



The effects of attention and task-relevance on the processing of syntactic violations during listening to two concurrent speech streams

Orsolya Szalárdy^{1,2} · Brigitta Tóth¹ · Dávid Farkas¹ · Annamária Kovács^{1,3} · Gábor Urbán^{1,4} · Gábor Orosz^{1,5} · Beáta Tünde Szabó^{1,6} · László Hunyadi⁷ · Botond Hajdu¹ · István Winkler¹

Published online: 14 June 2018
© Psychonomic Society, Inc. 2018

Abstract

The notion of automatic syntactic analysis received support from some event-related potential (ERP) studies. However, none of these studies tested syntax processing in the presence of a concurrent speech stream. Here we present two concurrent continuous speech streams, manipulating two variables potentially affecting speech processing in a fully crossed design: attention (focused vs. divided) and task (lexical – detecting numerals vs. syntactical – detecting syntactic violations). ERPs elicited by syntactic violations and numerals as targets were compared with those for distractors (task-relevant events in the unattended speech stream) and attended and unattended task-irrelevant events. As was expected, only target numerals elicited the N2b and P3 components. The amplitudes of these components did not significantly differ between focused and divided attention. Both task-relevant and task-irrelevant syntactic violations elicited the N400 ERP component within the attended but not in the unattended speech stream. P600 was only elicited by target syntactic violations. These results provide no support for the notion of automatic syntactic analysis. Rather, it appears that task-relevance is a prerequisite of P600 elicitation, implying that in-depth syntactic analysis occurs only for attended speech under everyday listening situations.

Keywords Concurrent speech streams · Syntactic violation processing · Lexical processing · Attention · N400 · P600 · N2b · P3

Introduction

The “cocktail party” situation (Cherry, 1953) poses one of the classical questions for experimental psychology. How can one follow the voice of one speaker while several others are talking at the same time? This problem includes two interrelated issues: (a) what are the computational algorithms that allow us to

decompose a mixture of similar sound streams (auditory scene analysis; Bregman, 1990; for the segregation of speech streams, see, Darwin, 2008)? and (b) how do we select one channel from the variety of inputs (selective attention; Broadbent, 1952)? Early studies of selective attention suggested that one can follow only one speech stream at a time (for an overview, see Broadbent, 1958). However, some studies

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13415-018-0614-4>) contains supplementary material, which is available to authorized users.

✉ Orsolya Szalárdy
szalardy.orsolya@tk.mta.hu

¹ Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar tudósok krt. 2, Budapest H-1117, Hungary

² Present address: Research Centre for Natural Sciences, Hungarian Academy of Sciences, P.O.Box 1519, Budapest H-1519, Hungary

³ Department of Telecommunications and Media Informatics, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Magyar tudósok krt. 2, Budapest H-1117, Hungary

⁴ Department of Cognitive Science, Faculty of Natural Sciences, Budapest University of Technology and Economics, Egy József u. 1, Budapest H-1111, Hungary

⁵ Faculty of Education and Psychology, Eötvös Loránd University, Kazinczy u. 23-27, Budapest H-1075, Hungary

⁶ Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Práter u. 50/a, Budapest H-1083, Hungary

⁷ Department of General and Applied Linguistic, University of Debrecen, Egyetem tér 1, Debrecen H-4032, Hungary

demonstrated that unattended speech streams are also semantically processed to some degree (Bentin, Kutas, & Hillyard, 1995; Lewis, 1970), while others showed that this is not always the case (see, e.g., Treisman, Squire, & Green, 1974). Yet other results suggest that information of high personal relevance, such as the listener's own name (Moray, 1959), can sometimes (though not always) be recognized within the unattended speech stream (Conway, Cowan, & Bunting, 2001). These controversies engendered the famous early versus late selective attention debate (for a review, see Bundesen & Habekost, 2007), which provided the focus for much of the research on selective attention for almost two decades. While modern theories of attention (Eysenck, 2012) conceptualize the issue of the processing bottleneck differently, the original questions remain unanswered. In the current study, we revisited the question of processing two concurrent speech streams from a linguistic processing point of view. Specifically, we tested how lexical, syntactic, and semantic information are processed within attended and unattended continuous natural speech streams under different task and attention conditions, using a fully crossed design measuring event-related brain potentials (ERP) and behavioral responses.

Syntactic processing is a necessary prerequisite of understanding speech and it has been associated with verbal working memory capacity (see, e.g., Caplan & Waters, 1999; Just & Carpenter, 1992). Two main ERP responses have been related to the processing of syntax: a negative component peaking at around 300- to 500-ms latency range (or sometimes earlier), and a later positive component. The negative waveform is usually identified as left-anterior negativity (LAN) or N400. LAN has been found for various types of morphosyntactic violations (Friederici, 2002, 2004). For word-category violations, Friederici and colleagues described an earlier negativity peaking between 100 and 300 ms from event onset (Friederici, 2002; see also Hahne & Friederici, 1999), the ELAN. The N400 has been typically associated with semantic violations (see, e.g., Kutas & Hillyard, 1983; Van Petten, 1995), phonological analysis, semantic memory access, and semantic/conceptual unification (for a review, see Kutas & Federmeier, 2011) and possibly reflects probabilistic prediction of the upcoming word (Kuperberg, 2016). Some studies found that this component was also present in response to syntactically incorrect sentences (Gunter & Friederici, 1999; Severens, Jansma, & Hartsuiker, 2008). In these cases, N400 has been distinguished from LAN by its more centrally dominant scalp distribution. However, it is not entirely clear from the literature which violations elicit N400 instead of LAN.

The late positive ERP component associated with syntactic processing has been termed P600 due to its typical peak latency of about 600 ms measured from the onset of the syntactic violation (Friederici, 2004; Hagoort, Wassenaar, & Brown, 2003; Osterhout, 1997). The P600 was found in several languages for phrase-structure (Friederici, Hahne, & Saddy, 2002; Hagoort,

Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992), number and case, and subcategorization violations (for reviews, see Friederici, 2004; Hagoort, Brown, & Osterhout, 1999). LAN/N400 is often followed by P600 for variety of syntactic errors (Friederici et al., 2002; Hagoort et al., 2003; Jolsvai, Sussman, Csuhaj, & Csépe, 2011) and probably reflects the revision of the syntactic structure (Friederici, 2002). P600 elicitation is modulated by attention (Hahne & Friederici, 1999) and it is related to task-relevant events (Schacht, Sommer, Shmuilovich, Casado Martienz, & Martín-Loeches, 2014). Kuperberg (2007) suggested that the P600 reflects a linguistic processing stream, which combines the semantic and syntactic information. Alternatively, the P600 might reflect target syntactic/semantic processing or categorization in general, similarly to the P3b (Kutas, Van Petten, & Kluender, 2006).

In Gunter and Friederici's (1999) study, two types of syntactic violations have been investigated: phrase structure violations and verb inflection violations. Sentences were presented visually, word-by-word. For phrase structure violations, an obligatory noun in a prepositional phrase was replaced by a verb form. Because the verb could not fill the syntactic role of the noun within the phrase, manipulated phrases had grammatically incorrect structure. For verb inflection violations, a correctly inflected verb was replaced by an incorrectly inflected form. In contrast to the phrase structure violations, these verb inflection violations did not modify the syntactic structure of the phrase. Both violations elicited N400 followed by P600 when the participants' task was to detect the violations with higher amplitudes for the phrase structure than for the verb inflection violations. However, when participants had to judge whether the last word of the sentence was presented in upper- or lower-case, these components were reduced or absent for the verb inflection violation, but still present for the phrase structure violation. Thus, this study showed processing differences between the two kinds of syntactic violations. The N400/P600 amplitude reduction for verb inflection violations showed that this violation was processed in a shallower manner during the visual feature discrimination task. In contrast, word category violations might be more difficult to integrate into the sentence. Therefore, a re-analysis of the syntactic structure is necessary even when the violation is task irrelevant. This result also suggests that the processing of phrase structure violations is less sensitive to attentional manipulations.

However, stimuli presented visually word-by-word do not provide a good comparison for the processing of continuous speech. Further, this procedure does not allow one to assess the processing of unattended syntactic violations. Participants in Maidhof and Koelsch' study (2011) were presented with speech and music concurrently and they were asked to either attend only to the speech or to the music stream. The speech material comprised discrete, grammatically incorrect German sentences with phrase structure violations. ELAN with similar amplitude was elicited whether or not the speech stream was

task-relevant. Pulvermüller et al. (2008) obtained compatible evidence by presenting matching and mismatching noun-word pairs in one ear and a simple tone in the other ear. The amplitude difference between rare grammatically incorrect and correct word pairs (termed syntactic mismatch negativity) was similar in a passive listening condition and when listeners performed a difficult perceptual task on the tones presented concurrently with the speech stimuli. Thus, these results support the notion that at least part of the processing of speech syntax is automatic. Hohlfeld et al. (2015) found that language processing (at the semantic level) is automatic using written sentences. However, neither tones nor music (despite its complex syntax-like structure) or a visual task activate all of the processes required for parsing speech. Further, neither of these studies used continuous speech with multiple semantically linked sentences. Hohlfeld, Sangals, and Sommer (2004) employed spoken word pairs investigating the effects of an additional task on speech processing and found that the language perception slowed down when there is a temporal overlap between the tasks. However, these also were not continuous speech materials. Therefore, not all of the processes of listening to natural continuous speech have been engaged in these studies.

In the current study, we tested the processing of natural continuous speech streams while independently manipulating the focus of attention and the task-relevance of the syntactic violations. Participants were instructed either to detect syntactic violations or performed a lexical task (responding to numerals) in the target stream (two different detection tasks). Further, participants were also tested on remembering information presented in one of the speech streams (tracking task). The tracking task also allowed us to compare between situations in which attention is focused on one stream (the target stream of the detection and the tracking task are the same: focused attention condition) or divided between the two streams (the target streams of the detection and tracking task are different: divided attention condition). Thus we could compare across the responses elicited by syntactic violations during focused and divided attention in the attended and unattended stream.

