# Capturing the Attentional Trade-off between Speech Planning and Comprehension

Cecília Hustá[1] and Antje Meyer[1,2]

## Abstract

■ In conversation, future speakers often plan speech simultaneously with comprehension, which means that they must divide attentional resources between these processes. In this EEG study, we used responses to linguistic attention probes (i.e., syllable "BA" presented during spoken sentences) to track temporal variations in attention to comprehension. Participants were asked to listen to prerecorded sentences with expected or unexpected sentence-final words. Each sentence was presented twice, once with and once without the attention probe starting 100 msec after the target word onset. Participants saw a picture 50 msec before the target word. Depending on the test block (picture naming or button press), participants either named the picture or pressed the space bar, both after an 850-msec delay. The probes elicited a negative potential approximately 100 msec after probe onset (i.e., an attention probe effect) in all probe conditions. Unexpectedly, neither word expectancy nor speech planning influenced the timing or strength of the attention probe effect. This indicates that expectancy of words in Dutch does not affect the allocation of attention toward these words 100 msec after their onset (i.e., the time of the probe presentation). Interestingly, engaging in speech planning does not seem to divert attentional resources away from comprehension at the moment of probe presentation. These findings imply that listeners are able to effectively distribute their attentional resources between comprehension and speech planning and carry out these processes at the same time. Considering these unexpected findings, using attention probes might not be the best approach to capture variations in temporal attention in dual-task paradigms. ■



## INTRODUCTION

Evidence from laboratory studies has shown that speech planning can take place simultaneously with comprehension (Bögels, Casillas, & Levinson, 2018; Bögels, Magyari, & Levinson, 2015; Sjerps & Meyer, 2015; Boiteau, Malone, Peters, & Almor, 2014). In line with this evidence, corpus studies found that interlocutors often overlap in their turns (e.g., Corps, Knudsen, & Meyer, 2022; Stivers et al., 2009). As both speech planning and listening require attention, attentional resources are shared between speech planning and comprehension during parts of conversations. Evidence from dual-tasking studies examining speech onset latencies as well as early auditory components (e.g., N100) during listening shows that there is mutual interference between speech planning and comprehension (Fargier & Laganaro, 2016, 2019; Daliri & Max, 2016; Bögels et al., 2015). However, studies utilizing the early auditory components only used single syllables as comprehension stimuli and word reading or picture naming as a task to assess speech planning (Fargier & Laganaro, 2019; Daliri & Max, 2016). Thus, there is so far no evidence that concurrent speech planning affects early auditory attention to more complex linguistic stimuli, such as spoken words or sentences. Listeners plan speech while listening to complex sentences in everyday conversations, making it essential to investigate whether speech planning diverts attention away from comprehension of more complex stimuli. We explored whether this was the case, using linguistic attention probes—an EEG method where an auditory comprehension stream is presented with and without the attention probes (i.e., a token of the syllable "BA"). The attention probe effect (i.e., the difference wave in the EEG signal computed by subtracting the signal elicited by a comprehension stimulus without the attention probe from the signal elicited by the same comprehension stimulus with the attention probe) is well suited for measuring auditory attention to comprehension at the moment of probe presentation (Li, Zhang, Li, Zhao, & Du, 2017; Li, Lu, & Zhao, 2014; Astheimer & Sanders, 2009). This is because its amplitude and latency reflect the listener's attentional engagement with the comprehension stream. We first tried to replicate that the attention probe effect is sensitive to modulations of temporal attention to comprehension in Dutch. This was done by overlapping the attention probe with expected compared with unexpected sentence-final words. The unexpected words should require more attention to be processed (e.g., Li et al., 2014; Ehrlich & Rayner, 1981). Importantly,

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, [2]Donders Institute for Brain, Cognition, and Behaviour, Nijmegen, The Netherlands




*https://doi.org/10.1162/jocn.a.97*

we also measured the attention probe effect with and without a concurrent speech planning task, namely, preparing for picture naming. We were interested in establishing whether people dedicated less attention to comprehension if they engage in speech planning. If the attention probe effect shows modulations due to speech planning, we can conclude that attention to comprehension is affected by simultaneous speech planning.

If speech planning leads to a decrease in attention to comprehension and thus a modulation of the attention probe effect, future studies could utilize the attention probe effect as an indirect method for investigating how and when speech planning takes place during comprehension in naturalistic conversations. If the paradigm works, modulations of the attention probe effect can be taken as an indication that speech planning is taking place at the time when the probe was inserted. Thus, this method has potential for investigating a variety of questions with regard to how speech planning is scheduled during comprehension. This is because the attention probes can be inserted at various places, for instance, at the beginning versus the end of an utterance, or overlaying a noun versus a verb. Attention is not constant and fluctuates during comprehension (Wöstmann, Maess, & Obleser, 2021; Jaeger, Mirkovic, Bleichner, & Debener, 2020; Li et al., 2014, 2017; Astheimer & Sanders, 2009; Magne et al., 2005). Specifically, listeners can pay more attention when they expect information of high importance or when listening is difficult. Considering that speech planning also requires attention (e.g., Mädebach, Jescheniak, Oppermann, & Schriefers, 2011; Cook & Meyer, 2008), it might be scheduled differently depending on the attentional fluctuations during comprehension. This means that listeners can choose when they will initiate speech planning. They might, for instance, schedule it to take place when they are comprehending less important parts of a message. If this is the case, the attention probe effect should be reduced at times when listeners have scheduled speech planning. Utilizing linguistic attention probes could bring us closer to understanding the role of temporal attention in the time course of speech planning during conversation. In the remainder of the introduction, we will discuss the role of various types of attention and their role in comprehension as well as conversation.

### Selective Attention in Conversation

Focusing attention on the relevant input is essential for conversation, as selecting the input that will receive more attentional resources allows interlocutors to tune out distracting or irrelevant stimuli. One of the best-known examples of this need to focus is the cocktail party effect, where in noisy situations listeners are successful at attending to a selected stream of speech while largely ignoring other streams of speech that are simultaneously heard (e.g., Brungart & Simpson, 2007; Conway, Cowan, & Bunting, 2001; Moray, 1959; Cherry, 1953).

The importance of selective attention for early auditory perception has been shown in studies using ERPs. ERPs are small voltage changes elicited by populations of neurons in response to events or stimuli. Early perceptual ERP components that start between 100 and 300 msec after stimulus onset are particularly sensitive to selective attention during listening (Woldorff et al., 1993; Schwent, Hillyard, & Galambos, 1976; Schwent & Hillyard, 1975; Hillyard, Hink, Schwent, & Picton, 1973). One of the most investigated components is the auditory N100, which is thought to reflect sound detection. Hillyard et al. (1973) used a dichotic listening paradigm, where participants listened to streams of standard tones, which were occasionally interrupted by oddball tones (i.e., tones with different pitch). Participants were asked to detect the oddball tones in the attended stream. The N100 had a larger amplitude for the oddball tones in the attended than the unattended stream. This indicates that the N100 was modulated by selective attention and that a time window 100 msec after the onset of a simple auditory stimuli shows promise in measuring attentional fluctuations.

More complex stimuli than tones can also elicit and modulate the N100 component. Hansen, Dickstein, Berka, and Hillyard (1983) showed that syllables elicit an N100 with similar morphology and scalp distribution as tones. The amplitude of the syllable-elicited N100 was also modulated by attention. Two recent studies investigated whether the N100 amplitude is reduced by speech planning. Fargier and Laganaro (2019) asked participants to listen to syllables in isolation or while preparing to name pictures that could appear 150, 300, or 450 msec before the syllable onset. They found that the N100 elicited by syllables had a lower amplitude when speech planning started closer to the syllables. The N100 latency was also affected. The N100 had a delayed latency, but only when the picture was preceded by the syllable by 450 msec. Daliri and Max (2016) found similar results, showing that N100 elicited by both syllables and tones had lower amplitudes when participants had to simultaneously read words out loud compared with reading silently. However, they did not find effects on the N100 latency depending on silent reading compared with reading aloud. N100 effects were sensitive to the timing of the N100-eliciting stimulus and the speech-planning stimulus. These studies show the promise of using early perceptual components to capture the impact of speech planning on auditory processing. The next step would be to investigate whether speech planning affects early auditory attention in more complex comprehension stimuli than syllables (i.e., words or sentences).

### Temporally Selective Attention

In the visual domain, it is well established that attention can be directed to a specific time and not just to a specific type of a stimulus (e.g., Correa, Lupiáñez, Madrid, & Tudela, 2006; Griffin, Miniussi, & Nobre, 2002; Miniussi, Wilding, Coull, & Nobre, 1999). For example, if participants hear a click or a cue before a visual target appears, they respond faster to the target, especially if the click is presented between 100 and 150 msec before the target (Bertelson, 1967). This is because their temporally selective attention can focus on the critical time window. Other cue–target contingency experiments showed that the P300 had a larger amplitude and an earlier onset when the target onset could be accurately predicted (Griffin et al., 2002; Miniussi et al., 1999). More relevant for the present study, temporally selective attention also affected early visual components (Correa et al., 2006), and specifically, the amplitude of visual N100 increased with more temporally predictable targets (Griffin et al., 2002). This indicates that temporally selective attention impacts early visual processing.

