

Tracking the Effects of Top–Down Attention on Word Discrimination Using Frequency-tagged Neuromagnetic Responses

Maxime Niesen, Marc Vander Ghinst, Mathieu Bourguignon, Vincent Wens, Julie Bertels, Serge Goldman, Georges Choufani, Sergio Hassid, and Xavier De Tiège

Abstract

■ Discrimination of words from nonspeech sounds is essential in communication. Still, how selective attention can influence this early step of speech processing remains elusive. To answer that question, brain activity was recorded with magnetoencephalography in 12 healthy adults while they listened to two sequences of auditory stimuli presented at 2.17 Hz, consisting of successions of one randomized *word* (tagging frequency = 0.54 Hz) and three acoustically matched *nonverbal* stimuli. Participants were instructed to focus their attention on the occurrence of a predefined *word* in the verbal attention condition and on a *nonverbal* stimulus in the nonverbal attention condition. Steady-state neuromagnetic responses were identified with spectral analysis at sensor and source levels. Significant sensor re-

sponses peaked at 0.54 and 2.17 Hz in both conditions. Sources at 0.54 Hz were reconstructed in supratemporal auditory cortex, left superior temporal gyrus (STG), left middle temporal gyrus, and left inferior frontal gyrus. Sources at 2.17 Hz were reconstructed in supratemporal auditory cortex and STG. Crucially, source strength in the left STG at 0.54 Hz was significantly higher in verbal attention than in nonverbal attention condition. This study demonstrates speech-sensitive responses at primary auditory and speech-related neocortical areas. Crucially, it highlights that, during word discrimination, top–down attention modulates activity within the left STG. This area therefore appears to play a crucial role in selective verbal attentional processes for this early step of speech processing. ■

INTRODUCTION

In our modern societies, humans are exposed daily to a high rate of spoken words, which are often mixed with different simultaneous conversations or with environmental nonspeech background sounds. Still, speech seems to be processed seamlessly. Indeed, language content is quickly recognized, and words are easily discriminated from nonspeech sounds. Nevertheless, to what extent attention influences this process remains a matter of debate (Yoncheva, Maurer, Zevin, & McCandliss, 2014; Shtyrov, Kujala, & Pulvermüller, 2010; Garagnani, Shtyrov, & Pulvermüller, 2009; Sabri et al., 2008). In particular, how selective auditory attention influences the first steps of speech processing (i.e., phonological, syllabic, or lexical processing) has been the topic of several functional neuroimaging studies relying on various experimental designs and which led to conflicting findings.

Some fMRI studies demonstrated that actively attending to words (compared with passive listening) increases response amplitude in specific laterotemporal neocortical areas around the STS, such as the superior temporal gyrus (STG) and the middle temporal gyrus (MTG; Woods, Herron, Cate, Kang, & Yund, 2011; Sabri et al., 2008; Alho

et al., 2006; Rimol, Specht, & Hugdahl, 2006). Other studies have shown that activity within these brain areas responds more strongly to words than to meaningless speech sounds (e.g., pseudowords, rotated speech, or reversed speech), especially during active listening and less so during passive listening (Harinen & Rinne, 2013; Sabri et al., 2008; Binder et al., 2000). Still, although fMRI has an excellent spatial resolution, it suffers from a limited temporal resolution precluding the study of the spectral and temporal dynamics of auditory inputs neural processing (for a review, see Poeppel & Hickok, 2015). Therefore, time-sensitive electrophysiological techniques, such as MEG and EEG, appear better suited to uncover the spatial, spectral, and temporal dynamics of top–down attentional effects on the first steps of neural speech processing (for a review, see Hari & Salmelin, 1997).

Recent electrophysiological data have shown that responses elicited by syllabic encoding is enhanced by focusing attention on speech auditory stream (Batterink & Paller, 2019) and that selective attention increases responses evoked by phonological processes (Yoncheva et al., 2014). Other evidence further suggests that applying lexical knowledge to group monosyllabic morphemes into meaningful disyllabic words requires selective auditory attention, but that monosyllabic morpheme encoding per se is much less influenced by top–down attentional processes

(Ding et al., 2018). This latter finding is supported by scalp EEG data demonstrating that the neocortical response at the syllabic rate is preserved during sleep, while parsing of higher order linguistic structures (i.e., words, sentences) is disrupted (Makov et al., 2017).

By contrast, results obtained from MMN paradigms suggest that the early stages of word processing are automatic (MacGregor, Pulvermüller, van Casteren, & Shtyrov, 2012; Shtyrov et al., 2010, 2012; Garagnani et al., 2009; Pulvermüller & Shtyrov, 2006; Shtyrov, Pihko, & Pulvermüller, 2005; Pulvermüller, Shtyrov, Kujala, & Näätänen, 2004). Briefly, these MMN studies relied on oddball paradigms with acoustically matched word and pseudoword stimuli. Auditory MMN responses to words were stronger compared with pseudowords confirming early activation of lexical memory traces (Shtyrov et al., 2005, 2010; Shtyrov, Osswald, & Pulvermüller, 2008; Pulvermüller & Shtyrov, 2006; Endrass, Mohr, & Pulvermüller, 2004; Pulvermüller et al., 2004; Shtyrov & Pulvermüller, 2002; Korpilahti, Krause, Holopainen, & Lang, 2001; Pulvermüller et al., 2001). Interestingly, only the late components (>200 msec) of MMN responses to words were influenced by top-down attention, arguing for the independence of the earlier stages of word processing from top-down attentional control and suggesting an automatic brain lexical access (MacGregor et al., 2012; Shtyrov et al., 2010, 2012; Garagnani et al., 2009; Garagnani, Wennekers, & Pulvermüller, 2008). Based on these conflicting electrophysiological data, it is unclear if syllabic and word neocortical processing are automatic or rather influenced by top-down attention.

In this study, we used the frequency tagging approach to bring novel insights into the role of top-down attention on the early stages of speech processing, that is, monosyllabic word processing and discrimination. The frequency tagging approach has been used to identify steady-state responses that are classically obtained by sensory stimulations periodically modulated in amplitude or frequency (Galambos, Makeig, & Talmachoff, 1981). Specific neural responses rise at the modulation frequencies, that is, “frequency tagging.” These neural responses illustrate the ability of the brain activity to “resonate” with the stimulus modulation rate (Picton, John, Dimitrijevic, & Purcell, 2003) and to track multiple simultaneous stimuli modulated at different frequencies (Ross, Borgmann, Draganova, Roberts, & Pantev, 2000; Lins & Picton, 1995). One of the major advantages of the frequency tagging approach is the ability to build strong a priori hypotheses about the relationship between a specific sensory stimulus and its related neural response (for a review, see, e.g., Norcia, Appelbaum, Ales, Cottureau, & Rossion, 2015). This method has been used to probe attention to speech using amplitude-modulated spoken vowels and reported that selective attention can affect the neural responses to speech streams (Bharadwaj, Lee, & Shinn-Cunningham, 2014). Interestingly, this approach can be mixed with a periodic oddball design by embedding a rare oddball stimulus in a continuous train of standard stimuli. Com-

pared with the classic MMN paradigm, it allows for objective identification of any response at the frequency of the oddball stimulus, without having to subtract responses to standard and oddball stimuli to highlight the discrimination response (Norcia et al., 2015). This method is very efficient and has been used in the visual modality (Norcia et al., 2015) to study face categorization (Rossion, Torfs, Jacques, & Liu-Shuang, 2015) and discrimination between words and nonwords (Lochy, Van Belle, & Rossion, 2015). It has also been used in the auditory domain to study auditory processing of frequency and interaural time contrast (Nozaradan, Mouraux, & Cousineau, 2017), tracking of musical beat (Lenc, Keller, Varlet, & Nozaradan, 2018; Nozaradan, 2014; Nozaradan, Peretz, & Mouraux, 2012; Nozaradan, Peretz, Missal, & Mouraux, 2011), and investigating the neural processes involved in the detection of regularities embedded in auditory streams (Farthouat, Atas, Wens, De Tieghe, & Peigneux, 2018; Farthouat et al., 2017).