Two ERP components are usually elicited by auditory target events (including speech stimuli): the N2b (Näätänen, Simpson, & Loveless, 1982; for a review of the N2 components, see Pritchard, Shappell, & Brandt, 1991) and the P3b (for reviews, see Donchin & Coles, 1988; Polich, 2007; Sutton, Braren, Zubin, & John, 1965). The N2b is a centrally maximal negative waveform typically peaking between 150 and 250 ms from stimulus onset, which has been suggested to index stimulus classification (Ritter, Simson, Vaughan, & Friedman, 1979; Näätänen, 1990). P3b is a parietally maximal positive waveform that often follows the N2b, typically peaking between 300 and 400 ms from stimulus onset. The P3b has been interpreted as reflecting context

updating (Donchin & Coles, 1988), closure of the target detection cycle (Verleger, 1988) and as a sign of interaction between working memory and attentional processes (Polich & Herbst, 2000). Further, comparing the effects of focused and divided attention, Parasuraman (1980) found that a slow negative shift (measured in the 50- to 400-ms latency range from stimulus onset) was affected by whether participants performed a focused or a divided attentional task. This waveform could have also been interpreted as N2. However, in young adult participants, Wild-Wall and Falkenstein (2010) found no N2b difference for target sounds (German vowels) when attention was divided between two spatial locations (left and right) compared with when only one side was attended.

Based on the results of the afore-reviewed studies, we formulated the following expectations for the effects of the various attention and task conditions on the ERP responses. If syntactic processing is at least partly automatic (Maidhof & Koelsch, 2011; Pulvermüller, Shtyrov, Hasting, & Carlyon, 2008), then the syntactic violation related ERP responses (LAN/N400 and/or P600) should be elicited regardless of the direction of attention or whether or not the syntactic violations are task-relevant. If, however, attention is required for the syntactic violations to be detected, then one should only expect these components elicited for attended speech streams. If task-relevance is a prerequisite of detecting the syntactic violations employed in our study, then LAN/N400 and/or P600 should only be elicited when the syntactic violations are task-relevant. Finally, if both attention and task-relevance are critical, then these ERP components should be elicited only in the attended stream by target syntactic violations. N2b and P300 were expected to be only elicited by targets. Based on the ERP results of Wild-Wall and Falkenstein (2010), there may be no effect of divided versus focused attention on the N2b.

Materials and methods

Participants

Twenty-six healthy young native Hungarian adults (12 male, 14 female, mean age: 21.88 years, SD: 2.05; 24 right-handed) participated in the study for modest financial compensation. None of them had a history of psychiatric or neurological symptoms. All participants had pure-tone thresholds within normal limits (< 25 dB, separately for the two ears and <10 dB difference between ears) for the frequencies ranging from 250 Hz to 4 kHz. An informed consent form was signed by all participants after the aims and methods of the study were explained to them. The study was conducted in full accordance with the World Medical Association Helsinki Declaration and all applicable national laws; it was approved by the institutional review board, the United Ethical Review

Committee for Research in Psychology (EPKEB). One participant's data were excluded from the EEG analysis due to more than two malfunctioning EEG channels.

Stimuli

Participants listened to two concurrent continuous Hungarian speech segments of ca. 6 min duration (mean duration: 352.15 s, SD: 9.34; mean word count: 636.41, SD: 84.87; mean number phonemes per word: 6.48, SD: 0.29) presented from two loudspeakers positioned symmetrically at 30° left and right from midline, 200 cm in front of the participant. The speech material was selected from a collection of news articles, which were reviewed by a dramaturge for correct grammar, natural text flow, and to avoid garden-path sentences. The information from which the articles were created was found on Hungarian news websites. The articles contained emotionally neutral, lesser known pieces of information. They were recorded from two male native Hungarian actors (20 articles, each) and edited by a professional radio technician. Each article was presented once during the experiment with two articles being concurrently played (one from each actor) in each stimulus block. The soundtracks were recorded at 48 kHz with 32-bit resolution and presented by Matlab R2014a software (Mathworks Inc.) on an Intel Core i5 PC with ESI Julia 24-bit 192 kHz sound card connected to Mackie MR5 mk3 Powered Studio Monitor loudspeakers. The speech segments were recorded in the same room where the experiment took place and they were delivered from approximately the same location where the actor sat during the recording session (i.e., the loudspeaker was placed at the approximate position of the actor's head). This confounded the location (side) of the loudspeaker with the identity of the speaker, as all articles from the same actor were recorded at the same location.

Each article contained 45–57 numerals ($M=50.7$, $SD=2.7$) consisting of 2–4 syllables. Thirty-two of the 40 articles also included 19–26 ($M=20.5$, $SD=1.4$) syntactic violations. The minimal distance between numerals and syntactic violations was three syllables (min. 290 ms, $M=2,820$ ms, $SD=452$ ms). Two types of syntactic violations were generated, verb inflection violation (1) and phrase structure violation (2), each with two subtypes. Verb inflection violation was realized by subject-predicate agreement mismatch ($M=12.2$ /article, $SD=3.0$, range: 6–19) with a plural subject noun and a singular predicate verb or vice versa, where the subject could either precede or follow the predicate (Table 1, top half). (Both orders are grammatically correct in Hungarian.) For phrase structure violations, subject-object reorganization errors were used ($M=8.3$ /article, $SD=3.1$, range: 2–14) with appending the object suffix to the noun that played the role of subject in the sentence (Table 1, bottom half).

For the first type of syntactic violations, the violation can be detected at the point of hearing the mismatching second

Table 1 Parts of sentences illustrating the syntactic violations employed in the study. Top half: Verb inflection violations with (1) the subject preceding and (2) following the predicate; Bottom half: Phrase structure violations with (1) the object suffix appended to the subject and (2) the object suffix removed from the object. The text is shown both in Hungarian and in English, with the syntax violation shown in square brackets

Verb inflection violation

1) A **videójáték-ipar** szintén **kiháználják**...

The **videogame-industry** [subject-singular] also **exploit** [predicate-plural]...

2) Havonta átlagosan hatvan órán át **használja** az **emperek**...

Monthly on average sixteen hour **use** [predicate-singular] the **people** [subject-plural]...

Phrase structure violation

1) Később azonban az oktatási intézmények, majd végül több ezer **civilt** is csatlakozott...

Later however the educational institutions and finally thousands of **civilians** [subject+object suffix] joined ...

element of the intended agreement pair with the effect of recalculating the number of actors as subjects. Both variants of the second type of syntactic violations result in the reorganization of the syntactic structure of the sentence, because at the time the listener encounters the intended subject with the object suffix appended or the intended object without the object suffix, the grammatical role of these words is incorrectly assigned. Then, at the point, where the syntactic violation is discovered (hearing the predicate verb), the listener needs to reassign the role of the affected word within the sentence. The distance between the mismatching words (subject-predicate or object-predicate) never exceeded four syllables (max two words). In a pilot study (Kocsis, Hajdu, Orosz, Winkler, & Honbolygó, 2017) it was found that, when sentences were presented visually, one word at a time, all three types of syntactical violations elicited the Left Anterior Negativity (LAN) and/or the P600 component.

Procedure

Listeners were tested in an acoustically attenuated and electrically shielded, dimly lit room at the Research Centre for Natural Sciences, MTA, Budapest, Hungary. In addition to the two loudspeakers, a 23-in. monitor was placed directly in front of the listener at a distance of 195 cm. Participants were instructed to keep eye blinks and all other motor activity to a minimum during the stimulus blocks by focusing on a fixation cross (the “+” sign) that was continuously present at the center of the monitor. For each stimulus block, two different short articles (one from each speaker) were randomly selected for simultaneous presentation. Thus, participants were listening to two concurrent speech streams produced by two different speakers from two spatial locations.

Six experimental conditions were delivered in which combinations of three different tasks were employed. For the

“tracking task,” listeners were informed that at the end of the stimulus block, they will be asked five questions regarding the contents of one or both of the speech streams. The tracking task was employed in each condition. There was either no other task (“only tracking”) or one of the detection tasks was employed. In the detection tasks, listeners were instructed to press a hand-held response key with their right thumb as soon as they detected the presence of a numeral word (“numeral detection task”) or a syntactic violation (“syntactic violation detection task”). Only numerals indicating the quantity of something within the context of the text were valid targets, words including a numeral as a component were not. The instruction for the syntactic violation detection task emphasized that the button should be pressed as soon as the listener detects that the sentence is grammatically incorrect. The assignment of the side of the stream for the detection task was constant within each listener; it was counterbalanced across listeners. The target speech stream of the tracking task was either the same as that of the detection task (“focused attention condition”) or the opposite (“divided attention condition”). In the only tracking task conditions (focused and divided attention), the articles contained no syntactic violations. Data of these two conditions are not reported here, as there were no detection task targets or syntactic violations, which could be expected to elicit attention or syntactic violation related ERP responses. As a result, data from the following four task conditions were analyzed for the current study (Fig. 1): (1) Focused attention – numeral detection task, (2) Divided attention – numeral detection task, (3) Focused attention – syntactic violation detection task, (4) Divided attention – syntactic violation detection task. Note that in this arrangement there is one target event (numeral or syntactic violation appearing in the stream designated for the detection task) and three types of non-target events. For disambiguation, we term the target type events appearing in the concurrent stream as distractors. The other two non-target events are termed task-irrelevant events: syntactic violations appearing during the numeral detection task and numerals appearing during the syntactic violation detection task. Task-irrelevant events were delivered both within the stream designated for the detection task as well as within the concurrent stream.