Similarly to the visual domain, in the auditory domain, attention also fluctuates with time. Three studies investigated effects of temporally selective attention on ERPs in the auditory domain (Sanders & Astheimer, 2008; Lange & Röder, 2006; Lange, Rösler, & Röder, 2003). Lange et al. (2003) presented participants with two white noise stimuli that were either spaced close together or far apart, creating short and long intervals, respectively, which were mixed within blocks. Throughout one block, participants were either asked to attend to the short or to the long intervals. Additionally, the second white noise stimulus was sometimes played at a higher intensity. The participants' task was to detect the stimulus of higher intensity in the intervals of the attended length. The N100 and the P300 (elicited by normal intensity stimuli) had larger amplitudes in the attended compared with unattended intervals. The N100 results were replicated in follow-up experiments, and it was shown that participants also responded faster to stimuli at the attended times (Sanders & Astheimer, 2008; Lange & Röder, 2006). This indicates that listeners can direct attention to a point in time, which affects early auditory perception and speeds up responses.

Temporally selective attention might also be crucial for auditory processing in conversations. This is because attentional demands of conversations likely vary as the interlocutor's turn progresses. Directing attention to the important time points in conversation can aid interlocutors' understanding and speed up their responses. Early perceptual components can be used to capture the involvement of temporally selective attention in conversations.

## Utilizing Linguistic Attention Probes to Study Attentional Fluctuations

How can we use temporally selective attention to study the attentional division between speech planning and comprehension? Researchers have shown that attention fluctuates throughout comprehension (Wöstmann et al., 2021; Astheimer & Sanders, 2009; Magne et al., 2005). Astheimer and Sanders (2009) inserted nonlinguistic attention probes (i.e., beeps) and linguistic attention probes (i.e., spoken syllable "BA") into spoken narratives. The probes were inserted at six locations: 100 and 50 msec before word onset, at word onset, 100 and 50 msec after word onset, and at random control intervals. The N100 elicited by the linguistic attention probes was larger for probes presented at and after word onsets versus probes presented before the word onsets or at control intervals. From here on, the N100 elicited by attention probes will be referred to as the attention probe effect. This study showed that listeners engaged temporally selective attention during comprehension and that their attentional engagement was higher during word-initial segments. Thus, the attention to the comprehension stream fluctuates with time. However, it is unclear whether the attention is directed more toward a particular stage of comprehension (e.g., phoneme selection, conceptual selection). The attention probe effect could not be observed using the nonlinguistic attention probes. This might have happened because listeners were asked to pay attention to linguistic stimuli, and thus, they might have ignored the nonlinguistic probes. This would indicate that linguistic probes are better suited for capturing temporally selective attention during comprehension.

Linguistic attention probes were also used to investigate how different linguistic factors affect listeners' temporally selective attention. Li et al. (2014) presented participants with Mandarin Chinese sentences where the critical words could vary in predictability (i.e., expectancy based on a cloze probability test) or accentuation (based on acoustic salience reserved for presenting important or new information). Each sentence was presented twice, once with a linguistic attention probe, which appeared 100 msec after critical word onset, and once without a linguistic attention probe. They found that the peak of the attention probe effect appeared earlier for the unpredictable and accented critical words, meaning that more attentional resources were dedicated to processing these words. In a follow-up study, Li et al. (2017) utilized a similar design to investigate how temporal attention varied when participants listened to Mandarin Chinese phrases. They manipulated whether the target character belonged to the same semantic event as the preceding characters (i.e., the same four-character idiom) or whether it extended across a semantic event boundary (2 two-character words) as well as the accent. The attention probe effect was larger and started earlier when the critical words were unaccented and belonged to the same semantic event (i.e., one idiom) compared with all other conditions. This shows that the accentuation effect on temporal attention interacts with end-of-phrase predictability. From these studies, it is still unclear whether the modulations of temporally selective attention during comprehension by linguistic factors are better reflected in variations of the amplitude or latency of the attention probe effect (Li et al., 2014, 2017). Thus, future studies should focus on amplitude as well as latency. Altogether, these studies show that linguistic factors impact the allocation of temporally selective attention during comprehension.

Would linguistic factors, such as expectancy, also affect temporally selective attention during comprehension in Dutch? It is well established that word expectancy affects language processing. In eye-tracking studies, highly expected words from the context attract fewer and shorter fixations compared with less expected words (Ehrlich & Rayner, 1981). Word expectancy is known to affect the EEG signal around 400 msec after word onset, indicating that word expectancy affects post-lexical semantic processing (e.g., Kutas & Federmeier, 2011). However, listeners can also anticipate upcoming words before their onset (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Altmann & Kamide, 1999), meaning that word expectancy should affect beginning stages of word processing. This is in line with Astheimer and Sanders (2011), who found that, in an artificial language, the N100 elicited by word onsets was larger than the N100 elicited by syllable onsets in the middle of the words, but only when the words were unpredictable. Thus, listeners tend to direct more temporally selective attention to word onsets because word onsets consist of more unpredictable parts of the speech. This means that listeners should also allocate more temporally selective attention to comprehension of unexpected compared with expected words, in line with Li et al. (2014).

## Current Study

In this study, we investigated whether speech planning impacted the allocation of attention during comprehension. We used linguistic attention probes (i.e., syllable "BA" added to the auditory signal) to tap into the distribution of the attentional resources of the listener. We first aimed to show that attention probes are sensitive to word expectancy in the comprehension stream in Dutch. This should establish that attention probes are sensitive to linguistic variables during comprehension and reflect temporal variations in attention toward the comprehension stream. Second, we investigated whether diverting attention from the comprehension stream toward speech planning also affected attention toward comprehension as reflected by the response to attention probes. In this experiment, participants were asked to listen to prerecorded auditory sentences, with expected or unexpected sentence-final words (see Figure 1 for the trial sequence). Each sentence was presented twice, once with and once without the attention probe that was played 100 msec after the target word. Each pair of sentences was presented either in a picture-naming block or a button press block. In both blocks, participants saw a picture, which appeared 50 msec before the target word; it was followed by a gray square and, finally, an exclamation mark. At the onset of the exclamation mark, participants were asked to name the pictures in the picture-naming blocks (i.e., delayed picture-naming condition), or they were asked to press the space bar in the button press blocks (i.e., delayed button press condition).

In the delayed picture-naming task, in the first 270 msec, it is thought that participants first process the picture visually and activate the associated conceptual and lexical information (Indefrey & Levelt, 2004). Thus, these stages should overlap with the onset of the probe, allowing us to investigate the effects of early speech planning on comprehension. Up until approximately 300 msec, speech planning stages (i.e., conceptual, lexical, and even the beginning of phonological retrieval) are thought to take place in the same manner in delayed naming (and covert naming) as in immediate naming tasks (Laganaro & Perret, 2011; Eulitz, Hauk, & Cohen, 2000). This makes this paradigm comparable to how these stages might be scheduled in naturalistic conversation. On the other hand, the delayed button press task keeps participant's engagement

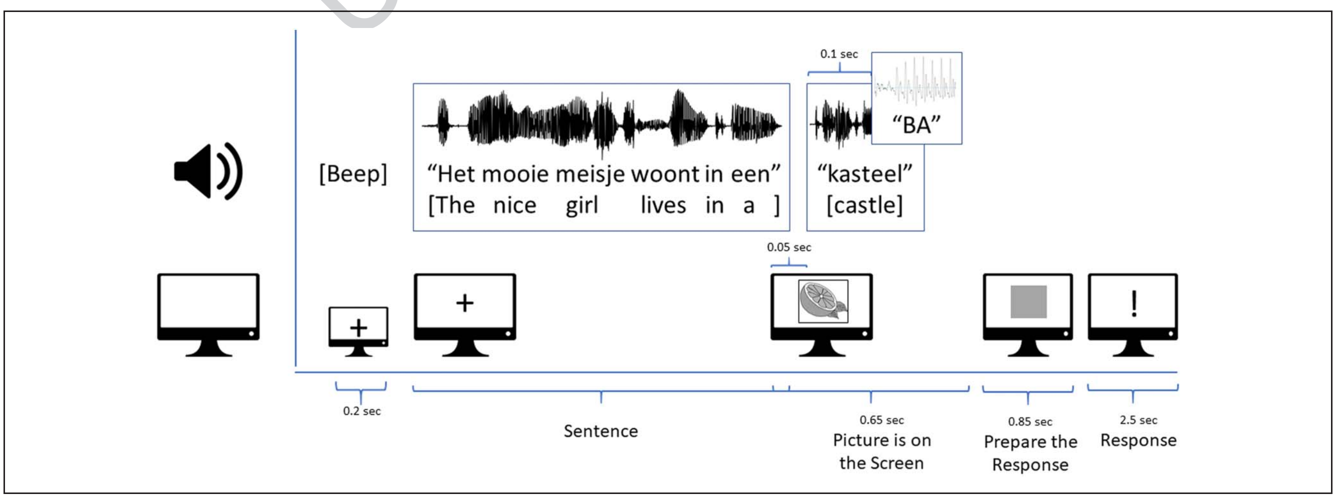


**Figure 1.** Experimental trial sequence of an unexpected probe condition up to the response interval. Note that the linguistic probe "BA" was only present in probe trials. Participants are instructed to name the picture after the onset of the exclamation mark as fast as possible in the picture-naming blocks and to press a spacebar after the onset of the exclamation mark as fast as possible in the button press blocks.

during the response interval but does not require any linguistic processing of the pictures. Note that the time interval between the onset of the picture and the exclamation mark was the same on every trial. Thus, in the delayed button press condition participants could be in a state of motor preparation and stimulus anticipation right before the onset of the exclamation mark.

