Behavioral/Cognitive

Left Superior Temporal Gyrus Is Coupled to Attended Speech in a Cocktail-Party Auditory Scene

Marc Vander Ghinst,[1,2]* Mathieu Bourguignon,[1,3,4]* Marc Op de Beeck,[1] Vincent Wens,[1] Brice Marty,[1] Sergio Hassid,[2] Georges Choufani,[2] Veikko Jousmäki,[3] Riitta Hari,[3] Patrick Van Bogaert,[1] Serge Goldman,[1] and Xavier De Tiège[1]

[1]Laboratoire de Cartographie fonctionnelle du Cerveau, UNI-ULB Neuroscience Institute, Université libre de Bruxelles, 1070 Brussels, Belgium, [2]Service d'ORL et de chirurgie cervico-faciale, ULB-Hôpital Erasme, Université libre de Bruxelles, 1070 Brussels, Belgium, [3]Brain Research Unit, Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, FI-00076-AALTO, Espoo, Finland, and [4]BCBL, Basque Center on Cognition, Brain and Language, 20009 San Sebastian, Spain

Using a continuous listening task, we evaluated the coupling between the listener's cortical activity and the temporal envelopes of different sounds in a multitalker auditory scene using magnetoencephalography and corticovocal coherence analysis. Neuromagnetic signals were recorded from 20 right-handed healthy adult humans who listened to five different recorded stories (attended speech streams), one without any multitalker background (No noise) and four mixed with a "cocktail party" multitalker background noise at four signal-to-noise ratios (5, 0, −5, and −10 dB) to produce speech-in-noise mixtures, here referred to as Global scene. Coherence analysis revealed that the modulations of the attended speech stream, presented without multitalker background, were coupled at ~0.5 Hz to the activity of both superior temporal gyri, whereas the modulations at 4–8 Hz were coupled to the activity of the right supratemporal auditory cortex. In cocktail party conditions, with the multitalker background noise, the coupling was at both frequencies stronger for the attended speech stream than for the unattended Multitalker background. The coupling strengths decreased as the Multitalker background increased. During the cocktail party conditions, the ~0.5 Hz coupling became left-hemisphere dominant, compared with bilateral coupling without the multitalker background, whereas the 4–8 Hz coupling remained right-hemisphere lateralized in both conditions. The brain activity was not coupled to the multitalker background or to its individual talkers. The results highlight the key role of listener's left superior temporal gyri in extracting the slow ~0.5 Hz modulations, likely reflecting the attended speech stream within a multitalker auditory scene.

Key words: coherence analysis; magnetoencephalography; speech in noise

Significance Statement

When people listen to one person in a "cocktail party," their auditory cortex mainly follows the attended speech stream rather than the entire auditory scene. However, how the brain extracts the attended speech stream from the whole auditory scene and how increasing background noise corrupts this process is still debated. In this magnetoencephalography study, subjects had to attend a speech stream with or without multitalker background noise. Results argue for frequency-dependent cortical tracking mechanisms for the attended speech stream. The left superior temporal gyrus tracked the ~0.5 Hz modulations of the attended speech stream only when the speech was embedded in multitalker background, whereas the right supratemporal auditory cortex tracked 4–8 Hz modulations during both noiseless and cocktail-party conditions.

Introduction

To follow and understand a single speaker among other competing voices ("cocktail-party effect") (Cherry, 1953), the human brain has to handle multiple acoustic cues and engage various sensory and attentional processes to tune in to a single speaker's voice while tuning out the noisy environs (McDermott, 2009).

Received May 1, 2015; revised Nov. 26, 2015; accepted Dec. 16, 2015.

Author contributions: M.V.G., M.B., S.H., G.C., R.H., P.V.B., S.G., and X.D.T. designed research; M.V.G., M.B., V.W., and B.M. performed research; M.O.d.B. and V.J. contributed unpublished reagents/analytic tools; M.V.G., M.B., M.O.d.B., V.W., and X.D.T. analyzed data; M.V.G., M.B., V.W., V.J., R.H., S.G., and X.D.T. wrote the paper.

This work was supported by Fonds de la Recherche Scientifique Research Credit J.0085.13, FRS-FNRS (Brussels, Belgium) and Fonds Erasme (Brussels, Belgium). M.V.G. was supported by Fonds Erasme (Brussels, Belgium). V.W. and X.D.T. were supported by Fonds de la Recherche Scientifique (FRS-FNRS, Brussels, Belgium). V.J. was supported by Institut d'Encouragement de la Recherche Scientifique et de l'Innovation de Bruxelles (Brains back to Brussels, Brussels, Belgium). We thank Nicolas Grimault and Fabien Perrin (Centre de Recherche en Neuroscience, Lyon, France) for providing us access to their entire audio material.