The two only tracking-task conditions received two stimulus blocks each, the other four conditions four blocks each. Thus, the experimental session consisted of 20 blocks (each with a unique pair of articles), with a mandatory break after the 10th block and occasional shorter breaks between blocks as requested by the participant. The blocks for the focused attention only tracking task condition were presented at the first and the 20th position, those for the divided attention only tracking task condition at the second and 19th position. The

rest of the stimulus blocks (the ones, whose data is reported here) were divided into two halves, each half containing two blocks of each condition, which were delivered in a pseudorandomized order with the constrain that the same condition should not appear twice in a row. The articles were randomly assigned to one of the two task conditions, separately for each speaker.

After each stimulus block, a recognition memory test was performed (the test for the tracking task). The test consisted of five multiple-choice questions with four possible answers, each. Each question corresponded to one piece of information that appeared within the article assigned to the tracking task. The experimenter read the question and the four possible answers and the listener was asked to verbally indicate the correct answer. The experimenter noted the participant’s choice and followed up with a request for confidence judgement with four alternatives: “I don’t remember I was just guessing”, “I am not sure, but the option I chose sounded familiar; I think I heard it during the last block”, “I am sure; I remember having heard it during the last block”, “I know the answer from some other source”. The confidence judgment was then recorded by the experimenter.

Data analysis

The four conditions that included a detection task form a 2×2 arrangement of Attention (Focused vs. Divided) \times Detection Task (Numeral detection vs. Syntactic violation detection).

Behavioral measures

Detection task performance Hits were initially searched for within a window of 0–5,000 ms from the onset of the target events: onset of the numeral word or the onset of the word at which the syntactic violation could be detected. In order to exclude responses, which were unlikely to have corresponded to the given event, separately for the two detection tasks, responses were rejected if they were longer than 95 % ($>1,885$ ms for numerals and $>2,214$ ms for syntactic violations) or shorter than 5 % (<453 ms for numerals and <513 ms for syntactic violations) of all responses (collapsed across the two Attention conditions and participants). From the remaining responses, mean reaction times (RTs) were calculated separately for each participant, Detection Task, and Attention condition. Next d' values (the standard measure for detection sensitivity; Green & Swets, 1988) were calculated from the accepted responses (“hits”) and the number of target events with no valid response (“misses”); for “false alarms” and “correct rejections,” time windows identical to the ones used for identifying hits were set for each distractor event (i.e., events of the same type occurring in the concurrent speech stream). The distractor effect was characterized by the ratio

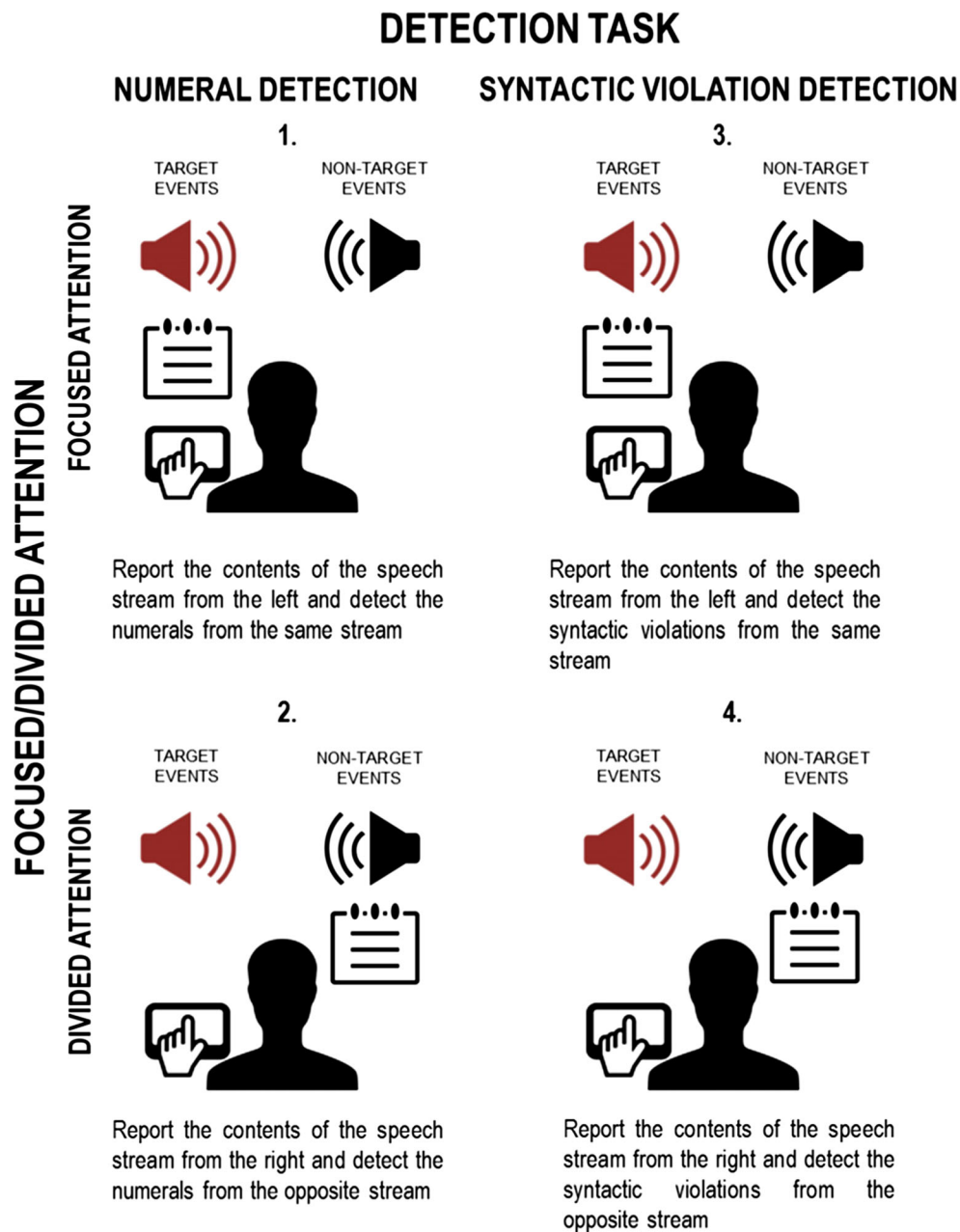


Fig. 1 Schematic illustration of the experimental conditions. Participants were presented with two concurrent speech streams under four different experimental conditions: (1) Focused attention – numeral detection task, (2) Divided attention – numeral detection task, (3) Focused attention – syntactic violation detection task, (4) Divided attention – syntactic violation detection task. The gist of the task instructions specifying the

target events for the detection task and the location of the target speech streams for each task are shown separately below each condition. The “text” pictograms indicate the target speech stream of the tracking task. Red “loudspeaker” and the “button-press” pictograms indicate the target stream of the detection task.

between the false alarms (FA; i.e., responses to distractors) and all non-hit responses, separately for each condition.

Tracking task performance Recognition performance was separately calculated for each participant and condition. In order to increase the sensitivity of this measure, items (questions) with an overall correct response rate (collapsed across all conditions and participants) above 95% or below 30% (25%

representing chance level) were excluded from the analyses. Note that due to the random assignment of the texts across the different conditions, the same text (and thus the same questions) could have appeared in different Attention/Detection Task conditions for different participants. Further, responses with the confidence judgment “I know the answer from some other source” were excluded from the calculation of recognition performance measure for the given participant.

Recognition performance was then calculated as the percentage of correct responses pooled across stimulus blocks, separately for each condition.

Behavioral data analysis

Separately for d' , RT, distractor effect, and the recognition index, an analysis of variance (ANOVA) was performed with the factors of DETECTION TASK (numeral vs. syntactic violation detection) \times ATTENTION (focused vs. divided) \times LOCATION (left vs. right detection task target stream), where Detection Task and Attention were within-subject factors, whereas Location a between-subject factor. For syntactic violations separate ANOVAs were conducted with the factors of ATTENTION (focused vs. divided) \times VIOLATION (phrase structure vs. verb inflection), separately for RT and d' . Statistical analysis was performed using Matlab 2015b (Mathworks, Inc.) and its Statistics and Machine Learning Toolbox 10.1. The alpha level was 0.05. All significant main effects and interactions are described. The p -values of post hoc pair-wise comparisons were adjusted using Bonferroni's correction.

EEG recording and preprocessing

EEG was continuously recorded from a few seconds before the beginning to a few seconds after the end of the two concurrent speech streams with a BrainAmp DC 64-channel EEG system with actiCAP active electrodes (Brain Products GmbH). EEG recordings were synchronized with the speech segments by matching an event trigger marked on the EEG record to the concurrent presentation of a beep sound in the audio stream with < 1 ms accuracy. Electrodes were placed according to the International 10/20 system with the addition of one electrode placed on the tip of the nose, and for EOG monitoring, one electrode placed lateral to the outer canthus of the right eye and another below the left eye. Electrode impedances were kept below 15 k Ω . During the recording, the FCz lead served as the reference electrode. The sampling rate was 1 kHz, and a 100-Hz online low-pass filter was applied.

EEG data analysis was performed using Matlab 2013a (Mathworks Inc., Natick, MA, USA). The continuous EEG signal was off-line band-pass filtered between 0.5 and 45 Hz by a finite impulse response (FIR) filter (Kaiser windowed, Kaiser $\beta=5.65$, filter length 4530 points) by the EEGLab 11.0.3.1.b toolbox (Delorme & Makeig, 2004). The 0.5-Hz high-pass filter was employed for removing the slow oscillatory drifts from the continuous data before estimating the ICA components. Maximum two bad EEG channels per subject were interpolated using the spline interpolation algorithm implemented in EEGLab. The Infomax algorithm of Independent Component Analysis (ICA) implemented in EEGLab was employed for artifact removal (Delorme et al., 2007). ICA

components constituting blink artifacts and horizontal eye-movements were removed via visual inspection of the topographical distribution and frequency contents of the components. Data were re-referenced to the electrode attached to the tip of the nose.