The linguistic attention probes should elicit the attention probe effect, which measures attention to comprehension at the moment of probe presentation (Li et al., 2014, 2017; Astheimer & Sanders, 2009). Thus, the amplitude and latency of the attention probe effect should indicate how much and when participants dedicate temporally selective attention to comprehension at the onset of the linguistic attention probes. The linguistic attention probes allowed us to investigate how much attention participants allocate to comprehension of expected versus unexpected target words, or how much attention participants allocated to comprehension while engaging in speech planning versus button press readiness. If successful, the paradigm can be used to address a variety of questions investigating how speech planning is scheduled during listening, as the probes can be inserted almost anywhere in spoken utterances. The ultimate goal of this study was to find a method for investigating how and when speech planning takes place in naturalistic conversations. We did not have explicit hypotheses regarding the three-way interaction between probe, expectancy, and task, and our study was not powered to find this effect.

### Hypotheses

1. We expected that the linguistic attention probes would elicit the attention probe effect. This means that the EEG signal should show a negative deflection with a peak between 80 and 280 msec after the probe onset (see Li et al., 2014, 2017; Astheimer & Sanders, 2009) in the probe as compared with the no-probe condition.
2. We anticipated that the expectancy of the target words would affect the attention probe effect and thus when and how much attention would be directed to comprehension.
   a. We expected to replicate a delayed attention probe effect (i.e., most negative component between 80 and 280 msec post probe onset) in the expected compared with the unexpected condition. This is because following an identical context, participants might direct attention to comprehension of an unexpected word faster as comprehending this word might require more processing capacity.
   b. We also expected to find a smaller amplitude (in μV) of the attention probe effect in the expected compared with the unexpected condition. At the moment of target presentation, participants might need to dedicate less attentional resources to the comprehension of expected words, as these might have been partially activated before their onset. This effect was not found in the previous study utilizing Mandarin sentences (Li et al., 2014); however, it might manifest differently in Dutch.
3. We anticipated that speech planning would modulate the attention probe effect and thus when and how much attention is allocated to comprehension.
   a. We expected a delayed attention probe effect (i.e., most negative component between 80 and 280 msec post probe onset) in the picture-naming blocks compared with the button press blocks, as we anticipated that participants might be slower in directing attentional resources to comprehension when they are simultaneously speech planning. This means that the difference between the probe and no-probe conditions should peak earlier in the button press as compared with the picture-naming condition.
   b. Additionally, we expected a smaller amplitude (in μV) of the attention probe effect in the picture-naming blocks compared with the button press blocks, as we anticipated that participants could dedicate less attentional resources to comprehension when they were simultaneously speech planning, which also requires capacity. This means that the difference between the probe and no-probe conditions should be larger in the button press as compared with the picture-naming condition.

## METHODS

The Stage 1 proposal for this report, experimental scripts, analysis scripts, and EEG data of participants who provided consent are available on https://osf.io/s25qm/. The present study was approved by the Ethics Committee of the Social Sciences Department of the Radboud University Nijmegen (ECSW-2019-019).

### Participants

We conducted an a priori power analysis based on the expected ERP amplitude interaction between probe and task using the SIMR package (Green & MacLeod, 2016, Version 1.0.6) in R (Version 4.1.3). The power analysis was based on a linear mixed-effects model fitted with the lmerTest package (Version 3.4; including the lme4 package; Bates, Mächler, Bolker, & Walker, 2015). To determine the minimum number of participants required to find the interaction between probe and task, we first created an empty data frame following our design (e.g., 24 participants, 200 trials, 200 items split over Lists A and B). Subsequently, we created a lmer using following formula: voltage ~ (probe * expectancy) + (probe * task) + (1 | subj) + (1 | target). We used dummy coding with no-probe and button press as the reference conditions. The size of the main effects of probe, expectancy, and task as well as the interactions between probe and expectancy and probe and task were estimated based on the condition means from Li et al. (2014, 2017). Note that even though Li and colleagues used linguistic probes in their design, they did not use a similar task; thus, our participants' estimate can only be taken as a rough guideline. We specified the intercept, slope of probe condition, slope of expectancy condition, slope of task condition, interaction between probe and expectancy, and interaction between probe and task at 0, −2, 0.5, 0, 0, and 1.5, respectively. From a previous unpublished study, we estimated the random intercepts for variance at 8 and 1 for subjects and targets, respectively, and standard deviation for residual variance was set at 8. After reducing the interaction to 80%, the power analysis with 1000 simulations showed that at the alpha level of .05 including 32 participants would give us more than 80% power to detect the interaction between probe and task.

Thirty-eight right-handed native Dutch-speaking participants without hearing impairments took part in the experiment for financial compensation. Data of four participants were excluded because of a technical error, and data of two participants were excluded because of artifacts in the EEG signal. The remaining 32 participants had a mean age of 22.59 years (range: 18–29), and two were male.

### Materials and Design

The comprehension materials were composed of 400 Dutch sentences containing sentence-final target words that either were contextually expected (e.g., *De mooie prinses woont in een kasteel* [The nice princess lives in a castle]) or unexpected (e.g., *Het mooie meisje woont in een kasteel* [The nice girl lives in a castle]), taken from previous studies (see Poulton & Nieuwland, 2021; Klaus, Schutter, & Piai, 2020; Roos & Piai, 2020; Piai, Roelofs, Rommers, & Maris, 2015; Piai, Roelofs, & Maris, 2014), as well as newly created. Note that the last two pretarget words were kept the same across conditions. To examine the expectedness of the target words, we conducted a cloze pretest where 10 participants read all sentences and filled in the last missing target word. The cloze probability was significantly higher for the sentences with the expected ($M$ = 0.90, $SD$ = 0.12, range: 0.6–1) compared with the sentences with the unexpected endings ($M$ = 0.04, $SD$ = 0.07, range: 0–0.3), $t(199) = 92.77$, $p < .001$, $d = 6.56$. The expected and unexpected contexts leading to the same targets were not presented to the same participant but were split over List A and List B. The sentences consist of seven to 11 words, and the target words had an average duration of 572.90 msec ($SD$ = 109.71, min = 369, max = 937). All sentences were prerecorded by a female speaker using 3.0.0 Audacity(R). The average intensity of the sentences was 73.08 ($SD$ = 1.43). A quarter of the sentences were associated with a yes–no comprehension question asking about the target (e.g., *Ging de zin over een kasteel?* [Was the sentence about a castle?]). The linguistic attention probe was constructed by creating a 50-msec excerpt from the spoken sentences when the speaker pronounced syllable "BA." The intensity of the probe was adjusted to 76 dB to match the loudest parts of the comprehension stimuli (Astheimer & Sanders, 2009). Each sentence was played twice, once with and once without the probe. The probes were played 100 msec after the target word onset.

Each sentence was paired with one of 12 target pictures depicting one of six fruits or one of six vegetables. All sentences were unrelated to the theme of food depicted in the target pictures. The target pictures were grayscale drawings from the MultiPic database (Duñabeitia et al., 2018; for the complete list, see Supplement A). The mean frequency of the picture names is 3.17 per million words (range: 1.55–5.33; Decuyper, Brysbaert, Brodeur, & Meyer, 2021). Each picture was displayed 20 times throughout the experiment. Using a homogeneous group of repeating pictures lead to comparable response times with little variation in early planning processes.

We used a two-by-two-by-two design with the blocked variable, task (picture naming vs. button press), and with a quasi-random within-block variables, probe (probe vs. no probe) and expectancy (expected vs. unexpected). We utilized 50 trials per condition. Expected and unexpected contexts associated with the same targets were distributed over two lists (Lists A and B), which were presented to different participants. We pseudorandomized the items with MIX (van Casteren & Davis, 2006), creating eight unique lists, four utilizing items from List A and four utilizing items from List B. The task condition (i.e., picture naming or button press) was blocked, meaning that 20 trials of the same condition type followed one another. The block type alternated throughout the experiment (e.g., Block 1: picture naming, Block 2: button press, Block 3: picture naming, …., Block 10: button press). The same probe condition (i.e., probe or no probe) and same expectancy condition (i.e., expected or unexpected) appeared on maximally four consecutive trials. There were maximally four sentences without a comprehension question presented after one another. Each picture was shown 20 times, with at least three pictures intervening between repetitions. The two tokens of the same sentence were shown with at least 20 intervening sentences (i.e., one block). For each randomized list, we switched the probe and no-probe conditions, creating eight new lists. For these 16 lists, we subsequently switched whether the participants started with the picture-naming condition or the button press condition, creating 16 new lists. Thus, we had 32 unique lists, one for each participant.

### EEG Acquisition

EEG was recorded using actiCap system (Brain Products) using 32-scalp electrodes arranged according to the 10–20 system. Twenty-five electrodes (F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, Oz) were positioned on the scalp. Six electrodes were positioned on the face to capture eye movements (two electrodes above and below the left eye and two electrodes on the left and right temples) and speech artifacts (two electrodes above and below the right side of the lips on the orbicularis oris muscle). The left mastoid was used as online reference, and one remaining electrode was placed on the right mastoid, so that an average mastoid reference could be computed. We used a 500-Hz sampling rate with a band-pass filter of 0.016 to 150 Hz for the recording. The impedances were adjusted below 10 kΩ for all electrodes.

### Procedure

We fitted participants with an EEG cap. Subsequently, they were familiarized with the pictures, by looking at each picture and reading its associated name. The experiment was run in Presentation software (Neurobehavioral Systems). Each trial started with a beep sound presented simultaneously with a fixation cross (200 msec). Subsequently, the comprehension sentence with expected and unexpected targets was played. On the probe trials, the linguistic probe (i.e., "BA") appeared 100 msec after the target word onset, so there was an overlap between the probe and target words. The pictures appeared in the center of the screen 50 msec before the onset of the target words and remained on the screen for 650 msec. The picture was replaced by a gray square presented for 850 msec. Subsequently, an exclamation mark appeared for 2500 msec. Participants were instructed to name the pictures or to press a space bar as fast as possible after the onset of the exclamation mark on the picture-naming and button press blocks, respectively. On the trials that included a comprehension question, participants pressed "C" (i.e., yes) or "N" (i.e., no), depending on whether the comprehension question correctly referred to the previously presented target word. Afterward a blinking interval was presented with duration jittered between 1200 and 1500 msec. Trials were separated into 10 blocks, each lasting about 4.5 min, and participants were able to take a break after each block.