The present MEG study relies on a frequency tagging oddball auditory design. We presented binaurally two different types of auditory stimuli, that is, monosyllabic *words* and *nonverbal* sounds, in rhythmic trains of stimuli. Auditory stimuli were presented at frequency f of 2.17 Hz, in series of one *word* (i.e., “oddball” stimuli) and three *nonverbal* stimuli (i.e., oddball frequency: $f/4 = 0.54$ Hz). We expected trains of auditory stimuli to elicit neocortical responses at auditory stimuli frequency but also at *word* stimuli frequency and its harmonics in speech sensitive brain areas. To investigate the effects of top-down attentional processes on these frequency-tagged responses, we manipulated selective auditory attention toward either *word* or *nonverbal* stimuli. Specifically, we hypothesized that focusing attention on *word* stimuli would increase neocortical responses at *word* frequency (i.e., 0.54 Hz), demonstrating that lexical access is an attention-sensitive process while focusing attention on nonverbal stimuli would decrease or abolish the amplitude of these responses.

METHODS

Participants

Thirteen healthy adult volunteers (seven women, age range = 24–33 years, mean age = 26 years) were included in this study. All participants were native French speakers, right-handed according to the Edinburgh Handedness Inventory questionnaire (score range = 45–100, mean score = 73; Oldfield, 1971). They all had normal hearing according to pure-tone audiometry (normal hearing threshold = 0–20 dB HL for all octave frequencies from 250 to 8000 Hz) and had no history of neurological or psychiatric disease. This study was approved by the Ethics Committee of the CUB-Hôpital Erasme. All participants gave written informed consent before their inclusion in the study.

One participant was excluded because of intractable MEG artifacts, leading to a final data set of 12 participants; a sufficient sample size based on previous studies using frequency tagging paradigms in auditory (Farthouat et al., 2017) or visual (Rossion et al., 2015) modalities.

Stimuli

Two types of auditory stimuli were used in this study: *word* and *nonverbal* stimuli, all lasting 460 msec.

As the power spectrum of MEG signals is dominated by low-frequency noise due to background brain activity and sensor noise (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993), the frequency tagging approach is best performed with brief stimuli. Accordingly, we chose monosyllabic words to avoid compromising on word intelligibility. *Word* stimuli consisted of 10 spoken consonant–vowel–consonant French words (i.e., *belle*, *coeur*, *compte*, *face*, *femme*, *longue*, *peine*, *peur*, *route*, *vite*) recorded by a native French speaker and downloaded from a free audio database (shtooka.net/). These words are part of the 600 most used words in the French language according to a French lexical database (Lexique 3.1, www.lexique.org; New, Pallier, Brysbaert, & Ferrand, 2004). They were recorded with AUDACITY (v2.0.6, audacity.sourceforge.net/) and encoded as 16 bit-mono WAV files (sampling rate = 44,100 Hz).

Nonverbal stimuli consisted mainly in *phase-scrambled* stimuli, but also in *noise* stimuli. *Phase-scrambled* stimuli consisted of 10 unintelligible sounds that were matched with *word* stimuli on spectral content and temporal envelope. The temporal envelope is important for (i) keeping the same power as the forward speech sound, (ii) keeping similar onset/offset dynamics, and (iii) because it conveys many linguistic cues like articulation and voicing (Rosen, 1992). To generate these *phase-scrambled* stimuli, the

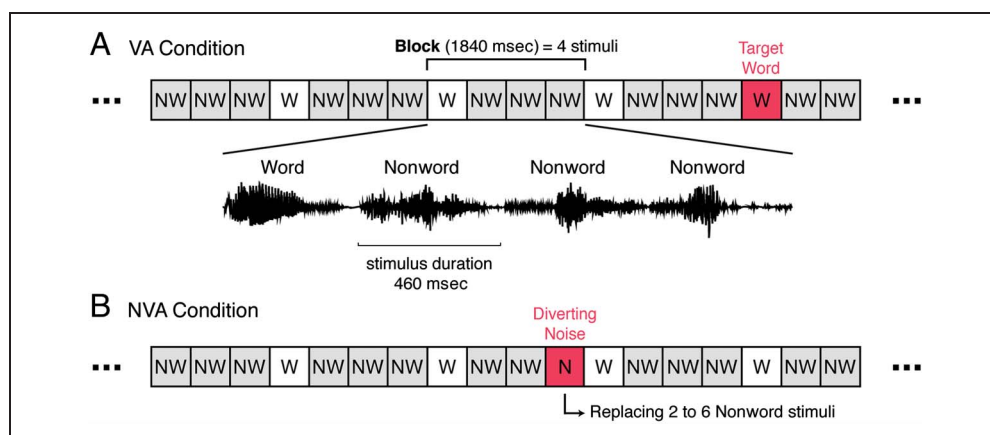
envelope of *word* stimuli was first extracted by rectifying and then low-pass filtering (<50 Hz) the forward speech signals. A phase-scrambled version of the word sound was then obtained using Fourier phase randomization (Faes, Pinna, Porta, Maestri, & Nollo, 2004) and finally modulated by the original word envelope. The *noise* stimuli (Perrin & Grimault, 2019) were characterized by a random phase and a power spectrum set to coincide with that of an average speech spectrum, with a flat spectral distribution until 1 kHz and then decreasing by 12 dB/octave following ANSI norm's specifications (1989).

Experimental Paradigm

Participants sat comfortably in the magnetically shielded room of the MEG with their arms resting on a table. They were instructed not to move or speak and to gaze at a cross drawn on the wall facing them. Participants underwent three conditions. The first condition was a 5-min rest (i.e., stimulation-free) during which they were asked to relax and stay still. The two next conditions were the “verbal attention” (VA) and the “nonverbal attention” (NVA) conditions. Their order was randomized across participants. They lasted 6 min and consisted of a continuous sequence of 200 stimulation blocks, each made up of four basic auditory stimuli without any acoustic gap (Figure 1). In both (VA and NVA) conditions, the first stimulus of each block was randomly selected among *word* stimuli, and the three subsequent stimuli were randomly taken from *phase-scrambled* stimuli. In addition, in the NVA condition, two to six *phase-scrambled* stimuli among the 200 blocks were randomly replaced by *noise* stimuli with a maximum of one *noise* stimulus per block. Therefore, VA and NVA sequences consisted of 800 stimuli each and differed by only two to six stimuli. The task was, however, different between conditions as it

Figure 1. (A) Stimulation paradigm for the VA condition. Each block was composed of four basic stimuli (each lasting 460 msec) without any acoustic gap. In total, 200 blocks were presented (i.e., 800 stimuli) in a continuous sequence of ~6 min. The first stimulus of each block was a *word* (W) randomly chosen from 10 monosyllabic words and the three subsequent stimuli were randomly chosen from 10 nonverbal (i.e., *phase-scrambled*) stimuli (NW). The task was, however, different as it manipulated the target of

selective auditory attention. In this condition, participants were asked to count the occurrences of a target word stimulus (highlighted in red, mean occurrence across the 200 stimulation blocks: 20 ± 4.5 SD, random selection of the target word across participants, constant target word per participant). (B) Stimulation paradigm for the NVA condition. This condition was identical to VA except that participants were asked to count the occurrence of *noise* stimuli (*n*, highlighted in red, mean occurrence: 4.5 ± 1.4 SD) randomly replacing *phase-scrambled* stimuli to divert attention from verbal stimuli.



manipulated the target of selective auditory attention. In the VA condition, participants were asked to count the occurrences of a target *word* stimulus (mean occurrence across the 200 stimulation blocks: 20 ± 4.5 SD, random selection of the target word across participants, constant target word per participant). In the NVA condition, participants were asked to count the occurrence of the *noise* stimulus (mean occurrence: 4.5 ± 1.4 SD). Although questionable, we decided to include less *noise* stimuli in the NVA condition than target words in the VA condition to minimize the allocation of attention to *word* stimuli in the NVA condition. Indeed, it has been suggested that low-probability targets involve more attentional demands and are less sensitive to attentional capture compared with high-probability targets (Hon, Ng, & Chan, 2016; Hon & Tan, 2013).