ERP data analysis

For analyzing the ERP responses, epochs were extracted from the continuous EEG record between -200 and +2,200 ms relative to the onset of numerals and syntactic violations ("events"; both being triggered from the onset of the word). Baseline correction was applied using the 200-ms pre-event interval. Artifact rejection with a threshold of 100- μ V voltage change was based on the whole duration of the epochs, separately for each electrode. As the number of syntactic violations were limited and for obtaining as clear data as possible further artifacts were eliminated manually with visual inspection of the data containing syntactic violations. For target events, only hits, for distractors, only correct rejections were analyzed; responses for all task-irrelevant events were analyzed.

Numeral targets elicited two consecutive ERP waveforms identified as N2b and P3 with maximal amplitudes at the Pz electrode. Based on the peak latency and width of the response, the time window for measuring the N2b component was set to 146–246 ms for the focused, and to 226–326 ms for the divided attention condition (i.e., 100 ms long windows centered on the group-average peaks). The same time windows were used for the corresponding non-target (distractor and task-irrelevant) numeral events. P3 amplitudes were measured in the 650–850 ms latency range for the focused attention condition and 700–900 ms for the divided attention condition (i.e., 200 ms long windows centered on the group-average peaks; the same amplitude measurement windows were used for the corresponding non-target events). Target and task-irrelevant phrase structure violations elicited two ERP responses: (1) a negative waveform peaking around 400 ms from event onset and (2) a later positive waveform peaking at ca 900 ms. Because of its centro-parietal distribution, the former was identified as N400, whereas the later as P600. Target verb inflection violations elicited only a P600, peaking at around 1,100 ms; however, a clear N400 was observed for task-irrelevant verb inflection violations. Because the time when the violation was actually detected could not be exactly established within continuous speech and the words allowing the detection of the violation consisted of only two to four syllables, the components were jittered during the averaging. However, due to the randomization of the texts across the different conditions, both the jitter and the distribution of different syntactic violations were approximately equal between the different conditions; therefore, the jitter did not affect the comparisons between conditions. As both of the components appeared with a parietal maximum, their amplitudes

were measured at the Pz electrode within the latency ranges of 280–500 ms for N400 and 800–1,000 ms for P600 elicited by phrase structure violations and 1,000–1,200 ms for those elicited by verb inflection violations. The same amplitude measurement windows were used for the corresponding non-target events. For testing the latency difference between the P600 to verb inflection and phrase structure violations, the latency at which the response reached its maximum amplitude within the 800–1,200 ms time window was separately measured for each participant.

For analyzing the ERP amplitudes elicited in the numeral detection task, data of all 25 participants could be used (after rejecting one participant out of 26 for more than two malfunctioning EEG channels). For the syntactic violation detection task, 11 more participants' data had to be rejected, because they had fewer than 15 artefact-free responses in at least one of the syntactic violation categories/condition (after eliminating misses and false alarms). Therefore, data of fourteen participants were entered into the analysis. (Note that no statistical analysis required ERP data from both detection tasks.) The average epoch number was 35.46 over all conditions (see Online Supplementary Tables 1 and 2 for the number of accepted epochs for each analyzed category). Visual inspection of the responses to numerals and syntactic violations appearing in the non-target stream showed that these events did not elicit any of the expected ERP components (N2b, P3, N400, and P600). Therefore, the responses to these events were only tested by one-sample *t*-tests (i.e., against zero). The presence of the ERP components in the target stream was verified by one-tailed *t*-tests. N2b and P3 amplitude amplitudes for numerals appearing in the target stream were entered into repeated-measures ANOVAs with the factors of DETECTION TASK (numeral vs. syntactic violation detection) \times ATTENTION (focused vs. divided). N400 and P600 amplitudes elicited by syntactic violations in the target streams were entered into repeated-measures ANOVAs with the factors of DETECTION TASK (numeral

vs. syntactic violation detection) \times ATTENTION (focused vs. divided) \times VIOLATION (phrase structure vs. verb inflection). The P600 latencies were tested by a repeated-measures ANOVA with the factors of ATTENTION (focused vs. divided) \times VIOLATION (phrase structure vs. verb inflection). The alpha level was 0.05. All significant main effects and interactions are described. Statistical analyses were conducted with the STATISTICA software; post hoc tests were performed using Tukey's HSD.

Results

Behavioral results

Figure 2 shows the summary of the results for the behavioral measures.

Detection task performance Analysis of d' values revealed significant main effects of DETECTION TASK ($F_{1,24}=235.066$; $p<0.001$, $\eta_p^2=0.907$) and ATTENTION ($F_{1,24}=27.256$; $p<0.001$, $\eta_p^2=0.532$). Listeners performed significantly better in the numeral detection than in the syntactic violation detection task and in the focused than in the divided attention condition. The distractor effect was significantly larger for numeral than for syntactic violation detection (main effect of the DETECTION TASK: $F_{1,24}=158.009$; $p<0.001$, $\eta_p^2=0.868$). The separate ANOVA assessing differences between the two subtypes of syntactic violations revealed main effects of ATTENTION ($F_{1,13}=9.792$; $p<0.01$, $\eta_p^2=.430$) and VIOLATION ($F_{1,13}=11.609$; $p<0.01$, $\eta_p^2=.472$). Detection performance was significantly better for phrase structure than for verb inflection violations, and for the focused than for the divided attention condition.

The analysis of RT's yielded a significant main effect of DETECTION TASK ($F_{1,24}=198.095$; $p<0.001$, $\eta_p^2=0.892$), which was due to listeners responding faster when detecting

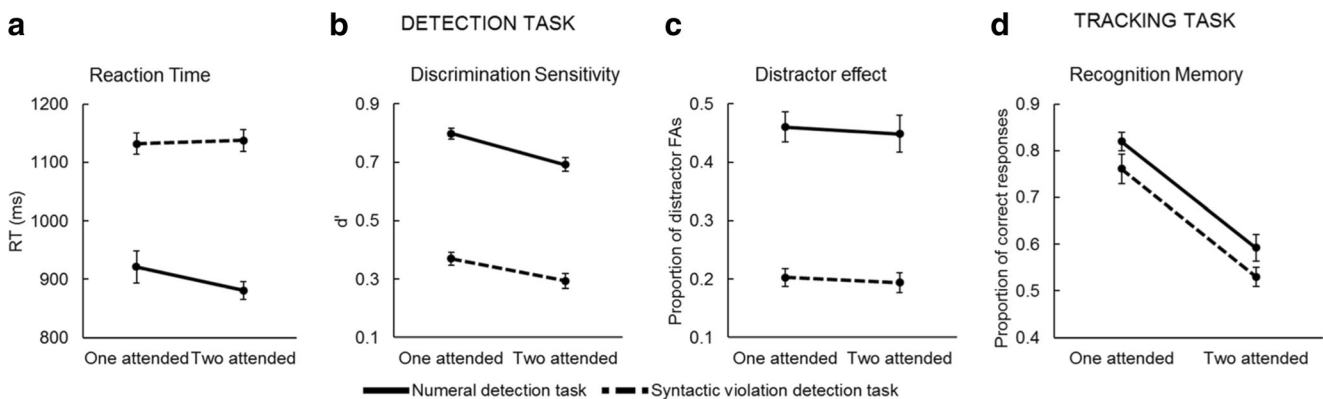


Fig. 2 Group average ($N=25$) performance in the detection task (indexed by RT and d' ; panels **A** and **B**, respectively), the effect of the distractors (assessed as the ratio between the number of responses to distractors and

the total number of the non-hit responses; panel **C**), and performance in the tracking task (recognition memory performance; panel **D**). Standard errors of mean are shown for each data point

numerals than syntactic violations. In the separate ANOVA comparing the two subtypes of syntactic violations, a main effect of VIOLATION was found ($F_{1,13}=72.437$; $p<0.001$, $\eta_p^2=.848$). This effect was caused by the faster RT's found for phrase structure than for verb inflection violations (1,026.3 ms, SD: 102.08 (focused attention) and 1,042.6 ms, SD: 121.26 (divided attention) for phrase structure violations and 1216.3 ms, SD: 100.98 (focused attention) and 1,207.8 ms, SD: 65.63 (divided attention) for verb inflection violations). In both cases, differences between reaction times measured for different target events may not be related to task differences, because reaction times were measured from word onsets and distance between word-onset and the point at which target detection could occur was not balanced between numerals and syntactic violations or between the two subtypes of syntactic violations.

Tracking task performance For the proportion of the correct answers to the questions about the news articles, significant main effects of ATTENTION ($F_{1,24}=97.153$; $p<0.001$, $\eta_p^2=0.802$) and DETECTION TASK ($F_{1,24}=11.258$; $p<0.005$, $\eta_p^2=0.319$) were found. More details of the speech stream were remembered by listeners in the focused than in the divided attention condition, and during the numeral than during the syntactic violation detection task. Further, significant interaction was obtained between ATTENTION and LOCATION ($F_{1,24}=13.625$; $p=0.001$, $\eta_p^2=0.362$). Post hoc pairwise comparisons showed higher recognition performance for the left than for the right target stream in the focused attention condition ($p<0.01$) with no significant difference in the divided attention condition.

