findings on early planning strategies in dialogue situations (e.g., Barthel et al. (2016) on German; Barthel and Levinson (2020) and Bögels (2020) on Dutch; Corps et al. (2018) on English). Notably, response latencies in the late planning condition were particularly long in trials in which the critical phoneme was not detected. Since phonemes in the late planning condition were certainly not missed because of concurrent speech planning, these (rare) misses must be due to temporarily low levels of attention spent on the experimental task, which also lead to slow verbal responses in these trials. Moreover, verbal response latencies and phoneme detection latencies were positively correlated, with trials with longer verbal response times showing later reactions to the encounter of the target phonemes. This correlation was not found for verbal response latencies and tone detection latencies. These results are in line with previous research finding that both language comprehension and speech planning require attention (e.g., Cohen, Salondy, Pallier, & Dehaene, 2021; Roelofs, 2021; Roelofs & Piai, 2011).

Taken together, the obtained findings clearly support the planning-prioritised hypothesis. Participants were shown to plan their responses to the questions as early as possible, which was in overlap with part of the question in the early planning condition. When processing capacities were supposedly insufficient, which was most frequent in early planning questions in Exp. 1a, the target phonemes were not detected. In late planning questions, detection misses were much less frequent and most probably due to a momentary lack of attention to the experimental task. In these rare cases, this lack of attention also lead to slower verbal response latencies. This conclusion finds support in the fact that questions that are particularly taxing, and thus get slower responses, were found to also show significantly reduced detection performance as compared to questions that get faster responses.

The attested pattern of results can be explained in two plausible, interconnected ways. As one possible, domain-general explanation, reduced phoneme detection performance in the early planning condition in Exp. 1a might be caused by divided attention and increased processing load during parallel speech planning and comprehension. Alternatively, reduced phoneme detection task performance might be driven by domain-specific interference between the processes of speech planning and language decoding, especially on the level of phonological processing. Each of these two accounts will be discussed in the following paragraphs.

Attention plays a crucial role in managing conversations in general and verbal turn taking in particular. Understanding and planning speech are complex cognitive tasks that rely heavily on attention (Chwilla, 2022; He et al., 2021; Hubbard & Federmeier, 2021; Jongman, Roelofs, & Meyer, 2015; Kristensen, Wang, Petersson, & Hagoort, 2013; Laganaro, Bonnans, & Fargier, 2019; Moisala et al., 2015; Shitova, Roelofs, Coughler, & Schriefers, 2017). Phases of speech planning in overlap with input decoding can therefore temporarily exhaust the limited capacity of attentional resources, as has been demonstrated in previous dual-task studies. For instance, Ferreira and Pashler (2002) demonstrated that in dual-tasks involving discriminating tones while preparing to name a picture, both picture naming performance and tone discrimination performance were reduced in difficult naming conditions (e.g., naming low-frequency words after low-cloze context sentences) as compared to easy naming conditions (e.g., naming high-frequency words after high-cloze context sentences), indicating that more difficult speech planning tasks lead to increased attentional demand. Investigating actual conversations, Boiteau et al. (2014) found that interlocutors' performance of continuously tracking a moving dot with their mouse was reduced during phases of speech planning, where speech comprehension and preparation regularly overlap. Similarly, Sjerps and Meyer (2015) found that participants who had to continuously tap their fingers in a pre-defined pattern while taking turns with the computer to name rows of pictures showed reduced tapping performance already shortly before the end of the incoming turn, i.e., just as they planned their own turn in overlap with speech comprehension. And a number of earlier experiments showed that both speech production and speech planning reduce concurrent driving performance (e.g., Kubose et al., 2006), which leads drivers to make use of numerous, partly multi-modal strategies to manage the attentional demand of the multi-task situation (Drews et al., 2008; Mondada, 2012). Thus, both language production planning and speech comprehension require attentional resources, which can lead to capacity limitations during phases of simultaneous processing.

The competition for attention of concurrent speech planning and comprehension has been shown to lead to increased processing load at turn transitions. In a controlled experimental setting in which participants exchanged conditionally relevant speaking turns with a confederate, Barthel and Sauppe (2019) found that participants' pupil sizes, an indicator for cognitive load (Beatty & Lucero-Wagoner, 2000), increased more intensely and for longer when they were planning their turns in overlap with the incoming turn as compared to when they planned their turn in silence after the incoming turn. This pattern has been confirmed in question-response sequences taken from unrestricted conversations (Barthel & Rühlemann, 2025), showing that conversational dual-tasking is cognitively taxing, even though it has been shown to be a common turn taking strategy (e.g., Bögels, 2020; Levinson & Torreira, 2015). Notably, the increase in processing load at turn transitions is not dependent on the actual articulation of the planned speech but has also been observed in potential next speakers in triadic conversations who do not actually produce the next turn but were merely co-selected as a potential next speaker (Rühlemann & Barthel, 2025).