Auditory stimuli were delivered through a 60×60 cm² high-quality flat panel loudspeaker (Panphonics SSH sound shower) placed 3 m away in front of the participant, at an average sound intensity of 60 dB as assessed by a sound level meter (Sphynx Audio System). At the end of each session, participants were probed on their count of the target *word* (VA condition) or *noise* (NVA condition) stimuli and to rate on separate visual analog scales their level of wakefulness and word identification, respectively (from 0 to 10, 0 = *asleep/tough*, 10 = *awake/effortless*).

Data Acquisition

Neuromagnetic data were recorded with a 306-channel whole-scalp covering MEG system installed in a light weight magnetically shielded room (Vectorview and MaxShield, MEGIN, Croton Healthcare); the characteristics of which have been described elsewhere (De Tiège et al., 2008). The MEG array is composed of 102 triplets of sensors, each consisting of one magnetometer and two orthogonal planar gradiometers. These MEG sensors are maximally sensitive to right beneath (planar gradiometers) or nearby (magnetometers) neural sources. Signals were filtered through 0.1–330 Hz and sampled at 1 kHz. Participants' head position inside the MEG helmet was continuously monitored using four head-tracking coils. The location of these coils and at least 250 head-surface points (on the nose, face, and scalp) with respect to anatomical fiducials were determined with an electromagnetic tracker (Fastrak, Polhemus). EOGs and electrocardiograms were also recorded time-locked to MEG signals.

Participants' 3D-T1 cerebral magnetic resonance image (repetition time = 9.8 msec, echo time = 4.6 msec, flip angle = 8°, resolution = 1 mm \times 0.795 mm \times 0.795 mm) was acquired after the MEG sessions on a 1.5 T MRI scan (Intera, Philips).

Data Processing and Analyses

Analyses were performed using in-house Matlab scripts (The MathWorks).

Preprocessing

MEG data were first preprocessed offline with the signal space separation method using the Maxfilter software (Maxfilter, MEGIN, Croton Healthcare; Version 2.2 with default parameters) to subtract external interferences and correct for head movements (Taulu, Simola, & Kajola, 2005). Ocular and cardiac artifacts were then removed using independent component analysis and temporal correlations between components, EOG, and electrocardiogram signals (Vigário, Särelä, Jousmäki, Hämäläinen, & Oja, 2000). The cleaned continuous MEG and sound signals were then split into 20.24-sec-long segments (corresponding to exactly 11 blocks). Epochs of MEG signals for which a channel exceeded 4 pT (magnetometers) or 1 pT/cm (planar gradiometers) were finally rejected from further analysis to avoid contamination of our data by high amplitude system artifacts.

Spectral Analysis

A fast Fourier transform was applied to data segments. Frequency resolution (i.e., the separation between adjacent frequency bins) was $1/(20.24 \text{ sec}) = 0.0494$ Hz. For each participant, condition and gradiometer sensor only, amplitude spectra were obtained as the modulus of the averaged Fourier-transformed sequences. These amplitude spectra were further combined within gradiometer pairs using the Euclidian norm. Spectra of signal-to-noise ratio (SNR) were then computed as the ratio between the amplitude at each frequency bin and the average amplitude at the 12 surrounding frequency bins (six on each side, excluding the immediately adjacent bins; see Barry-Anwar, Hadley, Conte, Keil, & Scott, 2018; Peykarjou, Hoehl, Pauen, & Rossion, 2017). SNR values present the valuable advantage of being maximally insensitive to the spectral properties of the noise.

Two sharp peaks of SNR were expected a priori in the SNR spectra in VA and NVA conditions: one at 2.17 Hz (f) due to the neocortical processing of each auditory (i.e., *word* and *nonverbal*) stimulus because they were presented every 460 msec, and another at 0.54 Hz ($f/4$) attributable to the neocortical detection/processing of *word* stimuli because they were presented every 1.84 sec (4×460 msec). Moreover, because the auditory stimulation was not a pure sine wave and because brain processing is not linear even if the stimulation is sinusoidal (Lochy et al., 2015; Regan, 1989), we expect to see harmonics at 1.09 Hz ($2 \times f/4$) and 1.63 Hz ($3 \times f/4$). Based on the hypothesis that these responses are linked (Nozaradan et al., 2017; Retter & Rossion, 2016), we averaged the SNR across 0.54, 1.09, and 1.63 Hz (f_{olds}).

Source Reconstruction

To locate the brain regions involved in sensors responses, we performed a similar spectral analysis at the source

level. To do so, individual MRIs were first segmented using the Freesurfer software (Martinos Center for Biomedical Imaging; Reuter, Schmansky, Rosas, & Fischl, 2012) and coregistered with MEG coordinate system using the three anatomical fiducial points for initial estimation and the head surface points to manually refine the surface coregistration. Then, a nonlinear transformation from individual MRIs to the Montreal Neurological Institute (MNI) brain was computed using the spatial normalization algorithm implemented in statistical parametric mapping (Ashburner & Friston, 1999; Ashburner, Neelin, Collins, Evans, & Friston, 1997). This transformation was used to map a homogenous 5-mm grid sampling of the MNI brain volume onto individual participants' brain. For each participant and grid point, a 3-D lead field (corresponding to the three spatial dimensions) was computed using one-layer boundary element model estimated with the MNE software suite (Gramfort et al., 2014) and further reduced to its two first principal components. Source time courses were then reconstructed based on the data from all sensors using MNE (Hämäläinen, Lin, & Mosher, 2010; Dale & Sereno, 1993), with noise covariance estimated from the resting-state data and regularization parameter fixed as a function of SNR using the prior consistency condition (Wens et al., 2015). We did not explicitly correct for the depth bias in these source data because our measure of SNR is scale invariant and thus unaffected by this bias. In fact, in this context, our source projection pipeline is rigorously equivalent to noise-normalized versions of MNE, such as standardized low-resolution tomography (Pascual-Marqui, 2002).

A spectral analysis identical to that conducted on sensor space data was used to estimate the SNR of the reconstructed sources at selected frequencies (f and f_{odds}).

Statistical Assessments

Behavioral Evaluation

Because some behavioral data were not normally distributed as indicated by Shapiro–Wilk tests (both $p < .05$ for identification ratings and wakefulness ratings, $p = .081$ for count rate errors), we performed a Wilcoxon signed-rank test to analyze behavioral results and identify statistical differences between the two conditions (i.e., VA and NVA). The significance threshold was set to $p < .05$, and count rate error was defined as the absolute value of the difference between correct count of items and participants' answer.

Identification of Auditory and Verbal Neocortical Responses

For each participant and condition, a nonparametric permutation-like test was used to estimate the statistical significance of the SNR at *word* frequencies (Nichols & Holmes, 2001). It specifically tested the null hypothesis

that responses to *word* stimuli were similar to those to all stimuli.