ERP responses to numerals

Figure 3 shows the ERP responses and the scalp distributions of the N2b and P3 components elicited by numerals. N2b and P3 were elicited for detected target numerals both in the focused and the divided attention condition ($p<0.05$ for the N2b, and $p<0.001$ for the P3 in both conditions). The amplitude of both components was highest at the Pz electrode with centroparietal scalp distributions. Numerals in the non-target streams did not elicit significant N2b ($p>0.12$ at least) or P3 components ($p>0.09$ at least). The ANOVAs of the ERP amplitudes elicited by numerals appearing in the target stream revealed significant main effects of DETECTION TASK ($F_{1,23}=14.257$; $p<0.001$, $\eta_p^2=0.373$ and $F_{1,23}=39.842$; $p<0.001$, $\eta_p^2=0.624$; for the N2b and P3 components, respectively). This was caused by numerals appearing during the numeral detection task eliciting significantly larger N2b and P3 responses than numerals appearing during the syntactic violation detection task. We found no significant main effect of ATTENTION or significant interactions between DETECTION TASK and ATTENTION. Unexpectedly, a

small but significant positive response ($p<0.01$; tested against 0 in the 400–600 ms time window) was elicited by task-irrelevant numerals in the target stream (Fig. 3, panel B).

ERP responses to syntactic violations

Figures 4 and 5 show the ERP responses and the scalp distributions of the N400 and P600 components elicited by syntactic violations. One-tailed *t*-tests verified that the N400 was elicited by detected target phrase structure violations in the focused and divided attention ($p<0.01$, both) as well as by task-irrelevant verb inflection violations in the focused attention numeral detection task condition ($p<0.05$). Task-irrelevant phrase structure violations also elicited the N400 in the focused attention task ($p<0.05$) and there was a strong tendency in the divided attention task as well ($p=0.052$). P600 was elicited by both types of detected target syntactic violations in both attention conditions ($p<0.001$, all). P600 showed typical parietal distribution whereas the N400 was centro-parietally distributed. Syntactic violations in the non-target streams did not elicit significant N400 ($p>0.14$, at least) or P600 components ($p>0.17$ at least) except for verb inflection violations in the focused attention numeral detection task condition ($t_{13}=2.649$, $p<0.05$ for P600). This exception could have been due to a slow positive shift appearing over the whole duration of the epoch rather than to the elicitation of P600. The ANOVA of the N400 amplitudes yielded a significant main effect of VIOLATION ($F_{1,13}=10.463$; $p<0.01$, $\eta_p^2=0.446$): phrase structure violations elicited larger N400 than verb inflection violations. Significant interaction was found between DETECTION TASK and VIOLATION ($F_{1,13}=7.624$; $p<0.05$, $\eta_p^2=0.370$). Post hoc tests revealed that the interactions was caused by significant difference between the N400 amplitudes elicited by phrase structure violations and verb inflection violations during the syntactic violation detection ($p<0.01$) but not during numeral detection ($p=0.215$). No other significant main effects or interactions were obtained ($p>0.09$, at least). For the P600 amplitudes, a significant main effect was found for DETECTION TASK ($F_{1,13}=71.882$, $p<0.001$, $\eta_p^2=0.847$), which was caused by the larger P600 amplitudes during the syntactic violation detection than during numeral detection. No other significant main effects or interactions were found. Statistical analysis of the P600 latencies revealed a main effect VIOLATION ($F_{1,13}=44.861$, $p<0.001$, $\eta_p^2=0.775$): P600 for phrase structure violations had a shorter latency than for verb inflection violations. No other significant main effects or interactions were found.

A post hoc time-frequency analysis was conducted for testing whether the lack of evoked activity for non-target syntactic violations was due to insufficient time-locking of such activity and the onset of the violating word. Significant induced delta and alpha activity was obtained for target syntactic violations in the P600 latency range, but not significant induced activity

Parietal (Pz) ERP responses to numerals

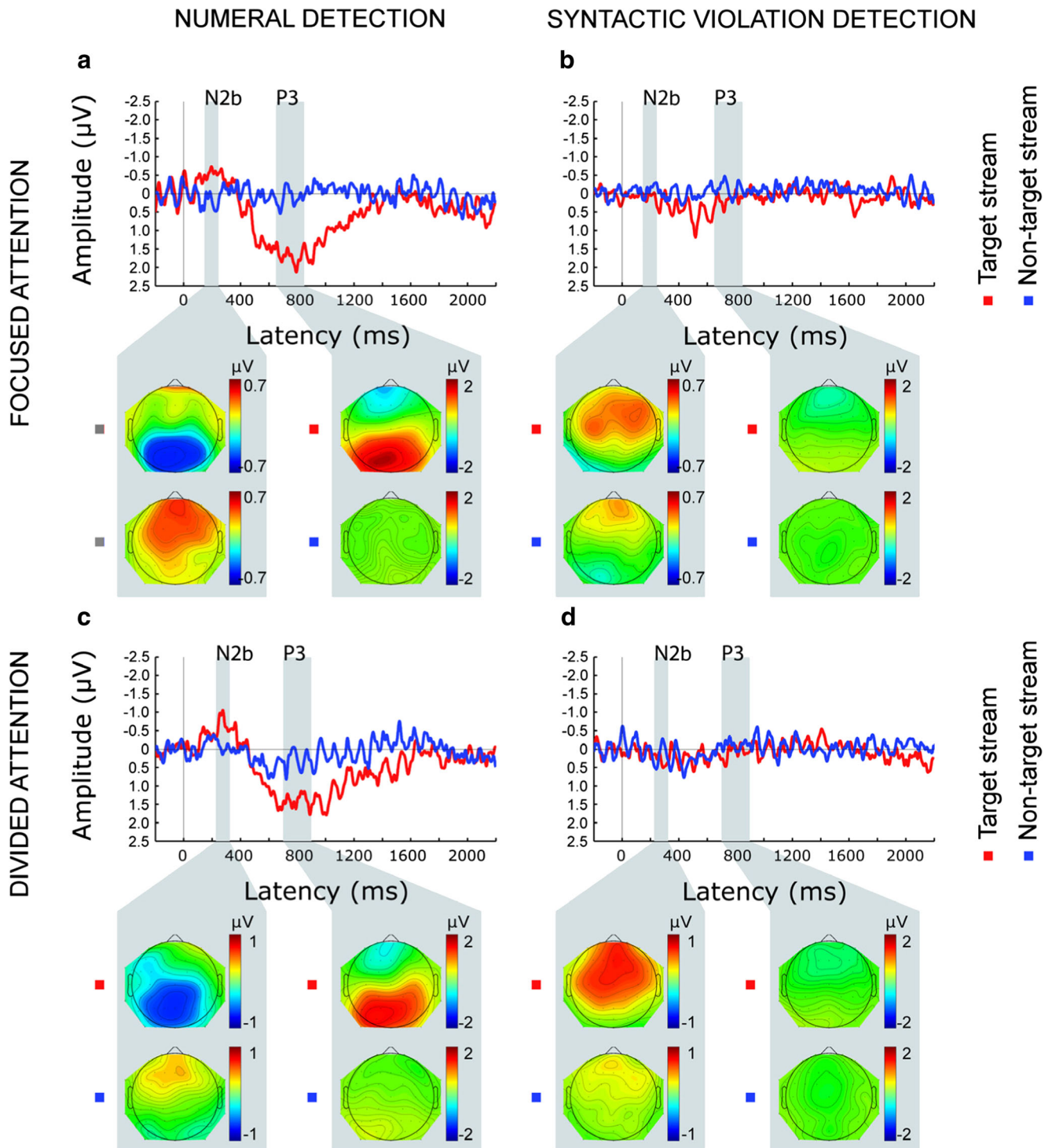


Fig. 3 Group-average ($N=25$) parietal (Pz) ERP responses elicited by numerals. The top half (panels **A** and **B**) shows the responses obtained for the focused, the bottom half (panels **C** and **D**) for the divided attention conditions. Left panels (**A** and **C**) show the responses recorded during the numeral detection, right panels (**B** and **D**) during the syntactic violation detection task. “Target stream” traces (red color) represent the responses to numerals appearing in the target speech stream; “Non-target stream”

traces (blue color) represent the responses to numerals in the non-target concurrent stream. Scalp topographies for the N2b and are presented below the ERP responses (upper row: target stream – red square; lower row: non-target stream – blue square). Calibration of the color scale is shown on the right side of the scalp topography maps. Maps were spline interpolated with a smoothing factor of 10^{-7}

was found for non-target violations in any frequency range (see the [Online Supplementary Material](#) for this analysis).

Discussion

In the present study, we measured ERP responses and task performance for assessing the processing of the lexical, syntactic, and semantic information in two concurrently delivered continuous speech streams using a fully crossed design between attention (focused vs. divided) and task type (lexical vs. syntactic). We found that selectively attending to one of the speech streams was more beneficial than dividing attention between two streams for processing all of these elements of speech, as participants were more accurate in detecting target numerals and syntactic violations as well as remembering more information from the tracked speech stream in the focused than in the divided attention condition. Detection performance (d' , and possibly RT – but see the cautionary note above) was superior for numeral targets than for syntactic violations and for phrase structure violations than for verb inflection violations. These results indicate that the detection of syntactic violations in the current study required more processing capacities, such as working memory, which is known to be affected by the direction of attention (Engle, 2002). Note that the difference in the difficulty between the two types of tasks also depends on the specific kind of syntactic violations tested. That is, numeral detection is not easier than syntactic violation detection, per se. Recognizing a word (such as a numeral) requires semantic categorization, whereas detecting syntactic violations involves fitting objects (words) into a hierarchical structure, the rebuilding of which (due to violation of especially the phrase structure) may be quite costly. Indeed, previous studies have shown that lexical-semantic and syntactic processes involve different brain mechanisms (see, e.g., Friederici, Opitz, & von Cramon, 2000). The difference in attentional capacity requirement between the two detection tasks probably also underlies the larger distractor effect found for the numeral than for the syntactic violation detection task (i.e., distractor numerals were more often confused for targets than distractor syntactic violations). This result shows that the non-target stream was processed to a higher degree during numeral than syntactic violation detection. Further, semantic information was more accurately reported during the numeral than the syntactic detection task, irrespective of whether tracking and detection was to be performed on the same or a different speech stream. This result can also be explained by the different capacity requirements of the two tasks.