### Behavioral Responses

For the planning blocks, each response was scored from participants' recordings as either correct (i.e., correctly named picture within 2500 msec) or incorrect (i.e., incorrectly named picture, omission, hesitation, or picture named outside the time window). For the button press blocks, Presentation scored the responses as correct (i.e., space bar was pressed within 2500 msec) or incorrect (i.e., space bar was not pressed within 2500 msec). Naming latencies as well as RTs for the button presses were automatically recorded by Presentation. Because of unforeseen microphone malfunction, the naming responses were recorded, but their onset was not accurately logged by Presentation. Thus, we annotated the naming onsets in Praat (Boersma & Weenink, 2023) from the recordings using an automated procedure. Only trials scored as correct were included in the RT and all EEG analyses.

### Preprocessing

BrainVision Analyzer (Version 2.2.0) was used for preprocessing. We removed noisy channels and reconstructed these channels using Spherical Spline interpolation. No more than five channels were excluded per participant. We used a band-pass filter between 0.1 and 30 Hz and a notch filter at 50 Hz, using 24 dB/octave roll-off. Subsequently, we re-referenced all channels to an average mastoid reference. The data were segmented with respect to the target word onset. The segments started from 200 msec before to 600 msec post target word onset. Note that the pictures appeared at 50 msec before the target, and the probes (if present) appeared at 100 msec after the target. A baseline correction was performed using the first 150-msec window of the segment. Trials with artificial noise (e.g., steep jumps in the amplitude) were excluded before running the independent component analysis (ICA). Restricted fast ICA implemented in the BrainVision Analyzer was performed to correct for blinks, eye movements, and steady muscle activity. The remaining physiological artifacts were removed using a semiautomatic procedure with a threshold of ±60 μV. A minimum of 40 trials were included per participant and per condition.

### Analysis

#### *Peak Latency Analysis*

We examined peak latency differences to investigate if the attention probe effect was delayed in the expected compared with the unexpected condition as well as in the picture-naming blocks compared with the button press blocks. It was essential to examine peak latency before examining amplitude differences, as finding effects in peak latency could have affected the amplitude analysis. One of the most frequent approaches to analyzing the peak latencies is to estimate the peak latency based on a given time interval per participant per condition (e.g., Dimitrijevic & Stapells, 2006; Sable, Low, Maclin, Fabiani, & Gratton, 2004). However, following Li et al. (2014), we performed peak latency analysis on the difference waves. We chose this approach because by subtracting activation elicited by sentences without probes from sentences with probes, we could account for the baseline activation of sentences and theoretically measure only the effect of the linguistic attention probes, which indicates the level of attention to comprehension at the probe presentation.

For this analysis, we specified an ROI from 180 to 380 msec post target onset (i.e., from 80 to 280 msec post probe onset) for centrally located electrodes (i.e., F3, Fz, F4, FC1, FCz, FC2, C3, Cz, C4, CP1, CP2, P3, Pz, P4). The ROI was chosen based on previous studies (Li et al., 2014, 2017; Astheimer & Sanders, 2009). The time window of interest was difficult to estimate from previous studies, as the peak of the attention probe effect seemed to vary widely, from 100 msec (Astheimer & Sanders, 2009) to 190 msec (Li et al., 2014). We anticipated that the peak of the attention probe effect in this study would more closely resemble the later attention probe effect. To capture potential individual differences in the attention probe peak, we chose to analyze a longer time window for the latency analysis.

Per participant, we first computed averages per condition. Subsequently, we generated difference waves between the probe and no-probe conditions separately for the button press expected, button press unexpected, picture-naming expected, and picture-naming unexpected conditions. We used peak detection to find the most negative peak in the ROI, for which we recorded the latency. We computed Bayesian linear mixed-effects model using the "brms" (Bürkner, 2017) package in R (Version 2022.07.2). We used a mixed-effects model with latency of the difference waves as the dependent variable, task and expectancy as main effects, and random intercept per subject (latency ~ task + expectancy + (1 | subj)). We used default priors for all parameters except for the main effects of task and expectancy. Following Dienes (2014), we used a prior centered on a mean of 0 and an *SD* of 30, as 30 msec was the approximate previous reported effect size for a similar effects (Li et al., 2014). For the main effects, we reported beta and the 95% credible interval and the Bayes factor evidence in support of the null hypothesis ($BF_{null}$). The credible interval indicates a 95% probability that the population estimate would lie within the interval, given the observed data. Thus, when comparing two groups, credible intervals that do not include 0 indicate that there is 95% likelihood that the population means are different from one another, given the observed data. The Bayes factor corresponds to the ratio between posterior density and prior density at 0. Values higher than 1 indicate a proportional increase in confidence in the null hypothesis, such that $BF_{null} = 3$ indicates a threefold increase of the evidence for the null hypothesis and $BF_{null} = 0.3$ indicates a threefold increase of the evidence for the alternative hypothesis. In this study, we considered $BF_{null}$ to designate the strength of the BF evidence: $BF_{null}$ between 1 and 3, $BF_{null}$ between 3 and 10, and $BF_{null}$ between 10 and 30 were considered as anecdotal, moderate, and strong evidence for the null hypothesis, respectively (Jeffreys, 1998).

### *ERP Amplitude Analysis*

We examined differences in the amplitude of the attention probe effect to investigate whether it was reduced in the expected compared with unexpected condition as well as in the picture-naming blocks compared with the button press blocks. We used an ROI with the same electrodes as in the latency analysis (see Peak Latency Analysis section). We specified a time window for the ROI per condition based on the peak latency analysis to account for possible latency effects. Per condition, we computed the mean latency of the attention probe effect and analyzed 100-msec time window around the peaks in each condition. To compare the amplitude between conditions, we computed mean voltage per each time window.

In a Bayesian linear mixed-effects model, we used voltages from the ROI as dependent variable. The model included the main effects of task, expectancy, and probe, as well as interactions between task and probe and expectancy and probe, and random intercepts for subjects and target words (voltage ~ (task * probe) + (expectancy * probe) + (1 | subj) + (1 | target)). We used default priors for all parameters except for the two interactions. Following Dienes (2014), for these interactions, we used a prior centered on a mean of 0 and an *SD* of 1, as 1 μV was the approximate previous reported effect size for a similar interaction (Li et al., 2017). Hence, there is a 95% prior probability that the population parameter lies between −1 and 1 μV. Note that a previous study did not find interaction between expectancy and probe on voltages (Li et al., 2014); however, we hypothesized to find this effect in this conceptual replication in Dutch. For this analysis, we again considered $BF_{null}$ to designate the strength of the BF evidence: $BF_{null}$ between 1 and 3, $BF_{null}$ between 3 and 10, and $BF_{null}$ between 10 and 30 were considered as anecdotal, moderate, and strong evidence for the null hypothesis, respectively (Jeffreys, 1998).

### *RT Analysis*

To examine the effects of speech planning, expectancy, and probe on RTs, we ran a Bayesian linear mixed-effects model with the same predictors as in the model examining ERP amplitude, but including RTs as dependent variable (RTs ~ (task * probe) + (expectancy * probe) + (1 | subj) + (1 | target)). We used the default priors for all parameters.

## RESULTS

### Accuracy and RTs

Overall accuracy was 95.89% (for accuracies of individual conditions, see Table 1). Incorrect trials were excluded from all subsequent analyses.

For the RT analysis, we also excluded responses where the RTs were 0 (indicating that naming latency was not estimated properly by Praat or that participants might have started speaking too early) or when there was no recorded response in the button press condition (indicating that participant might have failed to press a button or pressed a wrong button). The RTs for individual conditions are presented in Table 1. The model estimated that the RTs were

**Table 1.** Accuracy and RTs

| *Condition* | *Accuracy (%)* | *Accuracy (SD)* | *RT (Mean) (msec)* | *RT (SD) (msec)* |
|---|---|---|---|---|
| Expected picture naming with probe | 95.45 | 0.2085 | 406.77 | 192.01 |
| Expected picture naming without probe | 95.64 | 0.2043 | 424.32 | 189.32 |
| Expected button press with probe | 96.11 | 0.1933 | 368.32 | 208.05 |
| Expected button press without probe | 96.26 | 0.1898 | 386.84 | 220.07 |
| Unexpected picture naming with probe | 96.03 | 0.1954 | 417.27 | 205.04 |
| Unexpected picture naming without probe | 95.47 | 0.2081 | 445.38 | 221.93 |
| Unexpected button press with probe | 96.03 | 0.1954 | 371.65 | 199.58 |
| Unexpected button press without probe | 96.16 | 0.1923 | 391.27 | 218.46 |

This table presents accuracy in percentages and RTs in milliseconds, together with their respective standard deviations for each condition.

slower in the no-probe condition ($M$ = 411 msec, $SD$ = 214) compared with the probe condition ($M$ = 390 msec, $SD$ = 202; β = −0.05, *CrI* [−0.08, −0.03]). Button presses (i.e., RTs in the button press condition; $M$ = 380 msec, $SD$ = 212) were estimated to be faster than naming latencies (i.e., RTs in the picture-naming condition; $M$ = 423 msec, $SD$ = 203; β = 0.16, *CrI* [0.14, 0.18]). The credible interval for the expectancy effect on RTs had a zero boundary (β = 0.02, *CrI* [0.00, 0.05]), indicating little evidence for a meaningful difference between expected ($M$ = 396 msec, $SD$ = 204) and unexpected ($M$ = 406 msec, $SD$ = 213) conditions on RTs. We did not obtain evidence for interactions between probe and task conditions (β = 0.01, *CrI* [−0.02, 0.04]) or between probe and expectancy conditions (β = −0.00, *CrI* [−0.03, 0.03]) on the RTs.