In light of these previous findings, attention capacity limitations during phases of speech planning are a plausible candidate explanation for the observed reduced performance in phonological input processing. Both comprehension and production are known to be dependent on speakers' domain-general aptitude and information-processing speed (Hintz et al., 2020). Limitations in general processing resources or a processing bottleneck (Ferreira & Pashler, 2002; Kahneman, 1973; Navon & Miller, 2002; Pashler, 2000; Ruthruff, Pashler, & Klaassen, 2001; Tombu & Jolicœ ur, 2003, 2005) are thus conceivable causes of the reduced phoneme detection performance, and would explain the obtained pattern of results. Both phoneme detection accuracy and latencies were lower and longer, respectively, when subjects were planning their verbal response at the time when the target phoneme was encountered. Moreover, even in trials in which planning and comprehension did not overlap, verbal response latencies were longer when the critical phoneme was missed as compared to when it was detected. The same pattern holds the other way around, with late planning trials (where comprehension and planning did not overlap) showing a negative relation between phoneme detection rates and verbal response latencies, with more target phonemes being missed in trials with longer verbal response latencies. In line with these results, verbal response latencies and phoneme detection latencies were also found to be correlated, with trials with longer verbal reaction times also showing longer phoneme detection latencies. The combination of these findings indicates that a general momentary lapse of attention can affect both the processes of speech planning and phoneme detection.

An alternative, related approach to explain the reduced performance in phoneme detection during parallel speech planning centres around domain-specific dual-task interference. Speech production and comprehension are both assumed to operate on highly interconnected, if not partly identical mental representations (Indefrey & Levelt, 2004; Kempen, Olsthoorn, & Sprenger, 2012; MacKay, 1987), including coupled phonological representations (Buchsbaum, Hickok, & Humphries, 2001; Kittredge & Dell, 2016; Mitterer & Ernestus, 2008), and to operate on at least partly overlapping neural systems (Hagoort & Indefrey, 2014; Menenti et al., 2011; Silbert et al., 2014). When production and comprehension processes try to access the target representations at the same time, task performance is known to decrease. Abundant evidence for such interference effects comes from picture-word-interference experiments, which show that linguistic input can have detrimental effects on picture naming performance (e.g., Damian & Bowers, 2003; Glaser & Diingelhoff, 1984; Jescheniak, Schriefers, & Hantsch, 2003; Levelt et al., 1999). As a general finding of these studies, the observed interference of speech input with speech production is reduced when the input is phonologically similar to the intended output (Damian & Martin, 1999; Schriefers et al., 1990) (see also Barthel & Levinson, 2020, for related evidence in lexical decision). The standard explanation for this phonological similarity effect is that any input activates its respective phonological representations, which compete with the target representations that need to be selected to produce the output, causing phonological interference. Thus, any phonological overlap between input and output leads to activation of the output-relevant representations by the input, leading to reduced interference. This explanation is supported by findings of dual-task experiments that show that a picture naming task can be performed more easily with a concurrent tone-discrimination task than with a concurrent syllable-identification task (Fairs, Bögels, & Meyer, 2018); an effect that has been argued to be due to the fact that non-linguistic tones, contrary to syllables, do not activate competing linguistic representations. Corroborating results come from a phoneme monitoring study by Roelofs, Özdemir, and Levelt (2007), in which participants were presented with single spoken words and found to show shorter phoneme monitoring latencies when the name of a picture that was presented simultaneously also contained the target phoneme. Importantly, this phoneme priming effect was not found when subjects never had to react to the presented pictures in any way, which shows that speech production, including active phonological encoding, is the source of the obtained interference effect.

Based on these previous results, both attentional capacity limitations during dual-task phases as well as phonological interference of speech planning with phonological decoding are plausible candidate explanations for the reduction of phonological input processing efficiency during verbal response planning. Moreover, these two causes are not mutually exclusive and might in fact both have a share in the observed net effects. However, given the present results of the dual-task of tone detection and question answering in Exp. 2, language-specific effects seem to be the more likely (main) cause of the observed effects on phoneme detection performance in Exp. 1a. Tone detection accuracy was found to be equally high during phases of concurrent speech planning and phases without concurrent speech planning in Exp. 2, indicating that interference between response planning and input processing, especially on the phonological level, is the more likely reason for the observed decline in phoneme detection performance during speech planning in Exp. 1a than domain general attention capacity limitations. These results suggest that phonological input processing in particular is impaired during phases of concurrent speech planning in turn taking situations.