Judging by its latency and scalp distribution, the ERP component elicited by target phrase structure violations and task-irrelevant syntactic violations can be classified as N400. Supporting this assumption, Zawiszewski and Friederici (2009) found N400 that showed parietal distribution for

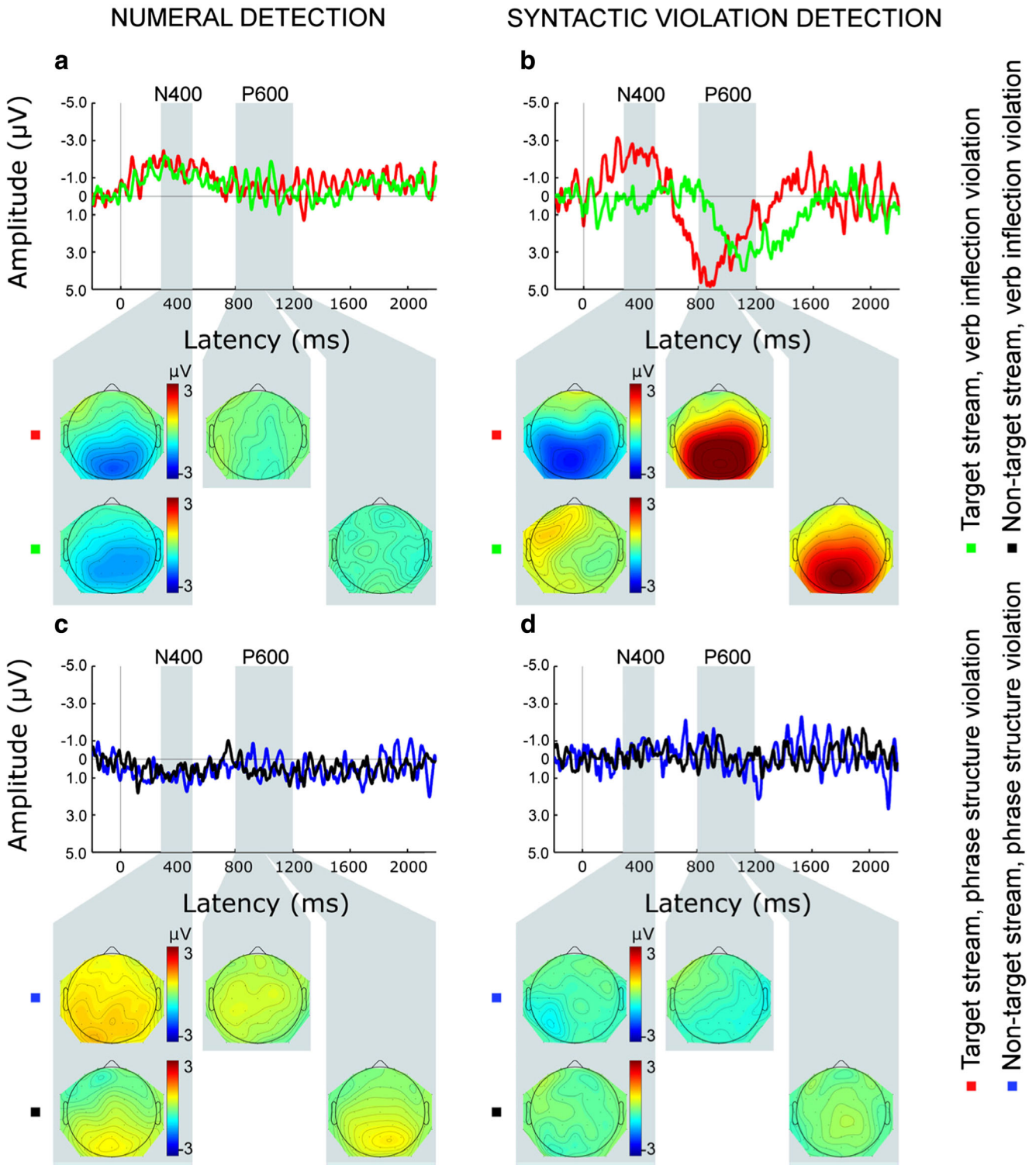
subject-verb and object-verb disagreements in Basque language. They presented grammatically correct and incorrect sentences to participants, who were instructed to tell whether the sentence was correct or not. The authors found that N400 was elicited by both subject-verb and object-verb disagreements, the former analogue to the current verb inflection, while the latter to the current phrase structure violations. Further, P600 was elicited by both types of violations with higher amplitude for object-verb than for subject-verb disagreement. On the other hand, Hungarian subject-verb disagreement has been previously found to elicit LAN and P600 (Jolsvai et al., 2011; Kocsis et al., 2017). Note, however, that in these studies, sentences were presented visually in a word-by-word manner rather than by continuous speech.

Syntactic violations were detected by the listener's brain in all task-relevant speech streams irrespective of whether or not they were task-relevant themselves (as attested by the elicitation of N400 and/or P600). The current experimental situation provided a good model of what humans do in everyday life during speech comprehension. Thus, it is likely that syntactic violations are detected whenever we listen to a speech stream. On the other hand, no N400 was obtained for syntactic violations in the non-target speech streams (with respect to the detection task), not even when participants were instructed to track the contents of the stream (divided attention condition). This result indicates that syntactic analysis occurs only when attention is primarily allocated to the speech stream. Thus, the current results do not provide support for the notion of attention-independent processing of speech syntax.

We observed a task effect on the N400 amplitude. The N400 amplitude was higher for phrase structure than for verb inflection violations in the syntactic violation detection task. In contrast, in the numeral detection task, both types of syntactic violations elicited the N400 with the same amplitude (although in the divided attention task phrase structure violations did not reach significance, rather we found a very strong tendency). This result contradicts those obtained by Gunter and Friederici (1999). These authors observed a sizable N400 response to verb inflection violations when syntactic violations were task-relevant and that the N400 amplitude was reduced when syntactic violations were task-irrelevant – a pattern opposite to the current one. This discrepancy might be due to the different modes of stimulus presentation: visual word-by-word in Gunter and Friederici's (1999) as opposed to continuous speech in the current study. It is possible that when listening to a speech stream in a multi-talker environment with the aim of extracting semantic information (as is typical in everyday situations), there is no strong differentiation between the syntactic violations due to shallower syntactic analysis (Kuperberg, 2007).

In contrast to the N400 component, P600 was only observed for target syntactic violations (task-relevant, target stream). Because listeners could clearly comprehend the full

Parietal (Pz) ERP responses to syntactic violations in the focused attention condition



speech stream even when they performed the numeral detection task (as attested by their tracking performance), the lack

of P600 may either be explained by assuming shallow syntactic analysis or that when listening to continuous speech

◀ **Fig. 4** Group-average (N=14) parietal (Pz) ERP responses elicited by syntactic violations in the focused attention condition. Left panels (A and C) show the responses during the numeral detection, right panels (B and D) during the syntactic violation detection task. “Target stream” traces (top half) represent the responses to target syntactic violations (red color for phrase structure violations, green for verb inflection violations); “Non-target stream” traces (bottom half) represent the responses to syntactic violations in the non-target concurrent stream (blue for phrase structure violations, black for verb inflection violations). Scalp topographies for the N400 and P600 are presented below the ERP responses (upper row, red/blue square: phrase structure violation; lower row, green/black square: verb inflection violation). Calibration of the color scale is shown on the right side of the scalp topography maps. Maps were spline interpolated with a smoothing factor of 10^{-7}

(possibly only in a multi-stream environment) syntactic reorganization is not strictly time-locked to the moment of detecting a syntactic violation. (Note that P600 was elicited when the same sentences appeared in the syntactic violation detection task. Thus, the lack of significant P600 cannot have been due to low S/N ratio.) However, a post hoc time-frequency analysis did not show significant induced activity for non-target syntactic violations. Therefore, the absence of P600 is not likely due to the lack of time-locking. It is also possible that part of the P600 belongs to the P3 component group, which reflects processes related to target detection (see, e.g., Coulson, King, & Kutas, 1998; Gunter, Stowe, & Mulder, 1997; Kutas et al. 2006). This would also explain why it was only elicited by target syntactic violations. Alternatively, the P600 elicitation could have been modulated by the required processing capacity. Similar to the current data, Schacht et al. (2014) found diminished P600 amplitudes when sentence-internal relationships were task-irrelevant. These authors suggested an alternative explanation of the diminished P600 amplitude: Without sufficient processing capacity allocated to the sentence-structure, no P600 is elicited. Our results are compatible with this explanation as the numeral detection task and tracking task could have engaged the processing capacities required analyzing the sentence structure.