The test sought for significant response in all sensors at once. The response considered was that at the *word* frequency and its two first harmonics (0.54, 1.09, and 1.63 Hz) in a first step and at each of these individual frequency bins separately in a second step. Practically, the genuine SNR averaged across tested frequencies (one value per sensor) was computed based on sequences in which either the first or last cycle was removed. A permutation distribution was then built by estimating 1000 times the maximum—across all sensors—of the mean SNR across tested frequency bins derived from sequences randomly trimmed by a duration corresponding to the $n = 0, 1, 2$, or 3 first sounds ($n \times 460$ msec) and $4-n$ last sounds. Doing so randomized the position of the *word* stimuli while preserving synchrony in auditory stimulation across sequences. The significance of the response at each sensor was obtained as the proportion of values in the permutation distribution that were above the genuine SNR value. This test, being akin to a permutation test (Nichols & Holmes, 2001), is exact and intrinsically deals with the multiple comparison issue.

The statistical significance of the SNR at the group level was assessed with a nonparametric permutation test (Nichols & Holmes, 2001). In practice, participant- and group-level rest SNR maps were obtained as done for VA and NVA conditions based on random extraction of non-overlapping 20.24 segments. Group-level difference maps were obtained by subtracting VA (or NVA) and rest SNR maps. Under the null hypothesis that SNR maps are the same whatever the experimental condition, the labeling VA (or NVA) and rest are exchangeable at the participant level before group-level difference map computation (Nichols & Holmes, 2001). To reject this hypothesis and to compute a threshold of statistical significance for the correctly labeled difference map for each hemisphere separately, the permutation distribution of the maximum absolute value of the difference map in each hemisphere was computed for the exhaustive set of permutations ($2^{12} = 4096$ permutations). The test assigned a p value to each voxel equal to the proportion of permutation values exceeding the genuine value in this voxel (Nichols & Holmes, 2001). We further identified the coordinates of local maxima in group-level SNR maps. Such local maxima are sets of contiguous voxels displaying higher SNR values than all neighboring voxels. We only report statistically significant local maximal of SNR, disregarding the extents of these clusters. Indeed, cluster extent is hardly interpretable in view of the inherent smoothness of MEG source reconstruction (Bourguignon, Molinaro, & Wens, 2018; Wens et al., 2015; Hämäläinen & Ilmoniemi, 1994).

Identification of Attentional Effects

To evaluate the effect of selective auditory attention and identify cortical areas in which SNR values are higher in

VA than NVA conditions at f and f_{odds} separately, we compared SNR maps using the same nonparametric permutation test described above, but between NVA and VA conditions instead of rest condition, leading to a VA—NVA difference map.

RESULTS

Behavioral Results

Table 1 provides an overview of the behavioral data in VA and NVA conditions. The Wilcoxon tests did not reveal any significant difference between the two conditions in count rate errors, wakefulness ratings, or identification ratings (p s > .05).

Auditory Simulation

Figure 2 illustrates the group-level amplitude spectra of the sound envelope (Figure 2A) and SNR spectra in both conditions (Figure 2B, C).

As expected given our experimental design, the sound envelope exhibited a clear peak at 2.17 Hz corresponding to the frequency of occurrence of auditory stimuli, independently of their type. This peak was also clearly disclosed in the neuromagnetic activity during auditory stimulation, in both VA ($p < .05$ for all participants) and NVA conditions ($p < .05$ for 11 of the 12 participants).

As illustrated in Figure 3A (left), the SNR at stimulation frequency was significant at MEG sensors covering bilateral temporal areas. In cortical maps, the SNR peaked in the right STG and left supratemporal auditory cortex in both conditions (see Figure 3A, right). Suprathreshold local maxima of SNR, their MNI coordinates, and their statistical significance are presented in Table 2.

Verbal Stimulation

The amplitude spectrum of sound envelope did not show any peak at word frequency and harmonics (Figure 2A). This ensures that *word* and *phase-scrambled* stimuli were spectrally undistinguishable and that responses at these frequencies can be securely ascribed to neocortical auditory processes that are sensitive to the contrast between monosyllabic words and nonwords.

As expected, a peak of SNR at 0.54 Hz corresponding to *word* presentation rate was also clearly observed in

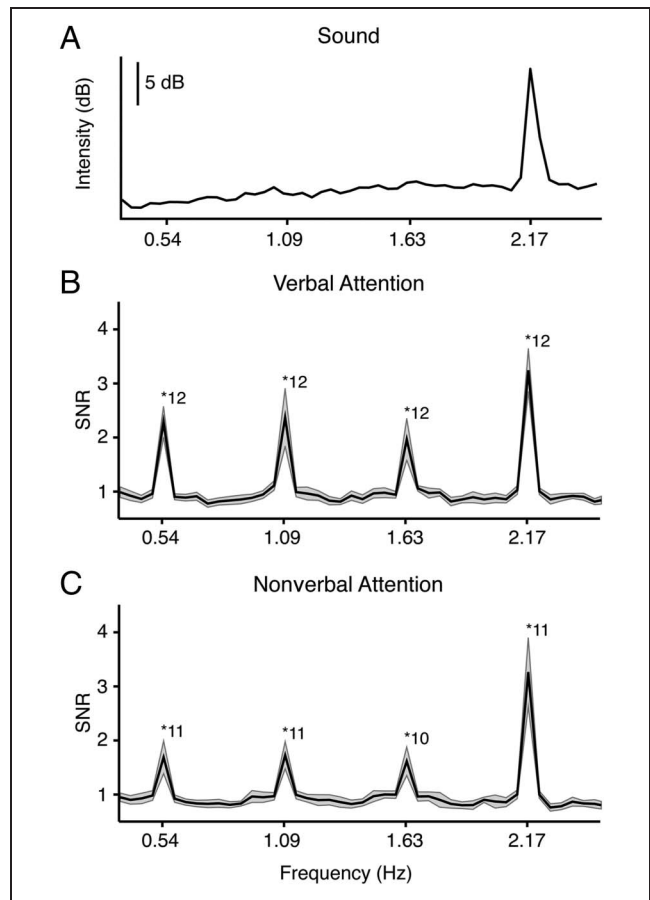


Figure 2. Group-level amplitude spectra. (A) Amplitude spectrum of sound envelope. (B, C) Mean SNR spectrum across all gradiometer pairs in VA (B) and in NVA (C). The black traces and gray-shaded areas indicate the mean and *SD* across participants. The number of participants showing significant response ($p < .05$) is indicated next to each peak.

both conditions ($p < .05$ for all participants in VA and for 11 of the 12 participants in NVA). Peaks of SNR were also significant at harmonics of the verbal stimulation, that is, at 1.09 Hz ($p < .05$ for all participants in VA and for 11 participants in NVA) and 1.63 Hz ($p < .05$ for all participants in VA and for 10 participants in NVA).

The SNR at f_{odds} was significant at sensors over bilateral temporal areas in VA and NVA (see Figure 3B). In source maps, the SNR peaked at bilateral STG in both conditions, as well as in the left inferior frontal gyrus (IFG) in NVA. All significant local maxima of SNR at 0.54, 1.09, and 1.63 Hz (Table 2) localized at the same cortical areas, that is, bilateral STG, left IFG, and left MTG.