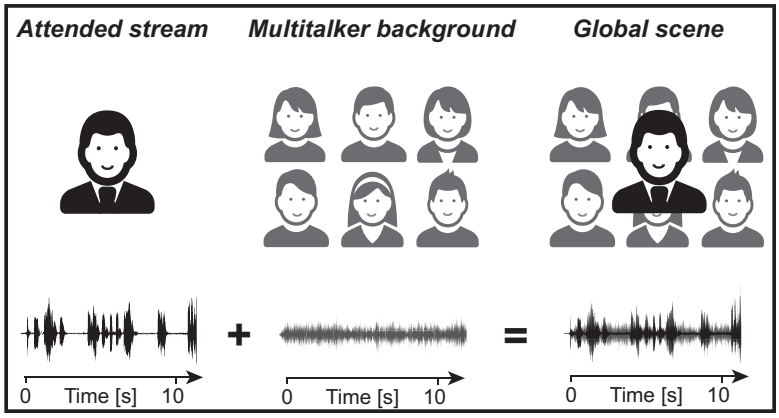

Figure 1. Experimental setup and the corresponding sounds (bottom traces). The Global scene is the combination of the Attended stream (black traces), the voice of the reader of a story, and of the Multitalker background (gray traces) obtained by mixing voices from six simultaneous French-speaking talkers (3 females and 3 males).

A major issue in comprehending neural speech-in-noise processing is to determine how the brain extracts the acoustic attributes of a specific speaker's voice from the whole auditory scene and how increasing background noise corrupts these neural processes (Zion Golumbic et al., 2013; Fishman et al., 2014; Simon, 2014).

Complex acoustic signals, such as speech sounds, can be decomposed into a temporal fine structure (TFS, referring to rapid phase fluctuations) and a temporal envelope (TE, amplitude modulations at frequencies <50 Hz) (Rosen, 1992). The speech TE carries critical information for speech comprehension (Shannon et al., 1995). Especially, the slow (<16 Hz) TE fluctuations, corresponding to the syllabic and phrasal rhythms of speech, play a key role in speech understanding in both quiet (Rosen, 1992; Greenberg et al., 2003) and noisy environments (Drullman et al., 1994a; Fullgrabe et al., 2009).

The rhythmic activity of the auditory cortex is typically coupled to the <10 Hz TE modulations of speech (Suppes et al., 1997, 1998, 1999; Ahissar et al., 2001; Luo and Poeppel, 2007; Wang et al., 2012; Bourguignon et al., 2013; Peelle et al., 2013). Moreover, in noisy environments (two competing speakers or spectrally matched stationary noise), the activity of the listener's auditory cortex follows the slow (1–8 Hz) temporal modulations of the attended speaker's voice regardless of the background noise level (Ding and Simon, 2012b, 2013; Zion Golumbic et al., 2013), suggesting a relative noise insensitivity of neural synchronization with slow TE modulations of speech (Ding and Simon, 2013). Acoustic components of the auditory scene are reflected in rhythmic cortical activity, but it is still unclear how increasing noise levels affect those brain rhythms in a typical cocktail-party auditory scene where the multitalker background noise varies in intensity (Zion Golumbic et al., 2013; Fishman et al., 2014; Simon, 2014).

In this magnetoencephalographic (MEG) study, we investigated, using an ecologically valid continuous listening task and corticovocal coherence analysis (Bourguignon et al., 2013), the frequency-specific coupling between the listener's cortical activity and the time course of different acoustic components of a multitalker auditory scene (Fig. 1). The corresponding auditory stimuli consisted of (1) the attended speech stream, that is, the reader's voice (Attended stream), (2) the unattended Multitalker background noise, and the voices of each individual background talkers, and (3) the whole acoustic scene, henceforth referred to as the Global scene (i.e., the combination of the Attended stream and the unattended Multitalker background). This study was specifically designed (1) to determine at which frequency and in which cortical regions, in ecological speech-in-noise conditions, the brain specifically tracks the TE of the Attended stream, the Multitalker background (and its components), and the Global scene, and (2) to assess the effect of the signal-to-noise ratio (SNR) on this frequency-specific coupling.

Materials and Methods

The methods used for MEG data acquisition, preprocessing, and analyses are derived from Bourguignon et al. (2013) and will be explained here only briefly.

Subjects

Twenty native French-speaking healthy subjects (mean age 30 years, range 23–40 years, 10 females and 10 males) without any history of neuropsychiatric or otologic disorder participated in this study. All subjects had normal hearing according to pure tone audiometry (i.e., normal hearing thresholds, between 0 and 20 dB HL for 250, 500, 1000, 2000, 4000, and 8000 Hz) and normal otomicroscopy. They were all right-handed (mean 83, range 65–100; left-right scale from −100 to 100) according to the Edinburgh Handedness Inventory (Oldfield, 1971). The study had prior approval by the ULB-Hôpital Erasme Ethics Committee. Subjects gave written informed consent before participation.

Experimental paradigm

During MEG recordings, the subjects comfortably sat in the MEG chair with the arms resting on a table positioned in front of them. They underwent five listening conditions and one Rest condition, each lasting 5 min. The order of the six conditions was randomized for each subject.

During the listening conditions, subjects were asked to attend to five different recorded stories, one in each of the five condition, narrated in French by different native French-speaking readers. The recordings were randomly selected from a set of six texts (readers' sex ratio 3/3) obtained from a French audiobook database (http://www.litteratureaudio.com) after written authorization from the readers.