We also found that the P600 latency was longer for verb inflection than for phrase structure violations. This latency difference has been accompanied by reaction time differences of similar magnitude: reaction times were at least 160 ms slower for verb inflection violations than for phrase structure violations. Although preparatory reaction-related activity could have overlapped the P600 amplitude measurements, the peak and onset latencies of the P600 components are more than one standard deviation (at least 150 ms) shorter than the reaction times, which suggest that the reaction-time difference is a consequence of the P600 latency difference rather than confounding the P600 latency measurement. These results are somewhat surprising given that, in contrast to phrase structure violations, the current verb inflection violations did not require significant syntactic reorganization of the affected sentence. As a post hoc explanation, we suggest that for verb inflection violations, reorganization is only

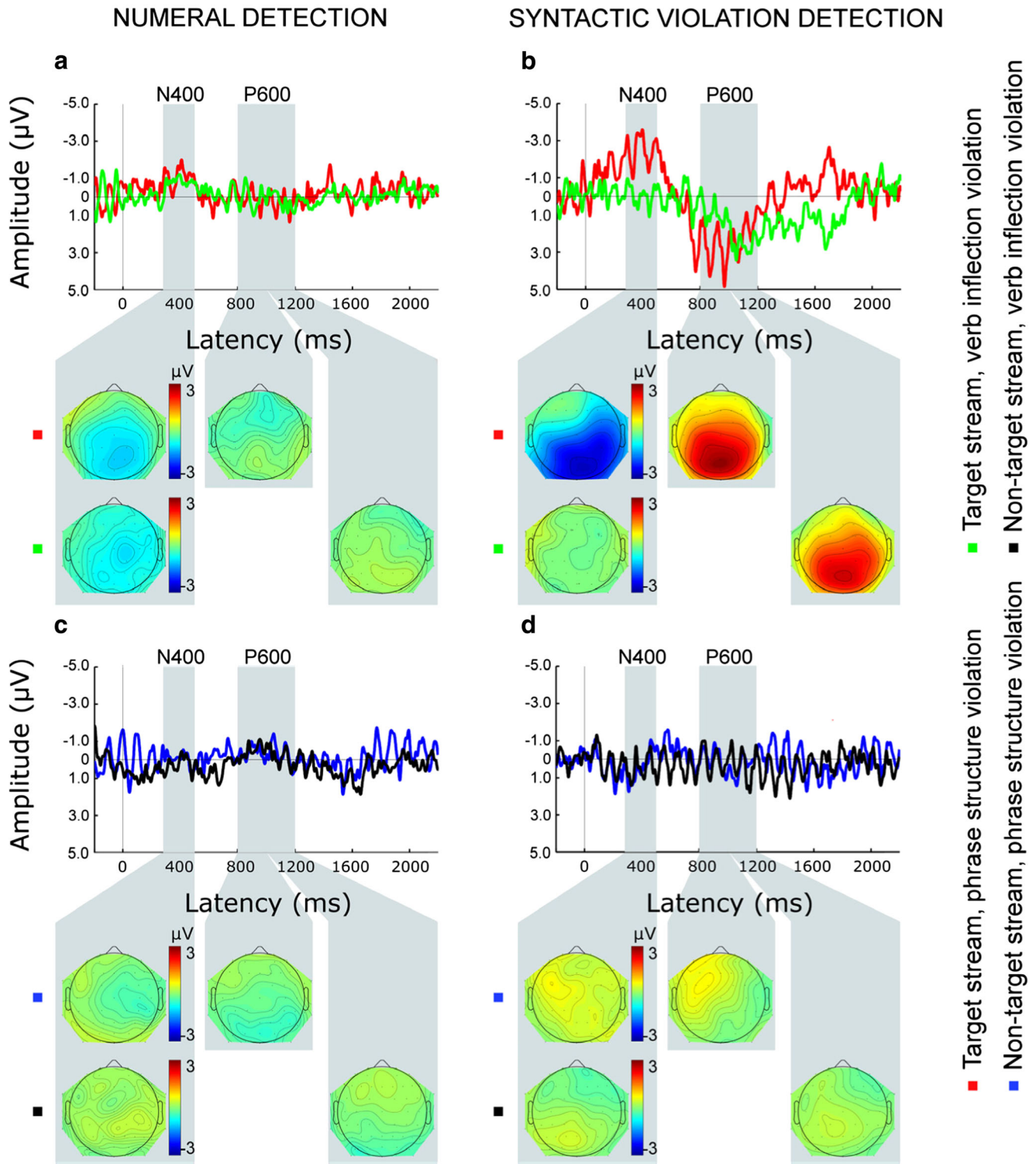
forced on the listener by the task requirement. That is, had they not been instructed to mark these violations, listeners would have simply skipped over these events without detrimental effects on speech comprehension. Thus, reorganization due to verb-inflection mismatch was not directly triggered by the violation itself, but rather mediated by the task set, causing delay in the execution of the process.

For target numerals, we found that N2b and P3 were elicited and their amplitudes showed no significant difference between the two attention conditions. This result is consistent with previous studies investigating the detection of target events (for a review, see Näätänen, 1990) as well as with the lack of N2b difference between focused and divided attention in young adults, such as our participants were (Wild-Wall & Falkenstein, 2010). None of these components were elicited for any of the non-target numerals (i.e., either for task-irrelevant or for distractor ones), except for a significant positive waveform peaking at about 500 ms from stimulus onset (thus earlier than the P300 to target numerals) elicited by attended non-target numerals. This response may mark that some of the numerals were noticed (despite that numerals were task-irrelevant), possibly due to the tracking task: Questions often inquired about numeric information. Therefore, it is likely that while participants detected the syntactic violations they also processed the numerals to some degree. However, behavioral results showed that numeral distractors were confused for targets more often than distractor syntactic violations. Because ERPs were only analyzed for correct rejections, there is no contradiction between the ERP and the behavioral results. The surprising finding of higher tracking performance for speech delivered from the left than from the right loudspeaker in the focused attention condition was probably due to the confound between location and actor. The actor, whose voice was always delivered from the left loudspeaker used more salient prosody than the other actor, as was shown by the difference in dynamic range¹ (~8.5 dB vs. ~4.5 dB for left and right, respectively) between the two voices. This explanation is also supported by the lack of significant recognition memory performance difference in the divided attention condition, in which the voice of both actors was task-relevant. In general, the behavioral results suggest that the attention and task manipulations were successful.

The current study employed a stimulus paradigm that was closer to real-life situations than that presented in most previous investigations of syntax processing: (1) syntactic violations were embedded in long continuous speech segments with coherent meaning; (2) a multi-talker situation was set up; (3) attention and task-relevance were fully crossed in the study design. Therefore, some of the differences in the results compared to previous studies may reflect the operation of

¹ Dynamic range is defined here as the average RMS power difference between the first two words and the rest of the sentence.

Parietal (Pz) ERP responses to syntactic violations in the divided attention condition



speech processing under everyday circumstances as opposed to artificial situations.

The current study also has some limitations, which may restrict the generality of the conclusions. The goal to keep

◀ **Fig. 5** Group-average (N=14) parietal (Pz) ERP responses elicited by syntactic violations in the divided attention condition. Left panels (A and C) show the responses during the numeral detection, right panels (B and D) during the syntactic violation detection task. “Target stream” traces (top half) represent the responses to target syntactic violations (red color for phrase structure violations, green for verb inflection violations); “Non-target stream” traces (bottom half) represent the responses to syntactic violations in the non-target concurrent stream (blue for phrase structure violations, black for verb inflection violations). Scalp topographies for the N400 and P600 are presented below the ERP responses (upper row, red/blue square: phrase structure violation; lower row, green/black square: verb inflection violation). Calibration of the color scale is shown on the right side of the scalp topography maps. Maps were spline interpolated with a smoothing factor of 10^{-7}

the speech segments relatively natural limited the number of syntactic violations that could be delivered in each condition. This has resulted in having to reject a relatively large number of participants due to low numbers of artifact-free events in some categories/conditions and reduced the S/N ratio, especially for the ERPs elicited by syntactic violations. However, collapsing across the two categories of syntactic violations produced qualitatively the same results (see Online Supplementary Fig. 1), which suggests that the data analyzed is reliable. Another constrain imposed by using natural continuous speech was that ERP responses and reaction times were referred to word onsets as opposed to the moment where the target event could be first detected. This has caused some temporal smearing of the ERPs and RTs, again reducing the S/N ratio. However, note that for most cases in which we found no significant evidence for the elicitation of a component, there is a comparable condition in which the same component was significant. This argues against interpreting the lack of the given component as being due to low S/N ratio. Furthermore, the use of 0.5-Hz high-pass filter could have reduced the amplitude of the P600. A recent study investigating the impact of high-pass filters on slower cortical components that are commonly recorded in experiments recording ERP components accompanying linguistic processes (Tanner et al., 2015) found a reduction in P600 amplitude with high cutoff filter values. We assume that since the effect of filtering is not different across task conditions, therefore the observed differences between the conditions cannot be attributed to distortions due to the filter parameters used in the study.

In summary, the current study demonstrated the utility of ERP responses for studying speech processing in multi-talker environments. Syntactic violations in the speech stream for which the participants performed an on-line task elicited N400 and/or P600, irrespective of whether or not the syntactic violations were task-relevant. Thus, some syntactic analyses are performed for any continuously monitored speech stream. However, neither of these components was elicited for syntactic violations occurring in the concurrent speech stream. Thus, the current results do not support the notion of automatic syntactic analysis. The lack of significant P600 response when the

speech stream was continuously monitored but syntactic violations were task-irrelevant may hint at syntactic reorganization being absent in everyday/multi-talker situations. Alternatively, it is possible that task-relevance is a prerequisite of P600 elicitation. Detecting target words elicited the typical ERP components of target detection (N2b and P3). Similar to some previous studies, we found no significant effect of focused versus divided attention on these components in our healthy young adult participants.

Acknowledgements This work was funded by the Hungarian Academy of Sciences’ Lendület project [LP2012-36/2012] awarded to István Winkler. The authors are grateful to Zsuzsanna D’Albini and Zsuzsanna Kovács for collecting the EEG data, to Ágnes Palotás for text editing, László Liszka for audio recording and editing, and to Ferenc Elek and Péter Scherer for voicing the texts.