The absence of an effect of concurrent speech planning of tone detection performance in Exp. 2 needs to be interpreted with some caution, though. In both the early and the late planning conditions, tone detection performance is almost at ceiling for many of the tested participants. This limits the certainty about these results' interpretation, due to the possibility that any detrimental effect of concurrent planning on tone detection was not detected because the tone detection task was too easy even in the early planning condition. Future studies could directly target this remaining uncertainty by testing a dual-task scenario that balances the level of difficulty between the language-related and the language-unrelated task more carefully. Furthermore, the phoneme detection task and the tone detection task are inherently not perfectly comparable, since in the tone detection task, participants need to pay attention to a part of the incoming signal that is not only not part of the phonological makeup of the input that needs to be comprehended for the quiz task, but also is it not part of the linguistic signal in general. Thus, the two tasks differ from one another in more than one respect, leaving some uncertainty about why exactly concurrent planning did not affect tone detection performance: only because the target tones are not part of the phonological signal of the input or because they are entirely non-linguistic. For the rational of this study's manipulations and comparisons, however, this was unavoidable.

Regardless of the respective shares of phonological versus general auditory processing in the observed net effect in the present dual-task study, phoneme detection has been shown to be more difficult during phases of concurrent speech planning, even though parallel processing of verbal input and output is a common strategy of interlocutors in turn taking situations (e.g., Barthel et al., 2017; Bögels et al., 2015). One reason why interlocutors commonly pursue this strategy in spite of these processing difficulties is that they are thereby able to start to articulate their responses with shorter gaps between turns of talk, making more efficient use of the time spent in conversation (Bögels & Levinson, 2017; Levinson & Torreira, 2015). A second reason for an overlapping processes strategy is that being able to produce timely well-aligned turns avoids any potentially unwanted mis-interpretations that might be triggered by speaking after a markedly long gap, which has been shown to lead to inferences of, for example, social distance (Blohm & Barthel, 2025) or low willingness to grant a request or accept an offer (Henetz, 2017; Roberts & Francis, 2013). A third and very fundamental reason lies in the basic systematics of turn allocation during conversation (Sacks et al., 1974). Whenever speaker transition becomes relevant during a conversation and the now-open speaking floor is not taken by the listener of the previous turn, the previous speaker might self-select again for the next turn and the present opportunity for the listener to take a turn is gone. This tension is obviously even more pronounced in multi-party interactions, where situations in which more than one current listener is entitled to take the next turn are common (Holler et al., 2021; Rühlemann & Barthel, 2025).

Table A.1
Model output of generalised linear mixed-effects model on detection rate in Experiment 1a.

	β	*SE*	*z*	*p*	
Intercept	2.578	0.265	9.728		
Planning_early	−1.323	0.212	−6.231	<0.001	***
verbalRT_centred	−0.160	0.071	−2.233	0.025	*
Phoneme_1	−0.275	0.264	−1.043	0.297	
Phoneme_2	−0.397	0.273	−1.455	0.145	
Phoneme_3	−0.083	0.256	−0.326	0.744	
Planning_early:verbalRT_c	0.162	0.111	1.453	0.146	

Formula = hit ~ 1 + planning * verbalRT_centred + phoneme + (1 + planning | subjectID) + (1 + planning | itemID); Family = binomial(link=logit).

In sum, planning in overlap is a frequently employed strategy, even though its execution is less efficient than planning in silence (Barthel & Sauppe, 2019). The presented evidence indicates that not only is planning less efficient in overlap with comprehension but comprehension efficiency is also reduced during concurrent planning (see also Fargier & Laganaro, 2019). After related evidence had been put forth for language comprehension on the level of semantic processing (Barthel, 2021), the present results show that also phonological input processing is reduced during parallel speech planning in dialogical turn taking situations, lending further support to the planning-prioritised hypothesis.

CRediT authorship contribution statement

Mathias Barthel: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

This research was partly funded by DFG grant InterTurn, Germany (grant No. 497498146) granted to the author, and partly by Berlin University Alliance grant Language Comprehension in Dialogue, Germany granted to the author.

Appendix

See Fig. A.1 and Tables A.1–A.3.

Data availability

The list of stimuli, raw data, and analysis scripts are available on OSF at https://osf.io/fzha2/?view_only=b7cfc71964d84864b9c5c660fec8cc6f.