The No noise condition was presented without Multitalker background. A specific SNR (i.e., Attended stream vs Multitalker background) was randomly assigned to each text: 5, 0, −5, and −10 dB, leading to four additional speech-in-noise SNR conditions (Fig. 2, left). The Multitalker background (Fonds sonores version 1.0) (Perrin and Grimault, 2005) served as a continuous cocktail party noise obtained by mixing the voices of six French speakers talking simultaneously in French (3 females and 3 males). This configuration of cocktail-party noise was selected because it accounts for both energetic and informational masking at phonetic and lexical level (Simpson and Cooke, 2005; Hoen et al., 2007).

The audio recordings were played using VLC media player (VideoLAN Project, GNU General Public License) running on a MacBook Pro (Apple Computer) and transmitted to a MEG-compatible 60 × 60 cm^2 high-quality flat-panel loudspeaker (Panphonics SSH sound shower, Panphonics) placed 3 m in front of the subjects. The average sound intensity was 60 dB SPL as assessed by a sound level meter (Sphynx Audio System). Subjects were asked to attend to the reader's voice and to gaze at a fixation

The authors declare no competing financial interests.

*M.V.G. and M.B. contributed equally to this work.

Correspondence should be addressed to Dr. Marc Vander Ghinst, Laboratoire de Cartographie fonctionnelle du Cerveau, UNI-ULB Neuroscience Institute, Université libre de Bruxelles, 808 Lennik Street, 1070 Brussels, Belgium. E-mail: Marc.Vander.Ghinst@erasme.ulb.ac.be.

DOI:10.1523/JNEUROSCI.1730-15.2016

 0270-6474/16/361597-11$15.00/0

point on the wall of the magnetically shielded room facing them. During the Rest condition, subjects were instructed to relax, not to move, and to gaze at the same fixation point.

At the end of each listening condition, subjects were asked to quantify the intelligibility of the attended reader's voice using a visual analog scale (VAS) ranging from 0 to 10 (0, totally unintelligible; 10, perfectly intelligible).

Data acquisition

Cortical neuromagnetic signals were recorded at ULB-Hôpital Erasme using a whole-scalp-covering MEG device installed in a lightweight magnetically shielded room (Vectorview and Maxshield, Elekta), the characteristics of which being described elsewhere (De Tiège et al., 2008; Carrette et al., 2011). The MEG device has 102 sensor chipsets, each comprising one magnetometer and two orthogonal planar gradiometers. MEG signals were bandpass-filtered through 0.1–330 Hz and sampled at 1 kHz. Four head-tracking coils monitored subjects' head position inside the MEG helmet. The locations of the coils and at least 150 head-surface (on scalp, nose, and face) points with respect to anatomical fiducials were digitized with an electromagnetic tracker (Fastrack, Polhemus). Electro-oculogram (EOG), electrocardiogram (ECG), and audio signals presented to the subjects were recorded simultaneously with MEG signals (bandpass 0.1–330 Hz for EOG and ECG, and low-pass at 330 Hz for audio signals; all signals sampled at 1 kHz). The recorded audio signals were used for synchronization between MEG and the transmitted audio signals, the latter being bandpassed at 50–22,000 Hz and sampled at 44.1 kHz. High-resolution 3D-T1 magnetic resonance images (MRIs) of the brain were acquired on a 1.5 T MRI (Intera, Philips).

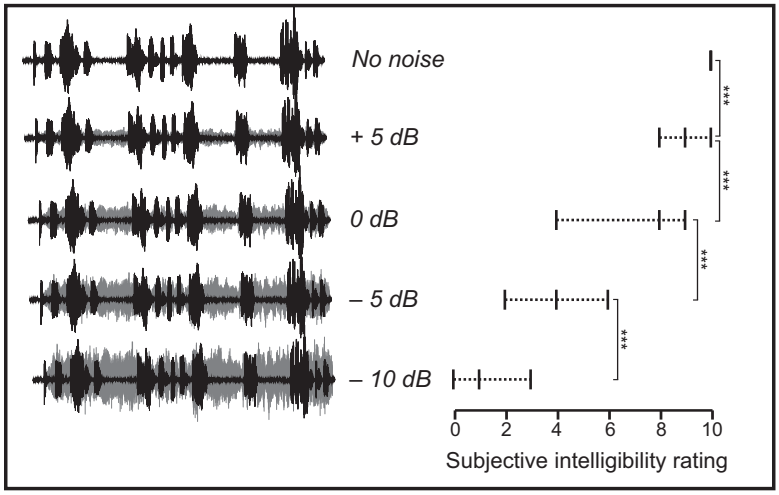

Figure 2. Global scene sound signals at different experimental SNRs and their corresponding relationship between subjective intelligibility scores: mean ± range, VAS ranging from 0 (totally unintelligible) to 10 (perfectly intelligible). Black represents the Attended stream. Gray represents the Multitalker background. The intelligibility of the Attended stream decreased significantly with decreasing SNRs. Vertical brackets represent the *post hoc* paired *t* tests between adjacent conditions. *** $p < 0.001$.