## Latency Analysis

In the latency analysis, we examined whether the peak of the difference waves (probe – no probe) depended on the expectancy or task (see Figure 2, showing difference waves, and Figure 3, showing the ERPs for each condition; Supplements B and C show the same information respectively for all electrodes). It can be seen from the plot that the attention probe effect for the selected ROI peaked marginally later in the expected ($M$ = 244, $SD$ = 20) compared with the unexpected condition ($M$ = 240, $SD$ = 19).

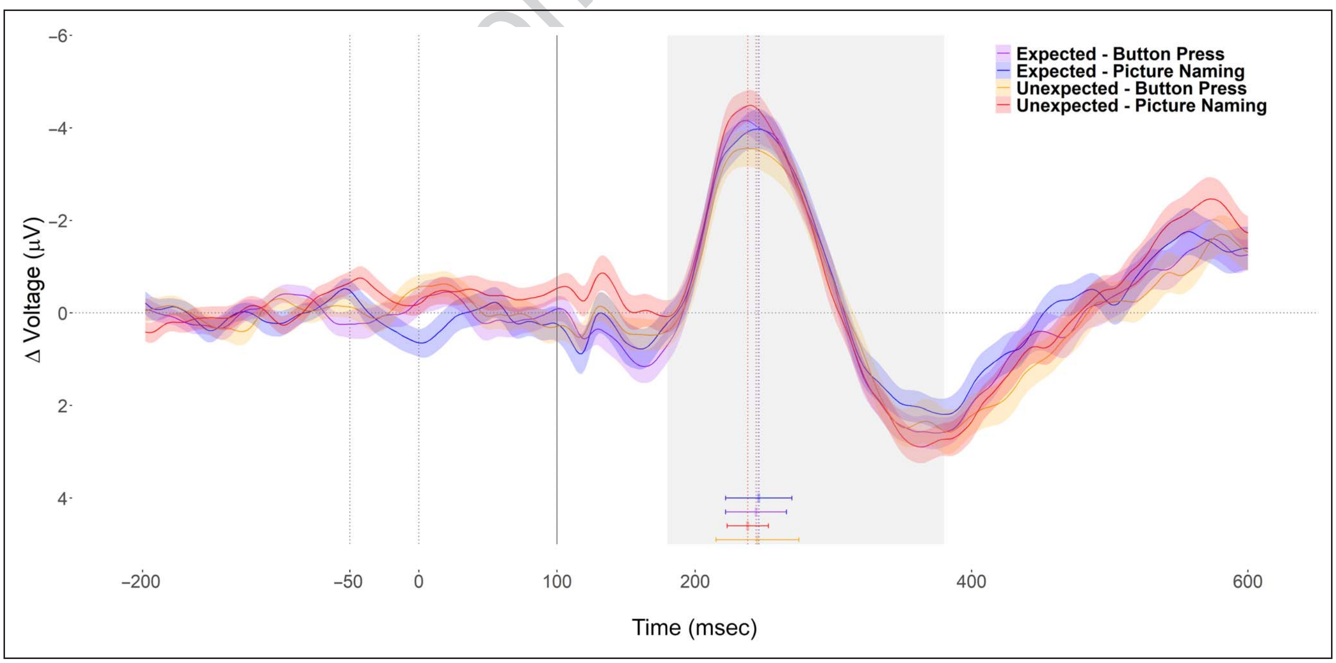


**Figure 2.** Difference waves between the probe and no-probe conditions depicted for the expected button press (purple), expected picture-naming (blue), unexpected button press (yellow), and unexpected picture-naming (red) conditions. The gray-shaded area represents the time window for which we performed the latency analysis. Vertical lines of the respective color represent the mean latency peak per condition. In this plot, 0 is the onset of the target word, −50 is the onset of the picture, and 100 is the onset of the probe (in the probe conditions).

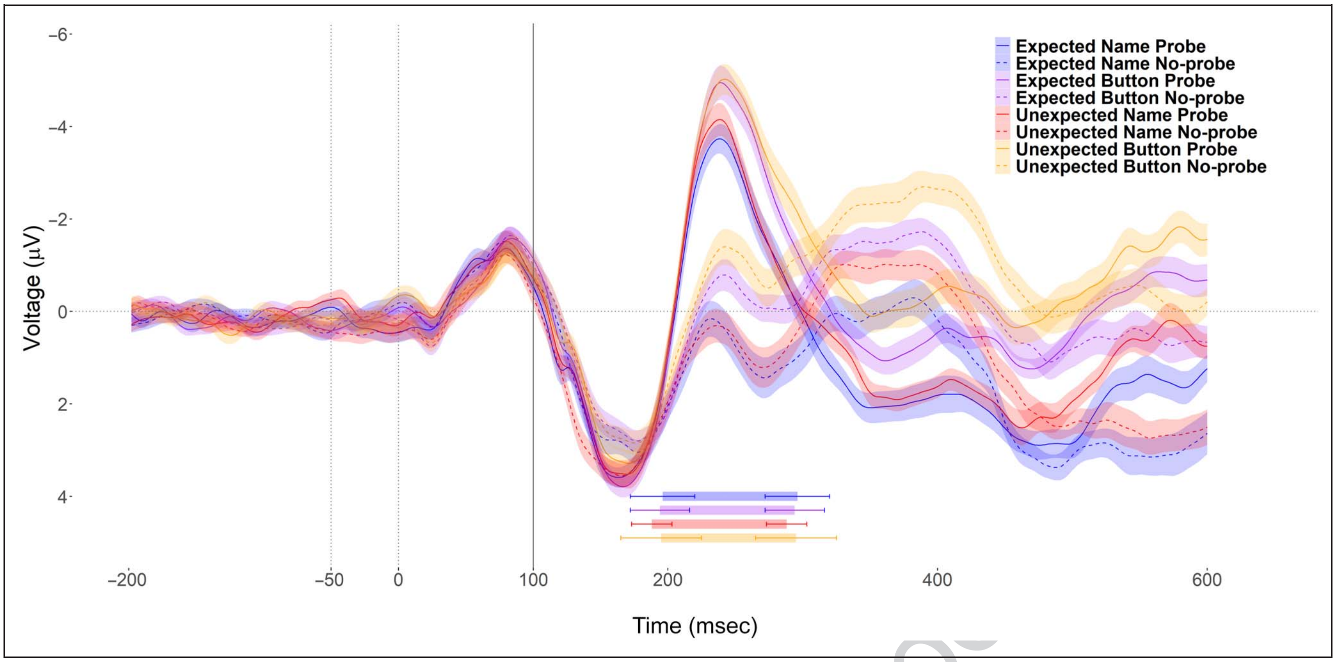


**Figure 3.** The expected picture-naming (blue), expected button press (purple), unexpected picture-naming (red), and unexpected button press (yellow) conditions, as well as the probe (solid) and no-probe (dashed) conditions (in μV), plotted for the centrally located electrodes. In this plot, 0 is the onset of the target word, −50 is the onset of the picture, and 100 is the onset of the probe (in the probe conditions). The horizontal bars of respective colors represent the mean onset and offset of the analysis windows, which varied per participant.

However, the analysis showed moderate support for the null hypothesis (β = 3.74, *CrI* [−1.71, 9.22], $BF_{null}$ = 4.52). The credible interval indicated that the population means do not differ from one another and the Bayes factor indicates that the null hypothesis is 4.52 more likely than the alternative. Thus, with these results, we failed to replicate Li et al.'s (2014) effect of expectancy on latency. Our results provide evidence that expectancy does not affect the latency of the attention probe effect.

Following the main analysis, we performed a sensitivity analysis to assess the effects further based on the chosen priors. We obtained moderate evidence for the null hypothesis when using a narrower prior (*SD* = 20; β = 3.69, *CrI* [−1.78, 9.18], $BF_{null}$ = 3.1), as well as a wider prior (*SD* = 40; β = 3.74, *CrI* [−1.83, 9.30], $BF_{null}$ = 6.05).

The latency of the attention probe effect did not seem to differ depending on the task (picture naming: *M* = 241, *SD* = 17; button press: *M* = 243, *SD* = 22; β = −1.76, *CrI* [−7.33, 3.74], $BF_{null}$ = 9.1). The credible interval indicated that the population means did not differ from each other, and the Bayes factor indicates that the null hypothesis was 9.1 more likely than the alternative, thus providing moderate evidence for the null hypothesis.

The sensitivity analysis with both narrower and wider priors supports these conclusions. Using a narrower prior (*SD* = 20) yielded moderate evidence for the null hypothesis (narrower prior: β = −1.75, *CrI* [−7.33, 3.79], $BF_{null}$ = 6.03), whereas using a wider prior (*SD* = 40) provided strong evidence for the null hypothesis (β = −1.73, *CrI* [−7.30, 3.87], $BF_{null}$ = 11.77).