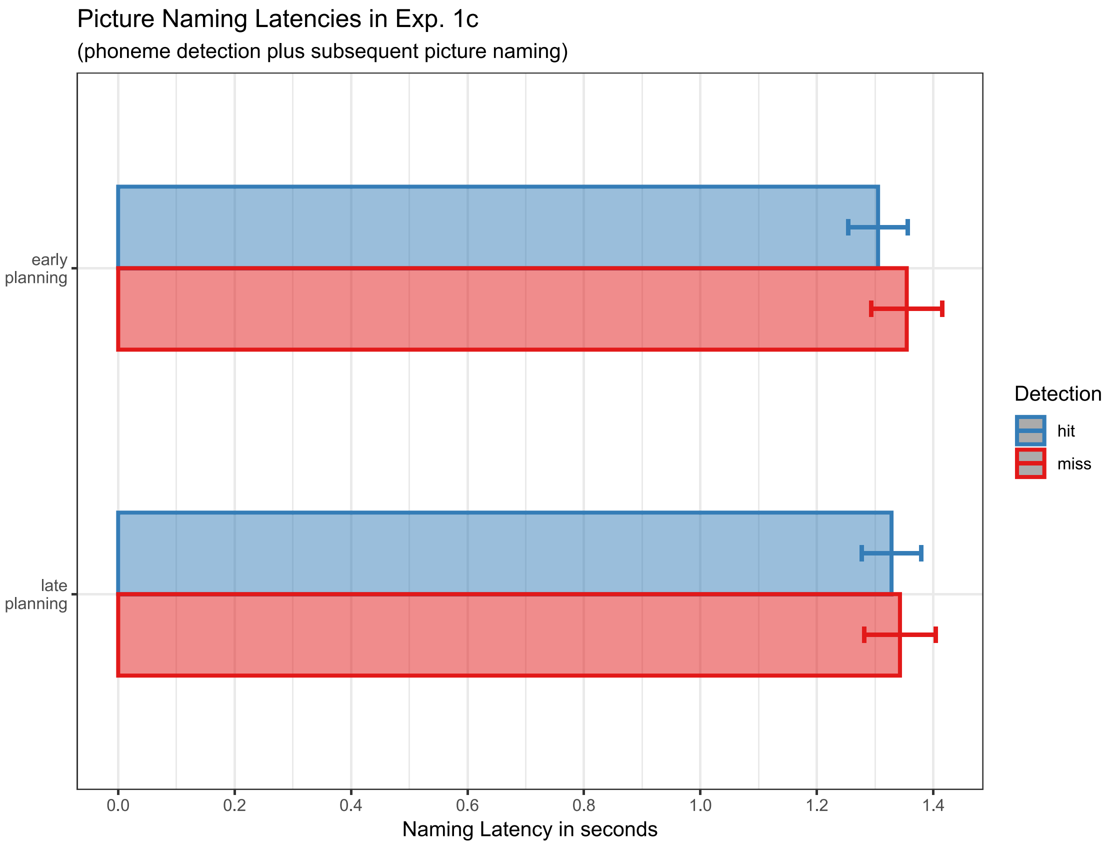

Fig. A.1. Model predictions of picture naming latencies in the Experiment 1c (see Table 7 for model details). Error bars represent one standard deviation from the mean. $n_{early/hit}$ = 1256; $n_{early/miss}$ = 231; $n_{late/hit}$ = 1372; $n_{late/miss}$ = 216.

Table A.2
Post-hoc tests on effect of verbal response latency on phoneme detection rate by condition in Experiment 1a.

Planning	*β*	*SE*	*z*	*p*	
early	0.002	0.087	0.026	0.999	
late	−0.160	0.071	−2.233	0.050	*

Table A.3
Model output of linear mixed-effects model on phoneme/tone detection latencies in Experiments 1a and 2.

	β	*SE*	*t*	*p*	
Intercept	1073.22	64.28	16.696		
Planning_late	−146.11	33.91	−4.308	<0.001	***
Experiment_2	−352.42	48.89	−7.209	<0.001	***
verbalRT_centred	159.83	24.43	6.543	<0.001	***
Phoneme_1	73.56	71.24	1.033	0.306	
Phoneme_2	−26.77	72.36	−0.370	0.712	
Phoneme_3	7.35	71.33	0.103	0.918	
Planning_late:Exp._2	32.13	42.57	0.755	0.451	*
Planning_late:verbalRT_c	−71.24	29.23	−2.437	0.014	*
verbalRT_centred:Exp._2	−132.43	29.17	−4.539	<0.001	***
Planning_late:verbalRT_centred:Exp._2	66.35	36.47	1.819	0.069	.

Formula = buttonRT ~ planning * experiment * verbalRT_centred + phoneme + (1 + planning | subjectID) + (1 + planning | itemID); Family = gaussian(link=identity).

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. http://dx.doi.org/10.1016/S0010-0277(99)00059-1.