Data preprocessing

Continuous MEG data were first preprocessed off-line using the signal-space-separation method (Taulu et al., 2005) to suppress external inferences and to correct for head movements. For coherence analyses, continuous MEG and audio signals were split into 2048 ms epochs with 1638 ms epoch overlap, leading to a frequency resolution of ~0.5 Hz (Bortel and Sovka, 2007). MEG epochs exceeding 3 pT (in magnetometers) or 0.7 pT/cm (in gradiometers) were excluded from further analysis to avoid contamination of the data by eye movements and blinks, muscle activity, or MEG-sensor artifacts. These steps led to an average of 685 (range 516–727) artifact-free epochs across subjects and conditions.

Coherence analyses in sensor space

For each listening condition, synchronization between rectified wideband (50–22,000 Hz) audio signals and artifact-free MEG epochs was assessed with coherence analysis in sensor space. The analysis covered frequencies from 0.1 to 20 Hz, as speech-amplitude modulations at these frequencies are critical for speech comprehension (Drullman et al., 1994b; Shannon et al., 1995; Greenberg et al., 2003; Fullgrabe et al., 2009). The coherence, an extension of Pearson correlation coefficient to the frequency domain, quantifies the degree of coupling between two signals, providing a number between 0 (no linear dependency) and 1 (perfect linear dependency) for each frequency (Halliday et al., 1995).

For the four speech-in-noise conditions (5, 0, −5, and −10 dB), coherence was separately computed between MEG signals and three acoustic components of the auditory scene: (1) the Global scene (i.e., Attended stream + Multitalker background), leading to Coh_{global}; (2) the Attended stream only (i.e., the reader's voice), leading to Coh_{att}; and (3) the Multitalker background only, leading to Coh_{bckgr}.

Additional coherence analyses were performed (as described here and below) to investigate a possible neural coupling with the individual components of the Multitalker background by computing the coherence between MEG signals and the individual audio recordings of the 6 different speakers composing the Multitalker background, leading to $Coh_{talker1\text{-}6}$.

Sensor-level coherence maps were obtained using gradiometer signals only, and signals from gradiometer pairs were combined as previously described by Bourguignon et al. (2015).

Previous studies demonstrated significant coupling between acoustic and brain signals in δ band (~0.5 Hz) (Bourguignon et al., 2013; Clumeck et al., 2014) and θ band (4–8 Hz) (Ding and Simon, 2012a; Peelle et al., 2013; Koskinen and Seppä, 2014). Accordingly, sensor-level coherence maps were produced separately in δ band (coherence at 0.5 Hz) and θ band (mean coherence across 4–8 Hz). These two frequency bands are henceforth referred to as frequency bands of interest.

Coherence analyses in source space

Individual MRIs were first segmented using Freesurfer software (Martinos Center for Biomedical Imaging). MEG and segmented MRI coordinate systems were then coregistered using the three anatomical fiducial points for initial estimation and the head-surface points to manually refine the surface coregistration. The MEG forward model, comprising pairs of two orthogonal tangential current dipoles, placed on a homogenous 5 mm grid source space covering the whole brain, was subsequently computed using MNE suite (Martinos Centre for Biomedical Imaging). As a preliminary step, to simultaneously combine data from the planar gradiometer and the magnetometer sensors for source estimation, sensor signals (and the corresponding forward-model coefficients) were normalized by their noise root mean square (rms), estimated from the Rest data filtered through 1–195 Hz. Coherence maps obtained for each subject, listening conditions (No noise, 5, 0, −5, and −10 dB), audio signal content (Global scene, Attended stream, Multitalker background), and frequency band of interest (δ and θ) were finally produced using the Dynamic Imaging of Coherent Sources approach (Gross et al., 2001) with Minimum-Norm Estimates inverse solution (Dale and Sereno, 1993). Noise covariance was estimated from the Rest data, and the regularization parameter was fixed in terms of MEG SNR (Hämäläinen and Mosher, 2010).

Group-level analyses in source space

A nonlinear transformation from individual MRIs to the standard MNI brain was first computed using the spatial normalization algorithm implemented in Statistical Parametric Mapping (SPM8, Wellcome Department of Cognitive Neurology, London) (Ashburner et al., 1997; Ashburner and Friston, 1999) and then applied to individual MRIs and every coherence map. This procedure generated normalized coherence maps in the MNI space for each subject, listening condition, audio signal, and frequency band of interest.

To produce coherence maps at the group level, we computed the across subjects generalized *f*-mean of normalized maps, according to $f(\cdot) = \arctanh(\sqrt{\cdot})$, namely, the Fisher *z*-transform of the square-root. This procedure transforms the noise on the coherence estimate to be approximately normally distributed (Rosenberg et al., 1989). Thus, the computed coherence is an unbiased estimate of the mean coherence at the group level. In addition, this averaging procedure avoids an over-contribution of subjects characterized by high coherence values (Bourguignon et al., 2012).

Local coherence maxima were subsequently identified in group-level coherence maps obtained for each frequency band of interest and listening condition (No noise, 5, 0, −5, and −10 dB). Local coherence maxima are sets of contiguous voxels displaying higher coherence value than all other neighboring voxels. Because the apparent source extent depends on the signal-to-noise ratio of the data, we only report local (and statistically significant) coherence maxima without paying attention to coherence that seems to cover wider areas.