References

- Bentin, S., Kutas, M., & Hillyard, S. A. (1995). Semantic processing and memory for attended and unattended words in dichotic listening: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 21(1), 54–67. doi:<https://doi.org/10.1037/0096-1523.21.1.54>
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Broadbent, D. E. (1952). Listening to one of two synchronous messages. *Journal of Experimental Psychology*, 44(1), 51–55.
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.
- Bundesden, C. M., & Habekost, T. (2007). Models of attention. In Q. Jing, M. R. Rosenzweig, G. d’Ydewalle, H. Zhang, H.-C. Chen, K. Zhang (Eds.) *Hove : Psychology Press*.
- Caplan, D. & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Brain and Behavioral Sciences*, 22, 77–126. doi:<https://doi.org/10.1017/S0140525X99001788>
- 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, 975–981.
- Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, 8(2), 331–335. doi:<https://doi.org/10.3758/BF03196169>
- Coulson, S., King, J. W. & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language & Cognitive Processes*, 13(1), 21–58. doi:<https://doi.org/10.1080/016909698386582>
- Darwin, C. J. (2008). Listening to speech in the presence of other sounds. *Philosophical Transactions of the Royal Society London B: Biological Sciences*, 363(1493), 1011–1021. doi:<https://doi.org/10.1098/rstb.2007.2156>
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. doi:<https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage*, 34(4), 1443–1449. doi:<https://doi.org/10.1016/j.neuroimage.2006.11.004>

- Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *The Behavioral and Brain Sciences*, *11*, 357–374. doi:<https://doi.org/10.1017/S0140525X00058027>
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23. doi:<https://doi.org/10.1111/1467-8721.00160>
- Eysenck, M. (2012). *Attention and arousal: Cognition and performance*. Springer Science & Business Media, Berlin.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Science* *6*(2), 78–84. doi:[https://doi.org/10.1016/S1364-6613\(00\)01839-8](https://doi.org/10.1016/S1364-6613(00)01839-8)
- Friederici, A. D. (2004). Event-related brain potential studies in language. *Current Neurology and Neuroscience Reports*, *4*(6), 466–470. doi:<https://doi.org/10.1007/s11910-004-0070-0>
- Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research*, *31*(1), 45–63. doi:<https://doi.org/10.1023/A:1014376204525>
- Friederici, A. D., Opitz, B., & von Cramon, D. Y. (2000). Segregating semantic and syntactic aspects of processing in the human brain: An fMRI investigation of different word types. *Cerebral Cortex*, *10*(7), 698–705. doi:<https://doi.org/10.1093/cercor/10.7.698>
- Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics*. Los Altos, CA: Peninsula.
- Gunter, T. C., & Friederici, A. D. (1999). Concerning the automaticity of syntactic processing. *Psychophysiology*, *36*(1), 126–137. doi:<https://doi.org/10.1017/S004857729997155X>
- Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, *34*(6), 660–676. doi:<https://doi.org/10.1111/j.1469-8986.1997.tb02142.x>
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language & Cognitive Processes*, *8*(4), 439–483. doi:<https://doi.org/10.1080/01690969308407585>
- Hagoort, P., Brown, C. M., & Osterhout, L. (1999). *The neurocognition of syntactic processing*. Oxford: Oxford University Press.
- Hagoort, P., Wassenaar, M., & Brown, C. M. (2003). Syntax-related ERP-effects in Dutch. *Cognitive Brain Research*, *16*(1), 38–50. doi:[https://doi.org/10.1016/S0926-6410\(02\)00208-2](https://doi.org/10.1016/S0926-6410(02)00208-2)
- Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11*(2), 194–205. doi:<https://doi.org/10.1162/089892999563328>
- Hohlfeld, A., Martín-Loeches, M., & Sommer, W. (2015). Is semantic processing during sentence reading autonomous or controlled? Evidence from the N400 component in a dual task paradigm. *Advances in Cognitive Psychology* *11*(2), 42–55. doi:<https://doi.org/10.5709/acp-0170-2>
- Hohlfeld, A., Sangals, J., & Sommer, W. (2004). Effects of additional task on language perception: An event-related brain potential investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* *30*(5), 1012–1025. doi:<https://doi.org/10.1037/0278-7393.30.5.1012>
- Jolsvai, H., Sussman, E., Csuhaj, R., & Csépe, V. (2011). Neurophysiology of Hungarian subject-verb dependencies with varying intervening complexity. *International Journal of Psychophysiology*, *82*(3), 207–216. doi:<https://doi.org/10.1016/j.ijpsycho.2011.06.010>
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149. doi:<https://doi.org/10.1037/0033-295X.99.1.122>
- Kocsis, Z., Hajdu, B., Orosz, G., Winkler, I., & Honbolygó, F. (2017). Szintaktikai sértések és a sértés nyelvtani elemei közötti távolság hatásának vizsgálata a magyar nyelvben. *Magyar Pszichológiai Szemle*, *72*(2), 149–162. doi:<https://doi.org/10.1556/0016.2017.72.2.1>
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research* *1146*, 23–49. doi:<https://doi.org/10.1016/j.brainres.2006.12.063>
- Kuperberg, G. R. (2016). Separate streams or probabilistic inference? what the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*, *31*, 602–616. doi:<https://doi.org/10.1080/23273798.2015.1130233>
- 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. doi:<https://doi.org/10.1146/annurev.psych.093008.131123>
- Kutas, M., & Hillyard, S. A. (1983). Event-related brainpotentials to grammatical errors and semantic anomalies. *Memory & Cognition* *11*(5), 539–550. doi:<https://doi.org/10.3758/BF03196991>
- Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II (1994–2005). (M. A. Gernsbacher & M. Traxler, Eds.), *Handbook of Psycholinguistics* (2nd ed.). New York: Elsevier.
- Lewis, J. L. (1970). Semantic processing of unattended messages using dichotic listening. *Journal of Experimental Psychology*, *85*(2), 225–228. doi:<https://doi.org/10.1037/h0029518>
- Maidhof, C., & Koelsch, S. (2011). Effects of selective attention on syntax processing in music and language. *Journal of Cognitive Neuroscience*, *23*(9), 2252–2267. doi:<https://doi.org/10.1162/jocn.2010.21542>
- Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60. doi:<https://doi.org/10.1080/17470215908416289>
- Näätänen, R. (1990). The role of attention in auditory information-processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*(2), 201–232. doi:<https://doi.org/10.1017/S0140525X00078407>
- Näätänen, R., Simpson, M., & Loveless, N. E. (1982). Stimulus deviance and evoked potentials. *Biological Psychology*, *14*(1–2), 53–98. doi:[https://doi.org/10.1016/0301-0511\(82\)90017-5](https://doi.org/10.1016/0301-0511(82)90017-5)
- Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, *59*(3), 494–522. doi:<https://doi.org/10.1006/brln.1997.1793>
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806. doi:[https://doi.org/10.1016/0749-596X\(92\)90039-Z](https://doi.org/10.1016/0749-596X(92)90039-Z)
- Parasuraman, R. (1980). Effects of information processing demands on slow negative shift latencies and N100 amplitude in selective and divided attention. *Biological Psychology*, *11*(3–4), 217–233. doi:[https://doi.org/10.1016/0301-0511\(80\)90057-5](https://doi.org/10.1016/0301-0511(80)90057-5)
- Polich, J. (2007). Updating p300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. doi:<https://doi.org/10.1016/j.clinph.2007.04.019>
- Polich, J., & Herbst, K. L. (2000). P300 as a clinical assay: Rationale, evaluation, and findings. *International Journal of Psychophysiology*, *38*(1), 3–19. doi:[https://doi.org/10.1016/S0167-8760\(00\)00127-6](https://doi.org/10.1016/S0167-8760(00)00127-6)
- Pritchard, W. S., Shappell, S. A., & Brandt, M. E. (1991). *Psychophysiology of N200/N400: A review and classification scheme* (Vol. 4). London: Jessica Kingsley.
- Pulvermüller, F., Shtyrov, Y., Hasting, A. S., & Carlyon, R. P. (2008). Syntax as a reflex: Neurophysiological evidence for early automaticity of grammatical processing. *Brain and Language*, *104*(3), 244–253. doi:<https://doi.org/10.1016/j.bandl.2007.05.002>
- Ritter, W., Simson, R., Vaughan, H. G., & Friedman, D. (1979). Brain event related to the making of a sensory discrimination. *Science*, *203*(4387), 1358–1361. Retrieved from <Go to ISI>://A1979GN50300035
- Schacht, A., Sommer, W., Shmuelovitch, O., Casado Martíenz, P., & Martín-Loeches, M. (2014). Differential task effects on N400 and

- P600 elicited by semantic and syntactic violations. *PLoS One*, 9(3), e91226. doi:<https://doi.org/10.1371/journal.pone.0091226>
- Severens, E., Jansma, B. M., & Hartsuiker, R. J. (2008). Morphophonological influences on the comprehension of subject-verb agreement: An ERP study. *Brain Research*, 1228, 135–144. doi:<https://doi.org/10.1016/j.brainres.2008.05.092>
- Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, 150(3700), 1187–1188. doi:<https://doi.org/10.1126/science.150.3700.1187>
- Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology*, 52(8), 997–1009. doi:<https://doi.org/10.1111/psyp.12437>
- Treisman, A., Squire, R., & Green, J. (1974). Semantic processing in dichotic listening? A replication. *Memory & Cognition*, 2(4), 641–646. doi:<https://doi.org/10.3758/BF03198133>
- Van Petten, C. (1995). Words and sentences: Event-related brain potential measures. *Psychophysiology*, 32(6) 511–525. doi: <https://doi.org/10.1111/j.1469-8986.1995.tb01228.x>
- Verleger, R. (1988). A critique of the context updating hypothesis and an alternative interpretation of P3. *Behavioral and Brain Sciences*, 11, 343–427. doi:<https://doi.org/10.1017/S0140525X00058015>
- Wild-Wall, N., & Falkenstein, M. (2010). Age-dependent impairment of auditory processing under spatially focused and divided attention: An electrophysiological study. *Biological Psychology*, 83(1), 27–36. doi:<https://doi.org/10.1016/j.biopsycho.2009.09.011>
- Zawiszewski, A., & Friederici, A. D. (2009). Processing canonical and non-canonical sentences in Basque: The case of object-verb agreement as revealed by event-related brain potentials. *Brain Research*, 1284, 161–179. doi:<https://doi.org/10.1016/j.brainres.2009.05.099>