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

Audacity Team (2024). Audacity.

Baddeley, A. (2003). Working memory and language: an overview. *Journal of Communication Disorders*, *36*(3), 189–208. http://dx.doi.org/10.1016/S0021-9924(03)00019-4.

Barthel, M. (2020). *Speech planning in dialogue - psycholinguistic studies of the timing of turn taking* (Ph.D. thesis), Nijmegen: Radboud University Nijmegen.

Barthel, M. (2021). Speech planning interferes with language comprehension: Evidence from semantic illusions in question-response sequences. In *Proceedings of the 25th workshop on the semantics and pragmatics of dialogue* (pp. 1–14). Potsdam, Germany.

Barthel, M., & Levinson, S. C. (2020). Next speakers plan word forms in overlap with the incoming turn: evidence from gaze-contingent switch task performance. *Language, Cognition and Neuroscience*, *35*(9), 1183–1202. http://dx.doi.org/10.1080/23273798.2020.1716030.

Barthel, M., Meyer, A. S., & Levinson, S. C. (2017). Next speakers plan their turn early and speak after turn-final "go-signals". *Frontiers in Psychology*, *8*, 393. http://dx.doi.org/10.3389/fpsyg.2017.00393.

Barthel, M., & Rühlemann, C. (2025). Pupil size indicates planning effort at turn transitions in natural conversation. In E. Zima, & A. Stukenbrock (Eds.), *vol. 351*, *New perspectives on gaze in social interaction: mobile eye tracking studies* (pp. 188–205). Amsterdam: John Benjamins Publishing Company, https://doi.org/10.1075/pbns.351.07bar.

Barthel, M., & Sauppe, S. (2019). Speech planning at turn transitions in dialog is associated with increased processing load. *Cognitive Science*, *43*(7), Article e12768. http://dx.doi.org/10.1111/cogs.12768.

Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. *Frontiers in Psychology*, *7*, 1858. http://dx.doi.org/10.3389/fpsyg.2016.01858.

Barthel, M., Tomasello, R., & Liu, M. (2024). Conditionals in context: Brain signatures of prediction in discourse processing. *Cognition*, *242*, Article 105635. http://dx.doi.org/10.1016/j.cognition.2023.105635.

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

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 142–162). New York: Cambridge University Press.

Bles, M., & Jansma, B. M. (2008). Phonological processing of ignored distractor pictures, an fMRI investigation. *BMC Neuroscience*, *9*(1), 20. http://dx.doi.org/10.1186/1471-2202-9-20.

Blohm, S., & Barthel, M. (2025). Why so cold and distant? Effects of inter-turn gap durations on observers' attributions of interpersonal stance. In *Proceedings of the 29th workshop on the semantics and pragmatics of dialogue* (pp. 1–9). Bielefeld, Germany.

Bögels, S. (2020). Neural correlates of turn-taking in the wild: Response planning starts early in free interviews. *Cognition*, *203*, Article 104347. http://dx.doi.org/10.1016/j.cognition.2020.104347.

Bögels, S., & Levinson, S. C. (2017). The brain behind the response: Insights into turn-taking in conversation from neuroimaging. *Research on Language and Social Interaction*, *50*(1), 71–89. http://dx.doi.org/10.1080/08351813.2017.1262118.

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, *5*(12881), 1–11. http://dx.doi.org/10.1038/srep12881.

Boiteau, T. W., Malone, P. S., Peters, S. A., & Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *Journal of Experimental Psychology: General*, *143*(1), 295–311. http://dx.doi.org/10.1037/a0031858.

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, *2*(1), 16. http://dx.doi.org/10.5334/joc.72.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 9. http://dx.doi.org/10.5334/joc.10.

Buchsbaum, B. R., Hickok, G., & Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, *25*(5), 663–678. http://dx.doi.org/10.1207/s15516709cog2505_2.

Bürki, A., Elbuy, S., Madec, S., & Vasishth, S. (2020). What did we learn from forty years of research on semantic interference? A Bayesian meta-analysis. *Journal of Memory and Language*, *114*, Article 104125. http://dx.doi.org/10.1016/j.jml.2020.104125.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, Article e62. http://dx.doi.org/10.1017/S0140525X1500031X.

Chwilla, D. J. (2022). Context effects in language comprehension: The role of emotional state and attention on semantic and syntactic processing. *Frontiers in Human Neuroscience*, *16*, http://dx.doi.org/10.3389/fnhum.2022.1014547.

Cohen, L., Salondy, P., Pallier, C., & Dehaene, S. (2021). How does inattention affect written and spoken language processing? *Cortex*, *138*, 212–227. http://dx.doi.org/10.1016/j.cortex.2021.02.007.