## Amplitude Analysis

We utilized the peaks per condition and per participant from the latency analysis in the amplitude analysis. In this analysis, we analyzed the amplitude of 100-msec-long intervals with the center at each of these individual peaks. We examined the attention probe negativity using voltages from the ROI as the dependent variable (Figure 3; for all electrodes, see Supplement C). As expected, there was a stronger negative potential in the probe condition as compared with the no-probe condition (β = −2.62, *CrI* [−3.12, –2.13]). There was a slow positive potential in the picture-naming condition as compared with the button press condition (β = 1.03, *CrI* [0.63, 1.44]). We did not find any credible amplitude differences based on expectancy (β = 0.38, *CrI* [−0.03, 0.78]). The interaction between probe and expectancy on the amplitude included zero, providing no credible evidence for the interaction. Consistent with this result, the Bayes factor indicated moderate evidence in support of the null hypothesis (β = −0.08, *CrI* [−0.64, 0.48], $BF_{null}$ = 3.36). Similarly, the interaction between probe and task on the amplitude also included zero, providing no credible evidence for the interaction. Consistent with this, the Bayes factor indicated anecdotal evidence in support of the null hypothesis (β = −0.26, *CrI* [−0.82, 0.30], $BF_{null}$ = 2.38).

The sensitivity analysis with the narrower priors (*SD* = 0.5) supported these conclusions, showing anecdotal evidence against the presence of an interaction between probe and expectancy (β = −0.07, *CrI* [−0.58, 0.43], $BF_{null}$ = 1.82) and an interaction between probe and task

($\beta = -0.21$, *CrI* [−0.71, 0.31], $BF_{null} = 1.38$). When using wider priors (*SD* = 1.5), the model showed moderate evidence against an interaction between probe and expectancy ($\beta = -0.08$, *CrI* [−0.66, 0.49], $BF_{null} = 4.91$) and an interaction between probe and task ($\beta = -0.27$, *CrI* [−0.86, 0.31], $BF_{null} = 3.31$).

Considering that the naming latencies varied widely, we ran an exploratory analysis where we excluded all trials where response latencies deviated by 3 *SD*s or more from the condition mean (1.67% of the data were excluded from the data used for the preregistered analysis). Thus, in this analysis, we repeated the main amplitude analysis but with fewer trials. The results did not change any of the conclusions drawn from the main analysis. We found main effects of probe ($\beta = -2.65$, *CrI* [−3.14, −2.15]) and task ($\beta = 1.08$, *CrI* [0.67, 1.49]), while no substantial evidence for the effect of expectancy was observed ($\beta = 0.41$, *CrI* [0.00, 0.83]). We found moderate evidence against an interaction between probe and expectancy ($\beta = -0.08$, *CrI* [−0.65, 0.49], $BF_{null} = 3.37$). There was anecdotal evidence against an interaction between probe and task ($\beta = -0.28$, *CrI* [−0.85, 0.29], $BF_{null} = 2.15$).

### Exploratory N400 Analysis

Since the attention probe effect showed no modulation by expectancy, we turned to the N400 time window to verify the effectiveness of the predictability manipulation, given its established link to expectancy effects (Kutas & Federmeier, 2011). We computed a Bayesian linear mixed-effects model using amplitude from centro-parietal (i.e., C3, Cz, C4, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8) electrodes and from the time window between 300 and 500 msec as dependent variable. The model included the main effects of task, expectancy, and probe, as well as the two-way interactions between all variables and random intercepts for subjects and target words (voltage ~ (expectancy * task) + (task * probe) + (expectancy * probe) + (1 | subj) + (1 | target)). We used default priors for all parameters.

The analysis of the N400 ROI (see the figure in Supplement E) showed that there was a stronger negative potential in the unexpected compared with expected condition ($\beta = 1.34$, *CrI* [0.78, 1.71]). There was a stronger positive potential in the probe condition as compared with the no-probe condition ($\beta = 1.54$, *CrI* [1.08, 1.99]), and there was a larger positive potential in the picture-naming condition as compared with the button press condition ($\beta = 2.15$, *CrI* [1.69, 2.62]). We did not find any credible amplitude differences for any of the interactions (Expectancy × Task: $\beta = -0.42$, *CrI* [−0.95, 0.11]; Task × Probe: $\beta = -0.06$, *CrI* [−0.59, 0.46]; Expectancy × Probe: $\beta = -0.36$, *CrI* [−0.90, 0.16]). Thus, the predictability manipulation was effective, and probe presence and task type independently influenced the amplitude in the N400 time window, without evidence for interactive effects. This suggests that differences in task demands or the presence of the probe do not modulate the semantic processes indexed by the N400.

## DISCUSSION

The current study examined the attention probe effect, defined as the difference between the signal elicited by a comprehension stimulus presented with and without a linguistic attention probe (the syllable "BA" presented during spoken sentences). The effect is believed to reflect attention to comprehension at the time the probe is presented (Li et al., 2014, 2017; Astheimer & Sanders, 2009). Previous studies have suggested that the amplitude and latency of the attention probe effect indicate how much and when participants dedicate temporally selective attention to comprehension at the presentation of the attention probes. In this sense, the attention probe serves as a "sonar" that pings the brain, and the amplitude and latency of the attention probe effect are the readouts that carry the information about the underlying comprehension processes taking place just as the brain was pinged.

The goals of the current study were twofold. First, we aimed to replicate the previous finding that word expectancy influences the allocation of temporally selective attention during comprehension (Li et al., 2014). Second, we investigated whether engaging in speech planning during comprehension affected this attentional allocation. We observed a strong attention probe effect across all probe conditions. However, neither word expectancy nor speech planning influenced the timing or strength of the attention probe effect.

### Comprehension Engagement

Our results revealed a clear attention probe effect, which peaked 142 msec after probe onset, and the mean amplitude of this difference wave was approximately −4 μV. Previous studies found large variability in the mean onset of the attention probe effect, with the mean varying between 120 msec (Astheimer & Sanders, 2009) and 180 msec (Li et al., 2014, 2017). Our effect lies in between the estimates from previous studies. Compared with previous studies, where the attention probe effect had an amplitude of approximately −2 μV (Li et al., 2014, 2017), our study found a larger effect, with a mean amplitude of around −4 μV. This suggests that the probes in our study elicited a stronger neural response. This difference may be attributed to factors such as experimental design, including the variation in comprehension sentences or tasks. Other possibilities include differences in the stimuli, such as probe perceptibility, or participant samples. Overall, these results suggest that the attention probe elicited a strong, reliable response, indicating that participants were effectively engaging with the comprehension task at the moment when the probe was presented.

The analysis of naming latencies showed that participants' responses were, on average, 450 msec (or less) in every condition. This is notably shorter than the typical time required to name an object, which is around 600 msec (e.g., Indefrey & Levelt, 2004). Moreover, the analyses of the naming latencies showed no effect of expectancy. Both of these findings are consistent with the assumption that participants prepared for naming before the cue, as they had been instructed to do.

### Expectancy Effects

The analysis provided moderate evidence against a difference in the latency and amplitude of the attention probe effect between the expected and unexpected conditions, failing to replicate previous findings by Li et al. (2014). Thus, our study suggests that expectancy does not influence temporal attention to comprehension, 100 msec after word onset. This means that participants do not allocate more or less attention to comprehension depending on whether or not the word was expected. However, we found a credible difference between the expected and unexpected conditions in the N400 time frame, as in an earlier study utilizing the same materials (Hustá, Nieuwland, & Meyer, 2023). The N400 effect has been linked to semantic prediction in comprehension (Grisoni, Tomasello, & Pulvermüller, 2021; Rabovsky, Hansen, & McClelland, 2018; Rabovsky & McRae, 2014), suggesting that the predictability manipulation did indeed work but simply did not show the expected attention probe effect. The lack of any expectancy effect on the attention probe effect is surprising given strong evidence that listeners do predict upcoming words, both in Dutch and in Mandarin (Kwon, Sturt, & Liu, 2017; Van Berkum et al., 2005). In fact, previous research has shown that listeners can anticipate upcoming words even before they are spoken, which implies that expectancy should affect early stages of word processing, especially how much attention participants decide to allocate to the comprehension of expected and unexpected words.

However, Li et al. (2014) are, to our knowledge, the only group that has utilized the attention probes to capture how expectancy affects early attention to comprehension. Thus, such early effects of expectancy might not be very strong and are in need for further replication. Additionally, it is not clear what factors might be relevant when utilizing the linguistic attention probes. There are several reasons why we might have encountered different expectancy results compared with Li and colleagues. These will be explored in the remainder of this section.

First, the two experiments differed in several design aspects, which may have influenced the results. The studies used different comprehension sentences with distinct cloze probability distributions. Specifically, this study used sentences with cloze probability values at the extremes of the distribution. However, this should lead to a stronger effect of expectancy. Moreover, the current study presented comprehension questions on only a quarter of the trials, which may have diverted participants' attention away from the comprehension task. Lastly, previous studies that utilized linguistic attention probes to examine expectancy or phrase boundaries also manipulated whether or not the critical word, where probe appeared, carried pitch accent. As pitch accent draws attention to the word, this may have affected the response to probe. By contrast, in our study, the sentences were read with natural intonation, without any specific emphasis on the sentence-final words.

Second, the discrepancy of our results and those obtained by Li and colleagues may be related to properties of the language studied, Mandarin Chinese and Dutch. Mandarin is a tonal language, meaning that pitch variations can change word meanings. As a result, listeners must track both phonemes and pitch contours simultaneously to access word meaning. When a word is expected, comprehension becomes easier because listeners can rely more on contextual cues and less on fine-grained tonal distinctions. In addition, Mandarin has a high degree of homophony due to its limited number of syllables, whereas Dutch exhibits more phonetic variation. In Mandarin, word meaning is often determined by context rather than phonological uniqueness (Li & Yip, 1996). Thus, expectancy might play a bigger role in determining how much attention listeners dedicate to comprehension in Mandarin than in Dutch.