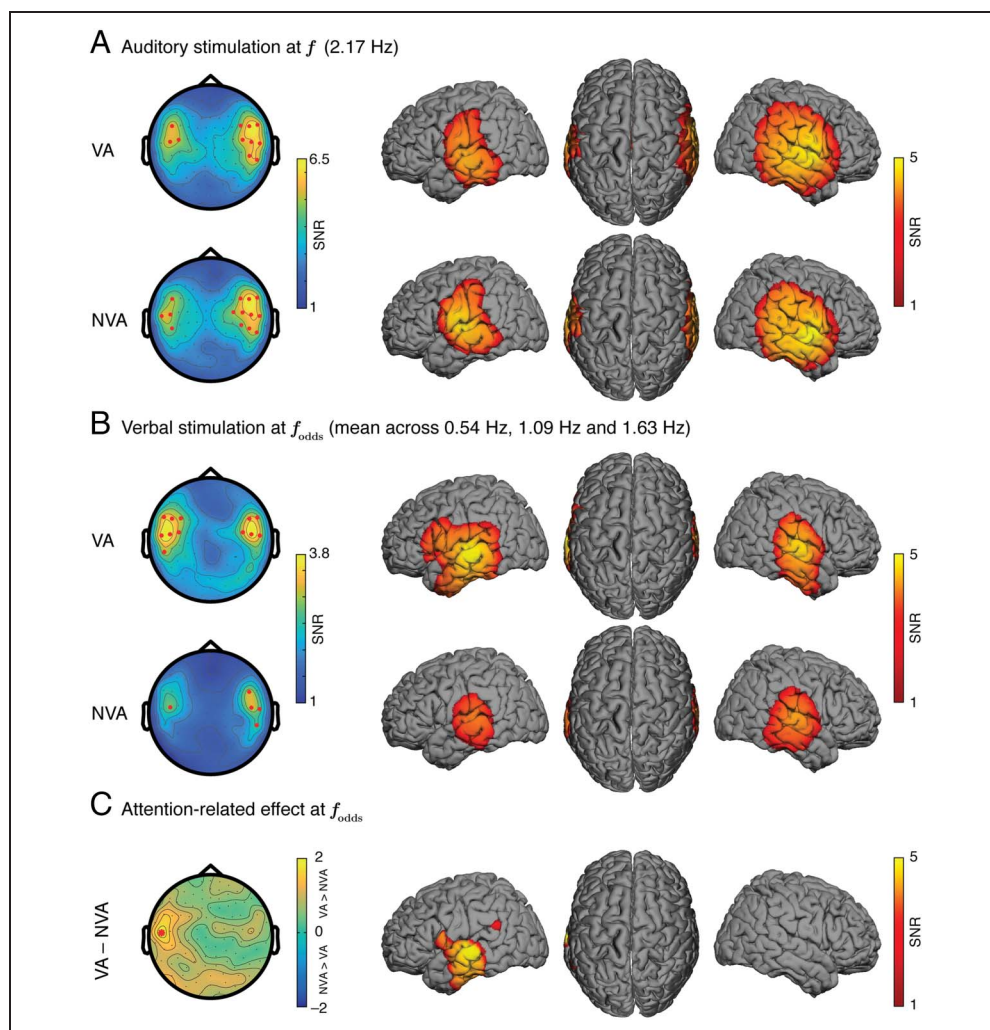
Attention-related Effect

Comparing SNR values between VA and NVA at f_{odds} revealed a significant effect of the target of attention (Figure 3C). Indeed, the amplitude contrast was stronger in VA than in NVA at one left temporal sensor ($p < .004$). Source reconstruction localized the selective auditory attention-related modulation of SNR contrast to the left posterior STG.

Table 1. Behavioral Results

	Error Count Rate		Wakefulness Ratings		Identification Ratings	
	Mean	SD	Mean	SD	Mean	SD
VA	1	1.86	7.5	1.68	9.28	0.51
NVA	0.58	0.79	7.5	1.83	8.75	1.42

Figure 3. Sensor and source distribution of the SNR. (A) SNR at stimulation frequency ($f = 2.17$ Hz) in both VA and NVA conditions. (B) SNR at frequencies tagging word processing (f_{odds} , i.e., the mean SNR across 0.54, 1.09, and 1.63 Hz). (C) The attentional effect (contrast between VA and NVA) at f_{odds} . Significant sensors are marked with a red star ($p < .05$). Cortical maps are thresholded at statistically significance level ($p < .05$) corrected for multiple comparisons. Upperscale thresholds are set to the maximum SNR value across VA and NVA (A and B) or to the maximum of the absolute value of its contrast (C). Please see Table 2 for exact coordinates of peak SNR.



As expected, no differences were found in SNR values between conditions at 2.17 Hz ($p > .9$). This indicates that participants processed auditory stimulations with similar attention in both conditions.

DISCUSSION

The present MEG study used a frequency tagging oddball auditory paradigm to examine the effects of top-down attentional processes on word discrimination. Findings highlight the existence of neural responses specific to the discrimination of words at primary and nonprimary auditory and speech-related neocortical areas. Importantly, they also demonstrate a clear effect of selective auditory attention on these neocortical auditory processes sensitive to words, especially at the left STG.

Cortical Discrimination of Verbal versus Nonverbal Sounds

Significant steady-state responses at 2.17 Hz were related with the spectral peak of the auditory stimulation, evidencing that these responses reflect unspecific (i.e., verbal and

nonverbal) auditory processing. This unspecific auditory processing was generated mainly at supratemporal auditory cortex and STG in both conditions. Although these areas are known to be involved in the earliest stages of cortical speech processing that entail spectrotemporal analyses of *verbal* stimuli (Hickok & Poeppel, 2007), our finding of *unspecific* auditory processing is not that surprising. Indeed, both types of stimuli (i.e., *words* and *phase-scrambled*) share similar spectrotemporal features. Still, previous studies have shown that the acoustic information is processed within central auditory pathways until it reaches cortical speech-sensitive areas where speech content is evaluated and analyzed (Osnes, Hugdahl, Hjelmervik, & Specht, 2011).

More interestingly, neuromagnetic signals revealed an additional steady-state response at 0.54 Hz (and harmonics), corresponding to the occurrence rate of monosyllabic *word* stimuli. Of importance, such a peak was absent from the amplitude spectrum of sound envelope, demonstrating that this response is specific to neocortical auditory processes involved in the discrimination of words. Given that *phase-scrambled* stimuli share similar spectrotemporal properties with *word* stimuli but have

Table 2. Anatomical Region, Coordinates, SNR Values, and Significance of the Local Maxima

Regions	MNI Coordinates			SNR	p
	x	y	z		
VA					
2.17 Hz					
R STAC	64	-1	4	7.28	<10 ⁻³
L STG	-60	-20	-3	6.24	.002
Mean 0.54 Hz and harmonics (<i>f</i> _{odds})					
L STG	-64	-23	-2	4.31	<10 ⁻³
R STG	54	-14	2	4.11	<10 ⁻³
0.54 Hz					
L STG	-53	-20	-1	4.57	<10 ⁻³
R STAC	45	-15	1	3.75	.005
1.09 Hz (1st harmonic)					
R STG	52	-14	4	4.93	<10 ⁻³
L STG	-66	-22	-2	4.57	.002
1.63 Hz (2nd harmonic)					
L MTG	-65	-25	-4	3.90	<10 ⁻³
R STG	61	-13	3	3.73	<10 ⁻³
L IFG	-60	11	13	3.33	.008
NVA					
2.17 Hz					
R STG	64	-2	-2	7.71	<10 ⁻³
L STAC/STG	-62	-13	10	7.17	<10 ⁻³
R STG (post)	65	-32	17	6.33	.002
R STG (post)	64	-31	19	6.33	.002
Mean 0.54 Hz and harmonics (<i>f</i> _{odds})					
R STG	62	-17	6	3.51	<10 ⁻³
L STG	-62	-27	6	3.18	<10 ⁻³
0.54 Hz					
R STAC	63	-17	10	3.76	<10 ⁻³
L STG	-57	-27	6	3.45	.002

Table 2. (continued)

Regions	MNI Coordinates			SNR	p
	x	y	z		
1.09 Hz (1st harmonic)					
R STG	62	-17	4	3.89	<10 ⁻³
L STG	-65	-27	6	3.28	.014
1.63 Hz (2nd harmonic)					
R STG	57	-18	2	2.96	<10 ⁻³
L STG	-65	-27	7	2.85	.002
L IFG	-55	21	6	2.57	.038
Effect of Attention					
VA - NVA					
Mean 0.54 Hz and harmonics (<i>f</i> _{odds})					
L STG	-64	-23	-2	4.31	<10 ⁻³
2,17 Hz					
<i>ns</i>					
NVA - VA					
Mean 0.54 Hz and harmonics (<i>f</i> _{odds})					
<i>ns</i>					
2,17 Hz					
<i>ns</i>					

STAC = supratemporal auditory cortex; STG = superior temporal gyrus; MTG = middle temporal gyrus; IFG = inferior frontal gyrus.

no phonological or lexical content, discrimination of words was most probably based on the detection of such content. In any case, the emergence of a peak at *word* frequency supports our hypothesis that different neocortical responses emerge depending on the nature of the auditory stimuli.