Statistical analyses

Number of artifact-free epochs. We tested for potential differences between listening conditions in the number of artifact-free epochs used to estimate the coherence using one-way repeated-measures ANOVA, with the 5 noise levels (No noise, 5, 0, −5, −10 dB) and with the number of artifact-free epochs as dependent variable.

Coherence in sensor space. For each listening condition, audio signal, and frequency band of interest, the statistical significance of individual coherence levels was assessed in sensor space with surrogate data-based statistics, which intrinsically deals with the multiple-comparison issue and takes into account the temporal autocorrelation within signals. For each subject, 1000 surrogate sensor-level coherence maps were computed, as was done for genuine sensor coherence maps but with the audio signals replaced by Fourier-transform-surrogate audio signals; the Fourier transform surrogate imposes power spectrum to remain the same as in the original signal, but it replaces the phase of Fourier coefficients by random numbers in the range $(-\pi; \pi)$ (Faes et al., 2004). Then, the maximum coherence value across all sensors was extracted for each surrogate simulation. Finally, the 0.95 percentile of this maximum coherence value yielded the coherence threshold at $p < 0.05$.

We also tested whether the number of subjects displaying significant coherence with at least one component of the Multitalker background was greater than expected by chance. Given our significance threshold for false positive coherence of 0.05 per subject and background component, the probability of observing spurious coherence with at least one of the six components per subject is 0.265 (= sum of probabilities of observing spurious coherence with 1, 2, 3, 4, 5, or 6 components in one subject). Consequently, the number of subjects displaying spurious Coh_{talker} with at least one background component follows a binomial distribution B(20, 0.265), and should be 8 or less at $p < 0.05$.

Coherence in source space. The statistical significance of the local coherence maxima observed in group-level source coherence maps was assessed with a nonparametric permutation test (Nichols and Holmes, 2002). The following procedure was performed for each listening condition and audio signal separately. First, subject- and group-level Rest coherence maps were computed, as was done for the different listening conditions coherence maps, but with MEG signals replaced by Rest MEG signals and sound signals unchanged. Group-level difference maps were obtained by subtracting *f*-transformed listening conditions coherence maps and Rest group-level coherence maps for each frequency band of interest. Under the null hypothesis that coherence maps are the same whatever the experimental condition, the labeling listening conditions and Rest are exchangeable at the subject-level before group-level difference map computation (Nichols and Holmes, 2002). To reject this hypothesis and to compute a threshold of statistical significance for the correctly labeled difference map, the permutation distribution of the maximum of the difference map's absolute value was computed from a subset of 10,000 permutations. The threshold at $p < 0.05$ was computed as the 95th percentile of the permutation distribution (Nichols and Holmes, 2002). All suprathreshold local coherence maxima were interpreted as indicative of brain regions showing statistically significant coupling with the audio signals.

Cortical processing of the auditory scene in speech-in-noise conditions

To identify cortical areas wherein activity would reflect more either the Attended stream or the Multitalker background than the processing of the Global scene, we compared Coh_{att} with Coh_{global} and Coh_{bckgr} with Coh_{global} coherence maps using the same nonparametric permutation test described above (permutation performed over the labels Global scene and Attended stream, or Global scene and Multitalker background instead of audio and Rest, leading to the $Coh_{att} - Coh_{global}$ and $Coh_{bckgr} - Coh_{global}$ difference maps).

The same method was used to search for cortical areas wherein activity would reflect more the processing of the Global scene than the processing of the Attended stream or the Multitalker background. This led to $Coh_{global} - Coh_{att}$ and $Coh_{global} - Coh_{bckgr}$ difference maps.

Effect of the SNR on corticovocal coherence and hemispheric lateralization in speech-in-noise conditions

The effect of the SNR on source-space coherence values and hemispheric lateralization was assessed with two-way, 5 noise levels (No noise, 5, 0, −5, −10 dB) × 2 hemispheric lateralizations (left, right), repeated-measures ANOVA. The dependent variable was the maximum coherence value within a sphere of 10 mm radius around the group-level difference-map maximum. To exclude the possibility that any significant effect could be related to variations in the coordinates of the local maxima, or to the inverse solution used to estimate source-space coherence, we performed a similar ANOVA with sensor-level coherence values. In that analysis, the dependent variable was the maximum coherence across preselected left/right hemisphere sensors (48 pairs of gradiometers for each side) widely covering the auditory cortices.

Effect of SNR on the intelligibility of attended stream, and the relationship between intelligibility and coherence levels

The effect of the SNR on the VAS scores was assessed with a one-way repeated-measures ANOVA, with the 5 noise levels (No noise, 5, 0, −5, −10 dB) with the VAS score as dependent variable.

To evaluate the correlation between intelligibility and coherence levels, we performed Pearson correlation analysis between coherence levels (separately in δ and θ bands) and VAS scores. This analysis was performed across and within the different SNR conditions (No noise, 5, 0, −5, −10 dB).