Corps, R. E., 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. http://dx.doi.org/10.1016/j.cognition.2018.01.015.

Damian, M. F., & Bowers, J. S. (2003). Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin & Review*, *10*(1), 111–117. http://dx.doi.org/10.3758/BF03196474.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(2), 345–361.

Davidson, J. (1984). Subsequent versions of invitations, offers, requests, and proposals dealing with potential or actual rejection. In J. Atkinson, & J. Heritage (Eds.), *Structures of social action: studies in conversation analysis* (pp. 102–128). Cambridge: Cambridge University Press.

Drews, F. A., Pasupathi, M., & Strayer, D. L. (2008). Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology: Applied*, *14*(4), 392–400. http://dx.doi.org/10.1037/a0013119.

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

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

Fairs, A., & Strijkers, K. (2021). Can we use the internet to study speech production? Yes we can! evidence contrasting online versus laboratory naming latencies and errors. In S. Sulpizio (Ed.), *PLOS One*, *16*(10), Article e0258908. http://dx.doi.org/10.1371/journal.pone.0258908.

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

Fargier, R., & Laganaro, M. (2019). Interference in speaking while hearing and vice versa. *Scientific Reports*, *9*(1), 5375. http://dx.doi.org/10.1038/s41598-019-41752-7.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. http://dx.doi.org/10.3758/BF03193146.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. http://dx.doi.org/10.1111/1467-8721.00158.

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

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. http://dx.doi.org/10.1111/j.1749-818X.2007.00007.x.

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Gambi, C., & Pickering, M. J. (2017). Models linking production and comprehension. In E. M. Fernández, & H. S. Cairns (Eds.), *The handbook of psycholinguistics* (pp. 157–181). Hoboken, NJ, USA: John Wiley & Sons, Inc., http://dx.doi.org/10.1002/9781118829516.ch7.

Gisladottir, R. S., Bögels, S., & Levinson, S. C. (2018). Oscillatory brain responses reflect anticipation during comprehension of speech acts in spoken dialog. *Frontiers in Human Neuroscience*, *12*, http://dx.doi.org/10.3389/fnhum.2018.00034.

Glaser, W. R., & Diingelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Learning & Memory*, *10*(5), 640–654.

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

Hagoort, P. (2019). The neurobiology of language beyond single-word processing. *Science*, *366*(6461), 55–58. http://dx.doi.org/10.1126/science.aax0289.

Hagoort, P., & Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual Review of Neuroscience*, *37*(1), 347–362. http://dx.doi.org/10.1146/annurev-neuro-071013-013847.

Halekoh, U., & Hojsgaard, S. (2014). A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models - the R package pbkrtest. *Journal of Statistical Software*, *59*(9), 1–30.

He, J., Meyer, A. S., & Brehm, L. (2021). Concurrent listening affects speech planning and fluency: the roles of representational similarity and capacity limitation. *Language, Cognition and Neuroscience*, 1–23. http://dx.doi.org/10.1080/23273798.2021.1925130.

Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & de Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences*, *119*(32), Article e2201968119. http://dx.doi.org/10.1073/pnas.2201968119.

Henetz, T. (2017). *Don't hesitate! the length of inter-turn gaps influences observers' interactional attributions* (Ph.D. thesis), Stanford: Stanford University.

Hintz, F., Jongman, S. R., Dijkhuis, M., van, V., McQueen, J. M., & Meyer, A. S. (2020). Shared lexical access processes in speaking and listening? an individual differences study. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *46*(6).

Holler, J., Alday, P. M., Decuyper, C., Geiger, M., Kendrick, K. H., & Meyer, A. S. (2021). Competition reduces response times in multiparty conversation. *Frontiers in Psychology*, *12*, Article 693124. http://dx.doi.org/10.3389/fpsyg.2021.693124.

Hubbard, R. J., & Federmeier, K. D. (2021). Dividing attention influences contextual facilitation and revision during language comprehension. *Brain Research*, *1764*, Article 147466. http://dx.doi.org/10.1016/j.brainres.2021.147466.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. http://dx.doi.org/10.1016/j.brainres.2015.02.014.

Hustá, C., Nieuwland, M., & Meyer, A. (2023). Effects of picture naming and categorization on concurrent comprehension: Evidence from the N400. *Collabra: Psychology*, *9*(1), 88129. http://dx.doi.org/10.1525/collabra.88129.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2*, 255. http://dx.doi.org/10.3389/fpsyg.2011.00255.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1–2), 101–144. http://dx.doi.org/10.1016/j.cognition.2002.06.001.