The dominant brain areas involved in 0.54-Hz responses in NVA and VA conditions were (i) bilateral STG, (ii) the left anterior MTG, and (iii) the left IFG; areas involved in different stages of speech processing (Hickok & Poeppel, 2007). The STG is known to be involved in the phonological processing of verbal stimuli (Mesgarani, Cheung, Johnson, & Chang, 2014; Chang et al., 2010). Previous studies have also highlighted stronger activity in response to verbal than nonverbal stimuli in the left posterior STS and midposterior STG, in line with the view that these regions are involved in phonetic and phonological processing, or at least sensitive to the

recognition of verbal stimuli (Möttönen et al., 2006; Dehaene-Lambertz et al., 2005; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005). Phonetic and nonphonetic contents recruit similar brain areas (i.e., STG and STS), but phonetic sounds (Osnes et al., 2011; Specht, Osnes, & Hugdahl, 2009) and words (DeWitt & Rauschecker, 2012; Turkeltaub & Coslett, 2010) activate these areas more strongly. The present results support the view that these areas could serve as key nodes in the verbal language network, which would process every sound in search of phonolexical content (as suggested by Specht et al., 2009) and thereby discriminate verbal stimuli from nonverbal ones. The left anterior MTG is known to be activated when listening to spoken words (DeWitt & Rauschecker, 2012; Peelle, 2012; Visser, Jefferies, & Lambon Ralph, 2010). Furthermore, according to the dual-stream model (Hickok & Poeppel, 2007), this region is involved in the combinatorial network of language perception, that is, the integrative processes that contribute to create syntactic and semantic bonds between words to extract meaning. Finally, cortical responses were also found at the left IFG, a brain region known for its major role in lexical and semantic access (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997). This area can be activated during target detection tasks (Steinschneider et al., 2014; Shahin, Alain, & Picton, 2006), but also during passive listening of words (Hagoort, 2005; Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004).

Attentional Effect on Monosyllabic Word Processing

To evaluate the effect of selective auditory attention on the early steps of neocortical speech processing, participants' attention was either focused on a specific monosyllabic *word* (VA condition) or on a *noise* randomly embedded in nonword stimuli (NVA condition) appearing in the auditory stream by asking them to count their occurrences. In the VA condition, attention was thus directed toward the verbal content of each (i.e., not only the targeted) word within the auditory stream, henceforth promoting word discrimination, recognition, and the subsequent activation of their lexical memory trace. Conversely, in the NVA condition, attention was focused on each (i.e., noise and scrambled stimuli) nonverbal (i.e., devoid of any verbal content) stimuli, rendering any discrimination of the word stimuli rather passive or automatic. Of note, VA and NVA conditions only differed by the insertion of two to six noise stimuli in the NVA condition that were not present in the VA condition. Although we cannot completely rule out that this subtle difference introduced some condition effects (in addition with the attentional effects of interest), this appears rather unlikely.

Crucially, we observed a stronger steady-state response in the VA than in the NVA condition, at the left STG. This finding is in line with previous studies showing that

increased top-down selective auditory attention to speech predominantly enhances activity within STG/STS regions (Brodbeck, Hong, & Simon, 2018; Nourski, Steinschneider, Oya, Kawasaki, & Howard, 2015; Alho, Rinne, Herron, & Woods, 2014; Yoncheva et al., 2014; Sabri et al., 2008). Increased activation of specific cortical regions in VA compared with NVA condition also supports the assumption that selective top-down auditory attention can modulate neocortical responses elicited by the early steps of speech processing (Yoncheva et al., 2014; Harinen & Rinne, 2013; Sabri et al., 2008; Binder et al., 2000) and demonstrates that monosyllabic word recognition and discrimination in particular is a process sensitive to top-down attentional control. Indeed, recent studies demonstrated that lexical processing of unattended speech is possible but word identification from phonetic or syllabic information needs selective attention (Brodbeck et al., 2018; Ding et al., 2018). Interestingly, this finding also further supports the interpretation given by Vander Ghinst et al. (2016) that the increased coupling between the attended speech stream and the left STG activity in a multitalker background observed in their speech-in-noise study was driven by top-down auditory attentional processes to promote speech processing in adverse auditory scenes.

Although significantly reduced compared with that in VA condition, the presence of a significant peak of SNR at 0.54 Hz in NVA condition is worth noting. It indeed demonstrates passive or automatic verbal processing of monosyllabic word stimuli in the NVA condition (see also Brodbeck et al., 2018; Ding et al., 2018) and thereby further supports the hypothesis put forward by MMN studies that the early stages of speech processing are automatic to a certain degree (MacGregor et al., 2012; Shtyrov et al., 2005, 2010, 2012; Garagnani et al., 2009; Pulvermüller & Shtyrov, 2006; Pulvermüller et al., 2004). This result is also in line with the study of Bharadwaj et al. (2014). These authors used a frequency tagging approach to present dichotically two sequences of spoken vowels characterized by amplitude modulations at specific frequencies (from 35 to 45 Hz), asking participants to focus either on the left or on the right auditory streams. Results showed that attending to one stream increased the spectral power of the corresponding frequency-specific responses. In that study, the peak corresponding to the amplitude modulation frequency of the unattended stream was still present but substantially reduced. These findings clearly demonstrate that the difference in the amplitude of frequency-tagged neocortical responses corresponds to the effect of selective auditory attention on nonprimary auditory cortex activity.

Taken together, our results evidence that the discrimination of verbal versus nonverbal stimuli can be carried out passively and automatically but can be significantly enhanced by top-down attentional control. Admittedly (though not undermining the relevance of our findings), this pattern of results might be linked to the specific

context of our experiment. In particular, referring to the “perceptual load” theory (Lavie, 1995), a possibility is that the use of the frequency tagging approach in this study most probably rendered the occurrence of word and nonverbal stimuli highly predictable in the train of auditory stimuli, inducing a low perceptual load. This context would have allowed for automatic or passive processing of unattended stimuli (Murphy, Spence, & Dalton, 2017). Further frequency tagging studies should vary the perceptual load (e.g., by varying the frequency or the position of nonword and word stimuli) associated with the task to examine to what extent the passive neocortical processing of speech depends on task load.

Conclusion

The present MEG study demonstrates that, among the early steps (i.e., phonological, syllabic, word processing) of speech processing, monosyllabic word processing is clearly sensitive to top-down attentional control. It also evidences that top-down attention modulates activity within the left STG during monosyllabic word discrimination, which suggests that this left-hemisphere neocortical area plays a crucial role in selective top-down verbal attentional processes influencing the early steps of speech processing.