Results

Differences in the number of artifact-free epochs between listening conditions

The comparison of the number of artifact-free epochs used to estimate the coherence spectra revealed no significant differences between the five listening conditions ($p = 0.23$, $F_{(4,76)} = 1.43$).

Effect of SNR on the intelligibility of attended speech stream

Figure 2 (right) illustrates the progressive decrease in the intelligibility of the attended speech stream (i.e., attended reader's voice) as a function of SNR, as demonstrated by the ANOVA ($F_{(4,76)} = 206$; $p < 0.0001$); the decrease was on average 88% from the No noise condition to −10 dB condition. *Post hoc* comparisons performed with pairwise *t* tests confirmed that this decrease was statistically significant between all adjacent SNR conditions ($p < 0.001$).

Corticovocal coherence in the absence of Multitalker background

Figure 3 illustrates the results obtained in sensor and source spaces for the No noise condition. In the sensor space (left panel), statistically significant δ band (~0.5 Hz) coherence between the reader's voice and the listeners' MEG signals peaked at sensors

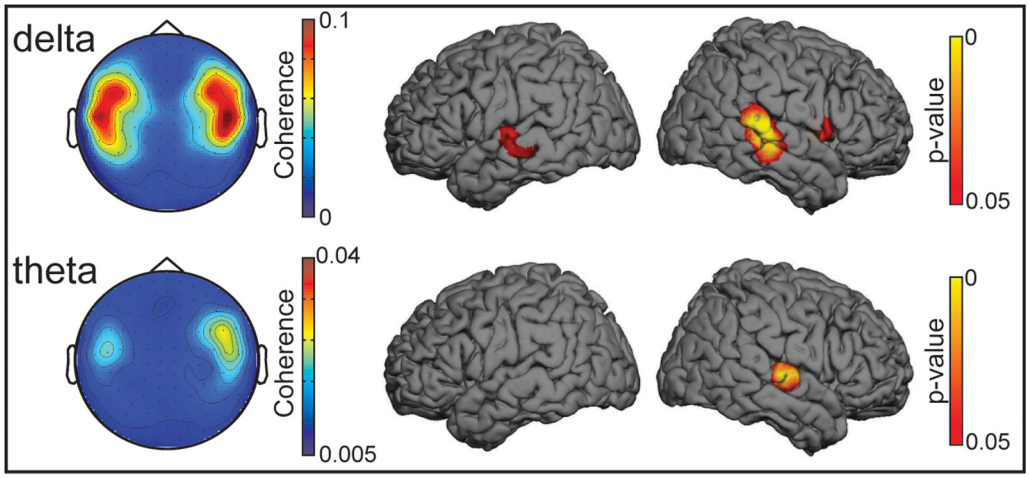

Figure 3. Sensor and source space results obtained in the No noise condition. Left, Spatial distribution of group-level sensor space coherence in the δ band (0.5 Hz; top) and θ band (4–8 Hz; bottom). In both frequency bands, the coherence maxima are located bilaterally at gradiometer sensors covering the temporal areas. The sensor array is viewed from the top. Right, Results obtained in the source space. Group-level statistical (*p* value) maps showing brain areas displaying statistical significant coherence. Maps are thresholded at $p < 0.05$. In the δ band, significant local maxima occur at the lower bank of the superior temporal gyrus bilaterally, with no significant hemispheric lateralization. In the θ band, a significant local maximum is seen only at right supratemporal auditory cortex.

Table 1. Coherence in sensor space

	Attended stream		Multitalker background		Global scene	
Condition	Coherence	*N*	Coherence	*N*	Coherence	*N*
δ band						
No noise	0.16 (0.04–0.42)	20				
5 dB	0.09 (0.05–0.17)	20	0.02 (0.01–0.03)	0	0.08 (0.04–0.15)	20
0 dB	0.11 (0.03–0.3)	18	0.02 (0.01–0.05)	1	0.08 (0.02–0.25)	18
−5 dB	0.07 (0.02–0.16)	15	0.03 (0.01–0.12)	1	0.04 (0.01–0.09)	6
−10 dB	0.04 (0.02–0.09)	5	0.02 (0.01–0.04)	0	0.03 (0.01–0.05)	2
θ band						
No noise	0.03 (0.01–0.12)	18				
5 dB	0.05 (0.02–0.1)	20	0.01 (0.01–0.03)	2	0.04 (0.02–0.09)	20
0 dB	0.04 (0.02–0.07)	20	0.02 (0.01–0.03)	8	0.03 (0.02–0.06)	19
−5 dB	0.02 (0.01–0.07)	13	0.02 (0.01–0.04)	3	0.02 (0.01–0.06)	10
−10 dB	0.02 (0.01–0.02)	3	0.02 (0.01–0.03)	1	0.02 (0.01–0.03)	3

Mean (and range) coherence values and the number of subjects (*N*) showing statistically significant sensor space coherence for each audio signal at various signal-to-noise ratios within δ (~0.5 Hz) and θ (4–8 Hz) bands.

covering temporal areas bilaterally; the coherence was statistically significant in all individuals (all $p < 0.05$; Table 1). Statistically significant θ band (4–8 Hz) coherence was observed in 18 of the 20 subjects (Table 1).