Third, prior studies used linguistic attention probes without additional overlapping tasks (Li et al., 2014, 2017; Astheimer & Sanders, 2009). In contrast, our study included additional task demands, as participants were always presented with a picture on the screen and they prepared for picture naming or a button press task simultaneously with comprehension. These tasks may have diverted listeners' attention away from word anticipation, potentially explaining why expectancy did not modulate the attention probe effect.

In summary, the lack of an expectancy effect in our study may stem from several methodological and linguistic factors. These include differences in experimental design, language properties, and additional task demands. Our results highlight the need for further exploration of factors that affect attentional allocation to comprehension in dual-task paradigms, particularly in cross-linguistic research.

### Task Effects

We found moderate evidence against a difference in latency and anecdotal evidence against a difference in amplitude of the attention probe effect between the picture-naming and button press conditions. This suggests that speech planning does not divert attentional resources away from comprehension at the moment of probe presentation more than preparation for a simple motor task does.

Our findings might indicate that participants can engage in both speech planning and comprehension simultaneously without reductions of attention to early comprehension. This aligns with previous research, showing that speech planning and comprehension can occur in parallel during conversation (He, Meyer, & Brehm, 2021; Bögels et al., 2015, 2018; Sjerps & Meyer, 2015; Boiteau et al., 2014). This body of research has shown that speech planning requires some, but not all of a speaker's attentional capacity. However, our results contrast with studies reporting effects of speech planning on auditory processing of syllables (Fargier & Laganaro, 2019; Daliri & Max, 2016). One possible explanation for this discrepancy is that word comprehension may be a more naturalistic process than syllable processing. In real-life communication, word or utterance planning often co-occurs with speech planning, and the concurrent execution of these processes could reduce the interference between them due to the frequent practice. Thus, the more holistic word comprehension may allow for better parallel processing with speech planning, while syllable processing may require more focused attention and thus be more susceptible to interference.

This study utilized a small set of pictures drawn from the same semantic category with homogeneous word frequency and visual characteristics. This was done to ensure that speech planning would take place in a similar manner across trials. Utilizing this set of pictures could have made parts of the speech planning task more automatic (e.g., lower visual processing demands, lower the activation threshold to initiate response planning; but see Belke, Shao, & Meyer, 2017). This means that the overall attentional demands might have been too low to find the trade-off between comprehension and speech planning. In natural conversations, speakers often plan longer, more complex utterances, which might impose a greater cognitive load, leading to stronger interference between speech planning and comprehension. Future studies could explore whether sentence-level planning imposes a greater attentional cost than word planning, thereby affecting temporal attention to comprehension more robustly.

It should be noted that participants saw the same pictures in both the picture-naming condition as well as the button press condition. This design ensured that condition differences could not be attributed to differences in visual input. We thought it was unlikely that participants would spontaneously retrieve the picture names in inner speech in the button press condition, particularly since a blocked design was used in this study. However, we cannot verify that this was the case. It is possible that the participants internally named the pictures and so processed them in similar ways in the button press and naming conditions. This would explain why the response to the attention probe did not differ across the two tasks.

In summary, we cannot conclusively determine whether the lack of a task effect resulted from participants effectively balancing the speech planning and comprehension tasks or from participants engaging with the picture in the button press task. However, the latter explanation seems unlikely, as the study employed a blocked design, which ensured that the pictures were not task-relevant in the button press block.

## Methodological Considerations

In this study, we explored a novel approach for examining when speech planning takes place during comprehension by using linguistic attention probes. Such an approach is needed to capture the dynamic allocation of attention in real time, as comprehension and speech planning unfold simultaneously in natural conversations. The probes are thought to reveal the attentional resources already dedicated to comprehension when the probe is presented. We hypothesized that the attention probes could pinpoint when speech planning is taking place during comprehension by detecting fluctuations in temporal attention at specific moments in conversation, as attention to comprehension should be decreased when the capacity demanding speech planning is taking place. The advantage of such an approach would be that the probes could "ping the brain" almost anytime during the comprehension utterance, revealing how much attention is dedicated to comprehension at the moment of probe presentation.

Although we observed a clear attentional probe effect, we did not find any significant modulation under the specific conditions tested. This suggests that the current version of the paradigm was not optimal for examining the attentional trade-off between comprehension and speech planning. Modifications of the experimental design, such as incorporating a more challenging comprehension task (e.g., asking participants in-depth comprehension questions about the content of the whole sentence) or introducing a harder speech planning task (e.g., by using a larger, more variable set of pictures with fewer repetitions or by eliciting more complex utterances), may very well yield different results. Such changes would increase the attentional load of the comprehension or/and production task and may provide evidence about the attentional trade-off between comprehension and speech planning in more demanding circumstances.

## Conclusion

In summary, we found evidence that expectancy and speech planning do not influence the latency or amplitude of the attention probe effect. This suggests that early attention to comprehension (i.e., 100 msec after word onset) is not delayed or modulated by expectancy of the comprehension material or the simultaneous occurrence of speech planning. However, it is important to note that in both conditions, speakers saw a picture and prepared for a response after the sentence offset. Therefore, the control button press condition may have also imposed an attentional load, potentially diverting resources away from comprehension. However, we consider this explanation unlikely given the use of a block design, which ensures that participants do not have to frequently switch between tasks. Thus, our findings likely imply that listeners are able to effectively carry out comprehension and speech-planning processes at the same time when planning simple words. While these results do not support our initial hypotheses, they provide valuable insights into the relationship between production and comprehension processes. Future research should consider using more complex linguistic tasks or alternative methods for measuring temporal attention to further clarify these effects.


### Acknowledgments

We thank Mante Nieuwland for comments on an earlier version of this manuscript, Ture Berg for assisting with the analysis of the N400 component, and Birgit Knudsen for performing the data collection.



Corresponding author: Cecília Hustá, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands, e-mail: cecilia.husta@mpi.nl.


### Data Availability Statement

Upon in-principle acceptance, the Stage 1 manuscript will be placed on the OSF together with the (to be collected) data and the already included analysis scripts (https://osf.io/s25qm/). Supplemental materials can be accessed on this article's homepage: https://doi.org/10.1162/jocn.a.97.


### Funding Information

This study was supported by Max-Planck-Gesellschaft (https://dx.doi.org/10.13039/501100004189).


### Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

## REFERENCES


Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264. https://doi.org/10.1016/s0010-0277(99)00059-1, PubMed: 10585516

Astheimer, L. B., & Sanders, L. D. (2009). Listeners modulate temporally selective attention during natural speech processing. *Biological Psychology*, *80*, 23–34. https://doi.org/10.1016/j.biopsycho.2008.01.015, PubMed: 18395316

Astheimer, L. B., & Sanders, L. D. (2011). Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia*, *49*, 3512–3516. https://doi.org/10.1016/j.neuropsychologia.2011.08.014, PubMed: 21875609

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–18. https://doi.org/10.18637/jss.v067.i01

Belke, E., Shao, Z., & Meyer, A. S. (2017). Strategic origins of early semantic facilitation in the blocked-cyclic naming paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*, 1659–1668. https://doi.org/10.1037/xlm0000399, PubMed: 28383951

Bertelson, P. (1967). The time course of preparation. *Quarterly Journal of Experimental Psychology*, *19*, 272–279. https://doi.org/10.1080/14640746708400102, PubMed: 6074169

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* (Version 6.1.16.) [Computer software]. www.praat.org

Bögels, S., Casillas, M., & Levinson, S. C. (2018). Planning versus comprehension in turn-taking: Fast responders show reduced anticipatory processing of the question. *Neuropsychologia*, *109*, 295–310. https://doi.org/10.1016/j.neuropsychologia.2017.12.028, PubMed: 29269305

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. https://doi.org/10.1038/srep12881, PubMed: 26242909

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*, 295–311. https://doi.org/10.1037/a0031858, PubMed: 23421443

Brungart, D. S., & Simpson, B. D. (2007). Cocktail party listening in a dynamic multitalker environment. *Perception & Psychophysics*, *69*, 79–91. https://doi.org/10.3758/BF03194455, PubMed: 17515218

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

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, *25*, 975–979. https://doi.org/10.1121/1.1907229

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*, 331–335. https://doi.org/10.3758/BF03196169, PubMed: 11495122

Cook, A. E., & Meyer, A. S. (2008). Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 886–899. https://doi.org/10.1037/0278-7393.34.4.886, PubMed: 18605876

Corps, R. E., Knudsen, B., & Meyer, A. S. (2022). Overrated gaps: Inter-speaker gaps provide limited information about the timing of turns in conversation. *Cognition*, *223*, 105037. https://doi.org/10.1016/j.cognition.2022.105037, PubMed: 35123218

Correa, Á., Lupiáñez, J., Madrid, E., & Tudela, P. (2006). Temporal attention enhances early visual processing: A review and new evidence from event-related potentials. *Brain Research*, *1076*, 116–128. https://doi.org/10.1016/j.brainres.2005.11.074, PubMed: 16516173

Daliri, A., & Max, L. (2016). Modulation of auditory responses to speech vs. nonspeech stimuli during speech movement planning. *Frontiers in Human Neuroscience*, *10*, 234. https://doi.org/10.3389/fnhum.2016.00234, PubMed: 27242494

Decuyper, C., Brysbaert, M., Brodeur, M. B., & Meyer, A. S. (2021). Bank of Standardized Stimuli (BOSS): Dutch names for 1400 photographs. *Journal of Cognition*, *4*, 33. https://doi.org/10.5334/joc.180, PubMed: 34327304

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. https://doi.org/10.3389/fpsyg.2014.00781, PubMed: 25120503