Jescheniak, J. D., Schriefers, H., & Hantsch, A. (2003). Utterance format effects phonological priming in the picture-word task: Implications for models of phonological encoding in speech production.. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 441–454. http://dx.doi.org/10.1037/0096-1523.29.2.441.

Jongman, S. R., Roelofs, A., & Meyer, A. S. (2015). Sustained attention in language production: An individual differences investigation. *Quarterly Journal of Experimental Psychology*.

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

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. http://dx.doi.org/10.1016/S0749-596X(03)00023-8.

Kempen, G., Olsthoorn, N., & Sprenger, S. (2012). Grammatical workspace sharing during language production and language comprehension: Evidence from grammatical multitasking. *Language and Cognitive Processes*, *27*(3), 345–380. http://dx.doi.org/10.1080/01690965.2010.544583.

Kendrick, K. H., & Torreira, F. (2014). The timing and construction of preference: A quantitative study. *Discourse Processes*, *52*(4), 1–35. http://dx.doi.org/10.1080/0163853X.2014.955997.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*(3), 983–997. http://dx.doi.org/10.2307/2533558.

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

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

Kubose, T. T., Bock, K., Dell, G. S., Garnsey, S. M., Kramer, A. F., & Mayhugh, J. (2006). The effects of speech production and speech comprehension on simulated driving performance. *Applied Cognitive Psychology*, *20*(1), 43–63. http://dx.doi.org/10.1002/acp.1164.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). **lmerTest** package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), http://dx.doi.org/10.18637/jss.v082.i13.

Laganaro, M., Bonnans, C., & Fargier, R. (2019). Word form encoding is under attentional demand: evidence from dual-task interference in aphasia. *Cognitive Neuropsychology*, *36*(1–2), 18–30. http://dx.doi.org/10.1080/02643294.2018.1564650.

Lenth, R. V. (2022). Emmeans.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01), 1–75. http://dx.doi.org/10.1017/S0140525X99001776.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*(731), 10–26. http://dx.doi.org/10.3389/fpsyg.2015.00731.

MacKay, D. G. (1987). In M. M. Sebrechts, G. Fischer, & P. M. Fischer (Eds.), *Cognitive science series*, *The organization of perception and action*. New York, NY: Springer New York, http://dx.doi.org/10.1007/978-1-4612-4754-8.

Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science*, *22*(9), 1173–1182. http://dx.doi.org/10.1177/0956797611418347.

Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition*, *35*(3), 494–503. http://dx.doi.org/10.3758/BF03193289.

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

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

Mondada, L. (2012). Talking and driving: Multiactivity in the car. *Semiotica*, *2012*(191), 223–256. http://dx.doi.org/10.1515/sem-2012-0062.

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language*, *53*(3), 359–377. http://dx.doi.org/10.1016/j.jml.2005.05.001.

Navon, D., & Miller, J. (2002). Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology*, *44*(3), 193–251. http://dx.doi.org/10.1006/cogp.2001.0767.

Pashler, H. (2000). Task switching and multitask performance. In S. Monsell, & J. Driver (Eds.), *Control of cognitive processes*. The MIT Press, http://dx.doi.org/10.7551/mitpress/1481.001.0001.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190; discussion 190–226. http://dx.doi.org/10.1017/S0140525X04000056.

Pickering, M. J., & Garrod, S. (2013). How tightly are production and comprehension interwoven? *Frontiers in Psychology*, *4*, 44681. http://dx.doi.org/10.3389/fpsyg.2013.00238, Publisher: Frontiers.

Pomeranz, A., Atkinson, J., & Heritage, J. (1984). Agreeing and disagreeing with assesments: some features of preferred/dispreffered turn shapes. In *Structures of social action* (pp. 53–101). Cambridge: Cambridge University Press.

Pomeranz, A., & Heritage, J. (2012). Preference. In J. Sidnell, & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 210–228). Chichester, UK: John Wiley & Sons, Ltd.

R. Core Team (2023). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Roberts, F., & Francis, A. L. (2013). Identifying a temporal threshold of tolerance for silent gaps after requests. *Journal of the Acoustical Society of America*, *133*(6), EL471–EL477. http://dx.doi.org/10.1121/1.4802900.

Roberts, F., Francis, A. L., & Morgan, M. (2006). The interaction of inter-turn silence with prosodic cues in listener perceptions of "trouble" in conversation. *Speech Communication*, *48*(9), 1079–1093. http://dx.doi.org/10.1016/j.specom.2006.02.001.

Roberts, S. G., & Levinson, S. C. (2017). Conversation, cognition and cultural evolution: A model of the cultural evolution of word order through pressures imposed from turn taking in conversation. *Interaction Studies*, *18*(3), 402–442. http://dx.doi.org/10.1075/is.18.3.06rob.