To identify the cortical generators of the sensor level coherence, source reconstruction was performed separately for δ and θ bands. The δ band coherence occurred in the superior temporal gyrus (STG) bilaterally (see Table 3; $p < 0.0001$), without hemispheric difference (paired *t* test, $p = 0.549$), and the maximum θ band coherence occurred in the right supratemporal auditory cortex (AC) (see Table 3; $p = 0.0235$).

Corticovocal coherence in speech-in-noise conditions

Table 1 provides the coherence values and number of subjects showing significant sensor-space Coh_{global}, Coh_{att}, and Coh_{bckgr} in δ and θ bands.

Figure 4 illustrates the results obtained in the sensor and the source spaces in all conditions.

Significant sensor-space δ band coherence with the Attended stream (Coh_{att}) was observed at sensors covering the temporal areas in the majority of subjects down to −10 dB where five subjects still showed significant coherence. Similar results were observed with the Global scene (Coh_{global}), except that the number of subjects displaying significant coherence substantially decreased from −5 dB onward (6 subjects at −5 dB; 2 subject at −10 dB). By contrast, ≤1 subjects per SNR condition exhibited significant δ band coherence with the Multitalker background (Coh_{bckgr}).

In the θ band, Coh_{att} and Coh_{global} were significant in most subjects down to −10 dB where 3 subjects still showed significant coherence. On the other hand, ≤3 subjects per SNR condition exhibited significant Coh_{bckgr} with the Multitalker background (except at 0 dB, where 8 subjects exhibited significant Coh_{bckgr}).

For both frequencies of interest, some subjects had statistically significant coherence with the voices of the individual talkers composing the Multitalker background. These results are detailed in Table 2. However, whatever the condition or the frequency bands of interest, the number of subjects displaying statistically significant coherence with at least one of the 6 talkers composing the Multitalker background was ≤8, and hence, compatible with what is expected by chance.

Based on the low occurrence of sensor-space Coh_{bckgr} and $Coh_{talker1\text{-}6}$ for both frequency bands of interest, Coh_{bckgr} or $Coh_{talker1\text{-}6}$ were not considered for further source-space analyses.

Table 3 provides the MNI coordinates of group-level global maxima and the corresponding coherence values in the source space for the two frequency bands of interest.

In the δ band, Coh_{att} and Coh_{global} maps displayed statistically significant ($p < 0.05$) local maxima at the STG bilaterally at 5 and 0 dB, but only at the left STG for the −5 dB condition. At −10 dB, a local coherence maximum was observed in the right STG, but it did not reach statistical significant ($p = 0.14$ for Coh_{global}; $p = 0.065$ for Coh_{att}).

In the θ band, Coh_{att} and Coh_{global} maps displayed statistically significant ($p < 0.05$) local maxima in the right AC in every condition. Furthermore, Coh_{att} maps had at −10 dB an additional significant local maximum in the left STG.

Cortical processing of the auditory scene in speech-in-noise conditions

Figure 5 (two top panels) illustrates cortical areas where Coh_{att} was statistically significantly stronger than Coh_{global}, meaning that