Dimitrijevic, A., & Stapells, D. R. (2006). Human electrophysiological examination of buildup of the precedence effect. *NeuroReport*, *17*, 1133–1137. https://doi.org/10.1097/01.wnr.0000223386.44081.ec, PubMed: 16837841

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*, 808–816. https://doi.org/10.1080/17470218.2017.1310261, PubMed: 28326995

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning & Verbal Behavior*, *20*, 641–655. https://doi.org/10.1016/S0022-5371(81)90220-6

Eulitz, C., Hauk, O., & Cohen, R. (2000). Electroencephalographic activity over temporal brain areas during phonological encoding in picture naming. *Clinical Neurophysiology*, *111*, 2088–2097. https://doi.org/10.1016/S1388-2457(00)00441-7, PubMed: 11068246

Fargier, R., & Laganaro, M. (2016). Neurophysiological modulations of non-verbal and verbal dual-tasks interference during word planning. *PLoS One*, *11*, e0168358. https://doi.org/10.1371/journal.pone.0168358, PubMed: 27992586

Fargier, R., & Laganaro, M. (2019). Interference in speaking while hearing and vice versa. *Scientific Reports*, *9*, 5357. https://doi.org/10.1038/s41598-019-41752-7, PubMed: 30926840

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*, 493–498. https://doi.org/10.1111/2041-210X.12504

Griffin, I., Miniussi, C., & Nobre, A. C. (2002). Multiple mechanisms of selective attention: Differential modulation of stimulus processing by attention to space or time. *Neuropsychologia*, *40*, 2325–2340. https://doi.org/10.1016/S0028-3932(02)00087-8, PubMed: 12417462

Grisoni, L., Tomasello, R., & Pulvermüller, F. (2021). Correlated brain indexes of semantic prediction and prediction error: Brain localization and category specificity. *Cerebral Cortex*, *31*, 1553–1568. https://doi.org/10.1093/cercor/bhaa308, PubMed: 33108460

Hansen, J. C., Dickstein, P. W., Berka, C., & Hillyard, S. A. (1983). Event-related potentials during selective attention to speech sounds. *Biological Psychology*, *16*, 211–224. https://doi.org/10.1016/0301-0511(83)90025-X, PubMed: 6615954

He, J., Meyer, A. S., & Brehm, L. (2021). Concurrent listening affects speech planning and fluency: The roles of representational similarity and capacity limitation. *Language, Cognition and Neuroscience*, *36*, 1258–1280. https://doi.org/10.1080/23273798.2021.1925130

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*, 177–180. https://doi.org/10.1126/science.182.4108.177, PubMed: 4730062

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

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144. https://doi.org/10.1016/j.cognition.2002.06.001, PubMed: 15037128

Jaeger, M., Mirkovic, B., Bleichner, M. G., & Debener, S. (2020). Decoding the attended speaker from EEG using adaptive evaluation intervals captures fluctuations in attentional listening. *Frontiers in Neuroscience*, *14*, 603. https://doi.org/10.3389/fnins.2020.00603, PubMed: 32612507

Jeffreys, H. (1998). *Theory of probability* (3rd ed.). Oxford, United Kingdom: Oxford University Press. https://doi.org/10.1093/oso/9780198503682.001.0001

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*, 643–650. https://doi.org/10.3758/BRM.42.3.643, PubMed: 20805586

Klaus, J., Schutter, D. J. L. G., & Piai, V. (2020). Transient perturbation of the left temporal cortex evokes plasticity-related reconfiguration of the lexical network. *Human Brain Mapping*, *41*, 1061–1071. https://doi.org/10.1002/hbm.24860, PubMed: 31705740

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123, PubMed: 20809790

Kwon, N., Sturt, P., & Liu, P. (2017). Predicting semantic features in Chinese: Evidence from ERPs. *Cognition*, *166*, 433–446. https://doi.org/10.1016/j.cognition.2017.06.010, PubMed: 28641220

Laganaro, M., & Perret, C. (2011). Comparing electrophysiological correlates of word production in immediate and delayed naming through the analysis of word age of acquisition effects. *Brain Topography*, *24*, 19–29. https://doi.org/10.1007/s10548-010-0162-x, PubMed: 20938730

Lange, K., & Röder, B. (2006). Orienting attention to points in time improves stimulus processing both within and across modalities. *Journal of Cognitive Neuroscience*, *18*, 715–729. https://doi.org/10.1162/jocn.2006.18.5.715, PubMed: 16768372

Lange, K., Rösler, F., & Röder, B. (2003). Early processing stages are modulated when auditory stimuli are presented at an attended moment in time: An event-related potential study. *Psychophysiology*, *40*, 806–817. https://doi.org/10.1111/1469-8986.00081, PubMed: 14696734

Li, X., Lu, Y., & Zhao, H. (2014). How and when predictability interacts with accentuation in temporally selective attention during speech comprehension. *Neuropsychologia*, *64*, 71–84. https://doi.org/10.1016/j.neuropsychologia.2014.09.020, PubMed: 25250708

Li, P., & Yip, M. C. W. (1996). Lexical ambiguity and context effects in spoken word recognition: Evidence from Chinese. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 228–232). Routledge.

Li, X., Zhang, Y., Li, L., Zhao, H., & Du, X. (2017). Attention is shaped by semantic level of event-structure during speech comprehension: An electroencephalogram study. *Cognitive Neurodynamics*, *11*, 467–481. https://doi.org/10.1007/s11571-017-9442-4, PubMed: 29067134

Mädebach, A., Jescheniak, J. D., Oppermann, F., & Schriefers, H. (2011). Ease of processing constrains the activation flow in the conceptual-lexical system during speech planning. *Journal of Experimental Psychology: Learning, Memory,*

*and Cognition*, *37*, 649–660. https://doi.org/10.1037/a0022330, PubMed: 21341927
Magne, C., Astésano, C., Lacheret-Dujour, A., Morel, M., Alter, K., & Besson, M. (2005). On-line processing of "pop-out" words in spoken French dialogues. *Journal of Cognitive Neuroscience*, *17*, 740–756. https://doi.org/10.1162/0898929053747667, PubMed: 15904541
Miniussi, C., Wilding, E. L., Coull, J. T., & Nobre, A. C. (1999). Orienting attention in time: Modulation of brain potentials. *Brain*, *122*, 1507–1518. https://doi.org/10.1093/brain/122.8.1507, PubMed: 10430834
Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*, 56–60. https://doi.org/10.1080/17470215908416289
Piai, V., Roelofs, A., & Maris, E. (2014). Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. *Neuropsychologia*, *53*, 146–156. https://doi.org/10.1016/j.neuropsychologia.2013.11.014, PubMed: 24291513
Piai, V., Roelofs, A., Rommers, J., & Maris, E. (2015). Beta oscillations reflect memory and motor aspects of spoken word production. *Human Brain Mapping*, *36*, 2767–2780. https://doi.org/10.1002/hbm.22806, PubMed: 25872756
Poulton, V., & Nieuwland, M. S. (2021). *Can you hear what's coming? Failure to replicate ERP evidence for phonological prediction*. Manuscript under review.
Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, *2*, 693–705. https://doi.org/10.1038/s41562-018-0406-4, PubMed: 31346278
Rabovsky, M., & McRae, K. (2014). Simulating the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition*, *132*, 68–89. https://doi.org/10.1016/j.cognition.2014.03.010, PubMed: 24762924
Roos, N. M., & Piai, V. (2020). Across-session consistency of context-driven language processing: A magnetoencephalography study. *European Journal of Neuroscience*, *52*, 3457–3469. https://doi.org/10.1111/ejn.14785, PubMed: 32432366
Sable, J. J., Low, K. A., Maclin, E. L., Fabiani, M., & Gratton, G. (2004). Latent inhibition mediates N1 attenuation to repeating sounds. *Psychophysiology*, *41*, 636–642. https://doi.org/10.1111/j.1469-8986.2004.00192.x, PubMed: 15189486
Sanders, L. D., & Astheimer, L. B. (2008). Temporally selective attention modulates early perceptual processing: Event-related potential evidence. *Perception & Psychophysics*, *70*, 732–742. https://doi.org/10.3758/PP.70.4.732, PubMed: 18556935
Schwent, V. L., & Hillyard, S. A. (1975). Evoked potential correlates of selective attention with multi-channel auditory inputs. *Electroencephalography and Clinical Neurophysiology*, *38*, 131–138. https://doi.org/10.1016/0013-4694(75)90222-9, PubMed: 45943
Schwent, V. L., Hillyard, S. A., & Galambos, R. (1976). Selective attention and the auditory vertex potential. II. Effects of signal intensity and masking noise. *Electroencephalography and Clinical Neurophysiology*, *40*, 615–622. https://doi.org/10.1016/0013-4694(76)90136-X, PubMed: 57047
Sjerps, M. J., & Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition*, *136*, 304–324. https://doi.org/10.1016/j.cognition.2014.10.008, PubMed: 25522192
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, U.S.A.*, *106*, 10587–10592. https://doi.org/10.1073/pnas.0903616106, PubMed: 19553212
Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 443–467. https://doi.org/10.1037/0278-7393.31.3.443, PubMed: 15910130
van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*, 584–589. https://doi.org/10.3758/BF03193889, PubMed: 17393828
Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., et al. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences, U.S.A.*, *90*, 8722–8726. https://doi.org/10.1073/pnas.90.18.8722, PubMed: 8378354
Wöstmann, M., Maess, B., & Obleser, J. (2021). Orienting auditory attention in time: Lateralized alpha power reflects spatio-temporal filtering. *Neuroimage*, *228*, 117711. https://doi.org/10.1016/j.neuroimage.2020.117711, PubMed: 33385562