Roberts, F., Margutti, P., & Takano, S. (2011). Judgments concerning the valence of inter-turn silence across speakers of American english, Italian, and Japanese. *Discourse Processes*, *48*(5), 331–354. http://dx.doi.org/10.1080/0163853X.2011.558002.

Rodd, J. M. (2024). Moving experimental psychology online: How to obtain high quality data when we can't see our participants. *Journal of Memory and Language*, *134*, Article 104472. http://dx.doi.org/10.1016/j.jml.2023.104472.

Roelofs, A. (2021). How attention controls naming: Lessons from Wundt 2.0. *Journal of Experimental Psychology: General*, *150*(10), 1927–1955. http://dx.doi.org/10.1037/xge0001030.

Roelofs, A., Özdemir, R., & Levelt, W. J. M. (2007). Influences of spoken word planning on speech recognition. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(5), 900–913. http://dx.doi.org/10.1037/0278-7393.33.5.900.

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

Rühlemann, C., & Barthel, M. (2024). Word frequency and cognitive effort in turns-at-talk: turn structure affects processing load in natural conversation. *Frontiers in Psychology*, *15*, Article 1208029. http://dx.doi.org/10.3389/fpsyg.2024.1208029.

Rühlemann, C., & Barthel, M. (2025). Speech planning depends on next-speaker selection: Evidence from pupillometry in question-answer sequences in naturalistic triadic conversation. *Discourse Processes*.

Ruthruff, E., Pashler, H. E., & Klaassen, A. (2001). Processing bottlenecks in dual-task performance: Structural limitation or strategic postponement? *Psychonomic Bulletin & Review*, *8*(1), 73–80. http://dx.doi.org/10.3758/BF03196141.

Ryskin, R., & Nieuwland, M. S. (2023). Prediction during language comprehension: what is next? *Trends in Cognitive Sciences*, http://dx.doi.org/10.1016/j.tics.2023.08.003, S1364661323001997.

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

Sauppe, S. (2017). Word order and voice influence the timing of verb planning in german sentence production. *Frontiers in Psychology*, *8*, 1648. http://dx.doi.org/10.3389/fpsyg.2017.01648.

Schnur, T. T., Costa, A., & Caramazza, A. (2006). Planning at the phonological level during sentence production. *Journal of Psycholinguistic Research*, *35*(2), 189–213. http://dx.doi.org/10.1007/s10936-005-9011-6.

Schriefers, H., Meyer, A. S., & Levelt, W. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*, 86–102.

Searle, S. R., Speed, F. M., & Milliken, G. A. (1980). Population marginal means in the linear model: An alternative to least squares means. *The American Statistician*, *34*(4), 216–221. http://dx.doi.org/10.1080/00031305.1980.10483031.

Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012). Shared syntax in language production and language comprehension–an fMRI study. *Cerebral Cortex*, *22*(7), 1662–1670. http://dx.doi.org/10.1093/cercor/bhr249.

Shitova, N., Roelofs, A., Coughler, C., & Schriefers, H. (2017). P3 event-related brain potential reflects allocation and use of central processing capacity in language production. *Neuropsychologia*, *106*, 138–145. http://dx.doi.org/10.1016/j.neuropsychologia.2017.09.024.

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, *111*(43), E4687–E4696. http://dx.doi.org/10.1073/pnas.1323812111.

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

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592. http://dx.doi.org/10.1073/pnas.0903616106.

Strijkers, K., & Costa, A. (2011). Riding the lexical speedway: A critical review on the time course of lexical selection in speech production. *Frontiers in Psychology*, *2*(356), 1–16. http://dx.doi.org/10.3389/fpsyg.2011.00356.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, *29*(6), 557–580.

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

Tombu, M., & Jolicœ ur, P. (2005). Testing the predictions of the central capacity sharing model. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(4), 790–802. http://dx.doi.org/10.1037/0096-1523.31.4.790.

Vogt, A., Hauber, R. C., Kuhlen, A. K., & Abdel Rahman, R. (2021). Internet based language production research with overt articulation: Proof of concept, challenges, and practical advice. http://dx.doi.org/10.31234/osf.io/cyvwf, (preprint), PsyArXiv.

Wilshire, C., Singh, S., & Tattersall, C. (2016). Serial order in word form retrieval: New insights from the auditory picture–word interference task. *Psychonomic Bulletin & Review*, *23*(1), 299–305. http://dx.doi.org/10.3758/s13423-015-0882-8.

Zehr, J., & Schwarz, F. (2018). PennController for internet based experiments (IBEX). http://dx.doi.org/10.17605/OSF.IO/MD832, Publisher: Open Science Framework.