Acknowledgments

Maxime Niesen and Marc Vander Ghinst were supported by the Fonds Erasme (Brussels, Belgium). Mathieu Bourguignon and Julie Bertels were supported by the program Attract of Innoviris (grant 2015-BB2B-10). Mathieu Bourguignon was supported by the Marie Skłodowska-Curie Action of the European Commission (Grant 743562) and by the Spanish Ministry of Economy and Competitiveness (grant PSI2016-77175-P). Xavier De Tiège was Postdoctoral Master Clinical Specialist at the Fonds de la Recherche Scientifique (FRS-FNRS, Brussels, Belgium). This study and the MEG project at CUB Hôpital Erasme were financially supported by the Fonds Erasme (Research Convention: “Les Voies du Savoir,” Fonds Erasme, Brussels, Belgium).

Reprint requests should be sent to Maxime Niesen, Laboratoire de Cartographie fonctionnelle du Cerveau, UNI-ULB Neurosciences Institute, Université libre de Bruxelles, 1070 Brussels, Belgium, or via e-mail: maxime.niesen@ulb.ac.be.

REFERENCES

Alho, K., Rinne, T., Herron, T. J., & Woods, D. L. (2014). Stimulus-dependent activations and attention-related modulations in the auditory cortex: A meta-analysis of fMRI studies. *Hearing Research, 307*, 29–41.

Alho, K., Vorobyev, V. A., Medvedev, S. V., Pakhomov, S. V., Starchenko, M. G., Tervaniemi, M., et al. (2006). Selective attention to human voice enhances brain activity bilaterally in the superior temporal sulcus. *Brain Research, 1075*, 142–150.

Ashburner, J., & Friston, K. J. (1999). Nonlinear spatial normalization using basis functions. *Human Brain Mapping, 7*, 254–266.

Ashburner, J., Neelin, P., Collins, D. L., Evans, A., & Friston, K. (1997). Incorporating prior knowledge into image registration. *Neuroimage, 6*, 344–352.

Barry-Anwar, R., Hadley, H., Conte, S., Keil, A., & Scott, L. S. (2018). The developmental time course and topographic distribution of individual-level monkey face discrimination in the infant brain. *Neuropsychologia, 108*, 25–31.

Batterink, L. J., & Paller, K. A. (2019). Statistical learning of speech regularities can occur outside the focus of attention. *Cortex, 115*, 56–71.

Bharadwaj, H. M., Lee, A. K. C., & Shinn-Cunningham, B. G. (2014). Measuring auditory selective attention using frequency tagging. *Frontiers in Integrative Neuroscience, 8*, 6.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex, 10*, 512–528.

Bourguignon, M., Molinaro, N., & Wens, V. (2018). Contrasting functional imaging parametric maps: The mislocation problem and alternative solutions. *Neuroimage, 169*, 200–211.

Brodbeck, C., Hong, L. E., & Simon, J. Z. (2018). Rapid transformation from auditory to linguistic representations of continuous speech report rapid transformation from auditory to linguistic representations of continuous speech. *Current Biology, 28*, 3976–3983.

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience, 13*, 1428–1432.

Dale, A. M., & Sereno, M. I. (1993). Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *Journal of Cognitive Neuroscience, 5*, 162–176.

De Tiège, X., Op de Beeck, M., Funke, M., Legros, B., Parkkonen, L., Goldman, S., et al. (2008). Recording of epileptic activity with MEG in a light-weight magnetic shield. *Epilepsy Research, 82*, 227–231.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage, 24*, 21–33.

DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences, U.S.A., 109*, E505–E514.

Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., & Zhang, J. (2018). Attention is required for knowledge-based sequential grouping: Insights from the integration of syllables into words. *Journal of Neuroscience, 38*, 1178–1188.

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Jr., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition, 92*, 145–177.

Endrass, T., Mohr, B., & Pulvermüller, F. (2004). Enhanced mismatch negativity brain response after binaural word presentation. *European Journal of Neuroscience, 19*, 1653–1660.

Faes, L., Pinna, G. D., Porta, A., Maestri, R., & Nollo, G. (2004). Surrogate data analysis for assessing the significance of the coherence function. *IEEE Transactions on Biomedical Engineering, 51*, 1156–1166.

Farthouat, J., Atas, A., Wens, V., De Tieghe, X., & Peigneux, P. (2018). Lack of frequency-tagged magnetic responses suggests statistical regularities remain undetected during NREM sleep. *Scientific Reports, 8*, 11719.

Farthouat, J., Franco, A., Mary, A., Delpouve, J., Wens, V., Op de Beeck, M., et al. (2017). Auditory magnetoencephalographic frequency-tagged responses mirror the ongoing segmentation processes underlying statistical learning. *Brain Topography, 30*, 220–232.

- Galambos, R., Makeig, S., & Talmachoff, P. J. (1981). A 40-Hz auditory potential recorded from the human scalp. *Proceedings of the National Academy of Sciences, U.S.A.*, *78*, 2643–2647.
- Garagnani, M., Shtyrov, Y., & Pulvermüller, F. (2009). Effects of attention on what is known and what is not: MEG evidence for functionally discrete memory circuits. *Frontiers in Human Neuroscience*, *3*, 10.
- Garagnani, M., Wennekers, T., & Pulvermüller, F. (2008). A neuroanatomically grounded Hebbian-learning model of attention-language interactions in the human brain. *European Journal of Neuroscience*, *27*, 492–513.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *Neuroimage*, *86*, 446–460.
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, *9*, 416–423.
- Hämäläinen, M. S., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetocephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, *65*, 413–497.
- Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical & Biological Engineering & Computing*, *32*, 35–42.
- Hämäläinen, M. S., Lin, F.-H., & Mosher, J. C. (2010). Anatomically and functionally constrained minimum-norm estimates. In *MEG: An introduction to methods* (pp. 186–215).
- Hari, R., & Salmelin, R. (1997). Human cortical oscillations: A neuromagnetic view through the skull. *Trends in Neurosciences*, *20*, 44–49.
- Harinen, K., & Rinne, T. (2013). Activations of human auditory cortex to phonemic and nonphonemic vowels during discrimination and memory tasks. *Neuroimage*, *77*, 279–287.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech. *Nature Neuroscience*, *8*, 393–402.
- Hon, N., Ng, G., & Chan, G. (2016). Rare targets are less susceptible to attention capture once detection has begun. *Psychonomic Bulletin & Review*, *23*, 445–450.
- Hon, N., & Tan, C. H. (2013). Why rare targets are slow: Evidence that the target probability effect has an attentional locus. *Attention, Perception, & Psychophysics*, *75*, 388–393.
- Korpilähti, P., Krause, C. M., Holopainen, I., & Lang, A. H. (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain and Language*, *76*, 332–339.
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 451–468.
- Lenc, T., Keller, P. E., Varlet, M., & Nozaradan, S. (2018). Neural tracking of the musical beat is enhanced by low-frequency sounds. *Proceedings of the National Academy of Sciences, U.S.A.*, *115*, 8221–8226.
- Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, *15*, 1621–1631.
- Lins, O. G., & Picton, T. W. (1995). Auditory steady-state responses to multiple simultaneous stimuli. *Electroencephalography and Clinical Neurophysiology*, *96*, 420–432.
- Lochy, A., Van Belle, G., & Rossion, B. (2015). A robust index of lexical representation in the left occipito-temporal cortex as evidenced by EEG responses to fast periodic visual stimulation. *Neuropsychologia*, *66*, 18–31.
- MacGregor, L. J., Pulvermüller, F., van Casteren, M., & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communications*, *3*, 711.
- Makov, S., Sharon, O., Ding, N., Ben-Shachar, M., Nir, Y., & Zion Golumbic, E. (2017). Sleep disrupts high-level speech parsing despite significant basic auditory processing. *Journal of Neuroscience*, *37*, 7772–7781.
- Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*, 1006–1010.
- Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., et al. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage*, *30*, 563–569.
- Murphy, S., Spence, C., & Dalton, P. (2017). Auditory perceptual load: A review. *Hearing Research*, *352*, 40–48.
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*, 516–524.
- Nichols, T. E., & Holmes, A. P. (2001). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, *15*, 1–25.
- Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottareau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, *15*, 4.
- Nourski, K. V., Steinschneider, M., Oya, H., Kawasaki, H., & Howard, M. A., III. (2015). Modulation of response patterns in human auditory cortex during a target detection task: An intracranial electrophysiology study. *International Journal of Psychophysiology*, *95*, 191–201.
- Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *369*, 20130393.
- Nozaradan, S., Mouraux, A., & Cousineau, M. (2017). Frequency tagging to track the neural processing of contrast in fast, continuous sound sequences. *Journal of Neurophysiology*, *118*, 243–253.
- Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience*, *31*, 10234–10240.
- Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *Journal of Neuroscience*, *32*, 17572–17581.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.
- Osnes, B., Hugdahl, K., Hjelmervik, H., & Specht, K. (2011). Increased activation in superior temporal gyri as a function of increment in phonetic features. *Brain and Language*, *116*, 97–101.
- Pascual-Marqui, R. (2002). Standardized low resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods & Findings in Experimental & Clinical Pharmacology*, *24*(Suppl. D), 5–12.
- Peelle, J. E. (2012). The hemispheric lateralization of speech processing depends on what “speech” is: A hierarchical perspective. *Frontiers in Human Neuroscience*, *6*, 309.
- Perrin, F., & Grimault, N. (2019). Fonds sonores (Version v 1.0). Geneva: Zenodo. <https://doi.org/10.5281/ZENODO.3265080>.
- Peykarjou, S., Hoehl, S., Pauen, S., & Rossion, B. (2017). Rapid categorization of human and ape faces in 9-month-old infants revealed by fast periodic visual stimulation. *Scientific Reports*, *7*, 12526.
- Picton, T. W., John, M. S., Dimitrijevic, A., & Purcell, D. (2003). Human auditory steady-state responses. *International Journal of Audiology*, *42*, 177–219.
- Poeppel, D., & Hickok, G. (2015). Electromagnetic recording of the auditory system. In M. J. Aminoff, F. Boller, & D. F. Swaab, *Handbook of clinical neurology* (1st ed., Vol. 129, pp. 245–255). Amsterdam: Elsevier.

- Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., et al. (2001). Memory traces for words as revealed by the mismatch negativity. *Neuroimage*, *14*, 607–616.
- Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology*, *79*, 49–71.
- Pulvermüller, F., Shtyrov, Y., Kujala, T., & Näätänen, R. (2004). Word-specific cortical activity as revealed by the mismatch negativity. *Psychophysiology*, *41*, 106–112.
- Regan, D. (1989). *Human brain electrophysiology: Evoked potentials and evoked magnetic fields in science and medicine*. New York: Elsevier.
- Retter, T. L., & Rossion, B. (2016). Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream. *Neuropsychologia*, *91*, 9–28.
- Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage*, *61*, 1402–1418.
- Rimol, L. M., Specht, K., & Hugdahl, K. (2006). Controlling for individual differences in fMRI brain activation to tones, syllables, and words. *Neuroimage*, *30*, 554–562.
- Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *336*, 367–373.
- Ross, B., Borgmann, C., Draganova, R., Roberts, L. E., & Pantev, C. (2000). A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. *Journal of the Acoustical Society of America*, *108*, 679–691.
- Rossion, B., Torfs, K., Jacques, C., & Liu-Shuang, J. (2015). Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain. *Journal of Vision*, *15*, 18.
- Sabri, M., Binder, J. R., Desai, R., Medler, D. A., Leitl, M. D., & Liebenthal, E. (2008). Attentional and linguistic interactions in speech perception. *Neuroimage*, *39*, 1444–1456.
- Shahin, A. J., Alain, C., & Picton, T. W. (2006). Scalp topography and intracerebral sources for ERPs recorded during auditory target detection. *Brain Topography*, *19*, 89–105.
- Shtyrov, Y., Kujala, T., & Pulvermüller, F. (2010). Interactions between language and attention systems: Early automatic lexical processing? *Journal of Cognitive Neuroscience*, *22*, 1465–1478.
- Shtyrov, Y., Osswald, K., & Pulvermüller, F. (2008). Memory traces for spoken words in the brain as revealed by the hemodynamic correlate of the mismatch negativity. *Cerebral Cortex*, *18*, 29–37.
- Shtyrov, Y., Pihko, E., & Pulvermüller, F. (2005). Determinants of dominance: Is language laterality explained by physical or linguistic features of speech? *Neuroimage*, *27*, 37–47.
- Shtyrov, Y., & Pulvermüller, F. (2002). Neurophysiological evidence of memory traces for words in the human brain. *NeuroReport*, *13*, 521–525.
- Shtyrov, Y., Smith, M. L., Horner, A. J., Henson, R., Nathan, P. J., Bullmore, E. T., et al. (2012). Attention to language: Novel MEG paradigm for registering involuntary language processing in the brain. *Neuropsychologia*, *50*, 2605–2616.
- Specht, K., Osnes, B., & Hugdahl, K. (2009). Detection of differential speech-specific processes in the temporal lobe using fMRI and a dynamic “sound morphing” technique. *Human Brain Mapping*, *30*, 3436–3444.
- Steinschneider, M., Nourski, K. V., Rhone, A. E., Kawasaki, H., Oya, H., & Howard, M. A. (2014). Differential activation of human core, non-core and auditory-related cortex during speech categorization tasks as revealed by intracranial recordings. *Frontiers in Neuroscience*, *8*, 240.
- Taulu, S., Simola, J., & Kajola, M. (2005). Applications of the signal space separation method. *IEEE Transactions on Signal Processing*, *53*, 3359–3372.
- Thompson-Schill, S. L., D’Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences, U.S.A.*, *94*, 14792–14797.
- Turkeltaub, P. E., & Coslett, H. B. (2010). Localization of sublexical speech perception components. *Brain and Language*, *114*, 1–15.
- Vander Ghinst, M., Bourguignon, M., Op de Beeck, M., Wens, V., Marty, B., Hassid, S., et al. (2016). Left superior temporal gyrus is coupled to attended speech in a cocktail-party auditory scene. *Journal of Neuroscience*, *36*, 1596–1606.
- Vigário, R., Särelä, J., Jousmäki, V., Hämäläinen, M., & Oja, E. (2000). Independent component approach to the analysis of EEG and MEG recordings. *IEEE Transactions on Bio-Medical Engineering*, *47*, 589–593.
- Visser, M., Jefferies, E., & Lambon Ralph, M. A. (2010). Semantic processing in the anterior temporal lobes: A meta-analysis of the functional neuroimaging literature. *Journal of Cognitive Neuroscience*, *22*, 1083–1094.
- Wens, V., Marty, B., Mary, A., Bourguignon, M., Op de Beeck, M., Goldman, S., et al. (2015). A geometric correction scheme for spatial leakage effects in MEG/EEG seed-based functional connectivity mapping. *Human Brain Mapping*, *36*, 4604–4621.
- Woods, D. L., Herron, T. J., Cate, A. D., Kang, X., & Yund, E. W. (2011). Phonological processing in human auditory cortical fields. *Frontiers in Human Neuroscience*, *5*, 42.
- Yoncheva, Y., Maurer, U., Zevin, J. D., & McCandliss, B. D. (2014). Selective attention to phonology dynamically modulates initial encoding of auditory words within the left hemisphere. *Neuroimage*, *97*, 262–270.