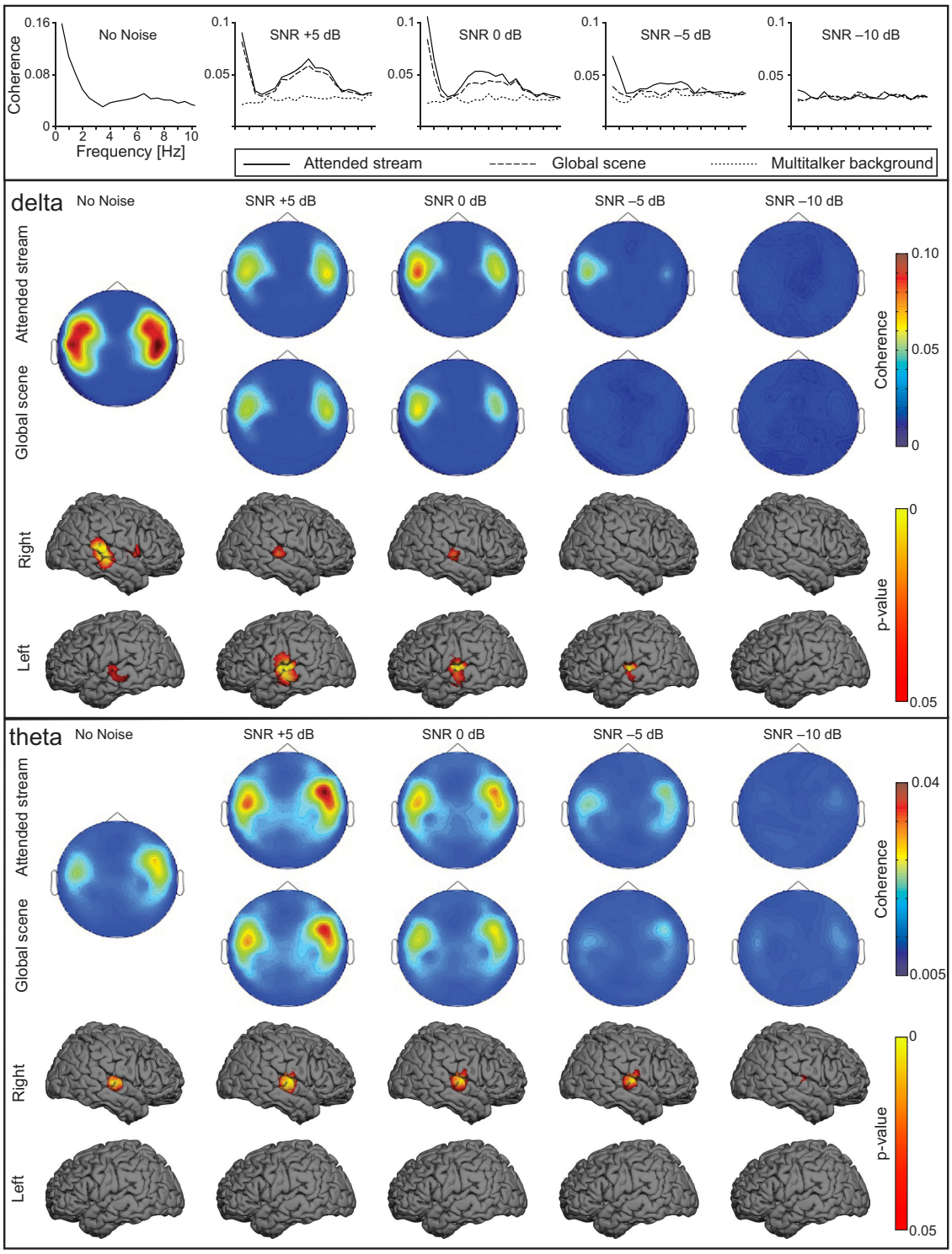

Figure 4. Sensor and source-space results obtained for the listening conditions. Top, Mean coherence spectra representing the arithmetic mean of the 20 individual maximum coherence spectra when coherence was computed between MEG signals and the envelopes of the different components of the auditory scene. Middle, Top, Group-level gradiometer sensor space coherence in the δ band ($\sim$0.5 Hz). Higher coherence values were found at the temporal-lobe sensors (with a left hemisphere dominance) when coherence was computed between MEG signals and the Attended stream (Coh_{att}) than with the Global scene (Coh_{global}). Coherence decreased as the level of the Multitalker background progressively increased. The sensor array is viewed from the top. Middle, Bottom, Group-level source space coherence in the δ band when the coherence was computed with the Attended stream (Coh_{att}). Group-level p value coherence map disclosed local coherence maxima at the superior temporal gyrus bilaterally with left hemisphere dominance in the noisy conditions. Bottom, Top, Group-level gradiometer mean sensor space coherence in θ band (4–8 Hz). Higher coherence values were found at the temporal-lobe sensors (with a right dominance) when coherence was computed between MEG signals and the Attended stream (Coh_{att}) than with the Global scene (Coh_{global}). Coherence values decreased with increasing noise level. Bottom, Bottom, Group-level source space coherence in θ band when the coherence was computed with the Attended stream (Coh_{att}). Group-level p value coherence map disclosed significant coherence maxima at the right supratemporal auditory cortex in every listening condition.

Table 2. Coherence with Multitalker background components

Condition	Subjects with significant coherence	Mean Coh_{att} − Coh_{Talker} value (± SD)	T_1	T_2	T_3	T_4	T_5	T_6
δ band								
5 dB	3	0.063 (±0.064)	1	0	1	0	0	1
0 dB	2	0.081 (±0.019)	1	0	0	1	0	0
−5 dB	2	0.04 (±0.002)	0	0	0	1	1	1
−10 dB	7	−0.0078 (±0.014)	1	0	0	3	1	2
θ band								
5 dB	8	0.031 (±0.028)	2	0	1	3	1	3
0 dB	6	0.03 (±0.018)	2	0	1	3	0	1
−5 dB	7	0.006 (±0.009)	4	0	1	3	1	0
−10 dB	5	−0.004 (±0.003)	1	0	0	4	0	2

For each condition within δ (~0.5 Hz) and θ (4–8 Hz) bands: left, the number of subjects having significant coherence with at least one of the component of the Multitalker background; middle, mean ± SD of difference between the coherence with the attended stream (Coh_{att}) and the maximal coherence with the single talkers composing the Multitalker background; and right, the number of subjects having significant coherence with each of the talker (T) composing the Multitalker background.

the MEG signals followed significantly more the Attended stream than the Global scene. In the δ band (top), these areas included the left STG in every condition ($p < 0.05$, except in the unintelligible −10 dB condition) and the right STG in the 5, 0, and −10 dB conditions. In the θ band (middle), the right AC fulfilled this criterion ($p < 0.05$ in all other conditions, except the unintelligible −10 dB condition). Left STG also displayed significantly stronger Coh_{att} than Coh_{global} in the 0 and −5 dB condition, even if coherence at left auditory cortices was not significant in Coh_{att} and Coh_{global} both at sensor and source levels.

Figure 5 (bottom) illustrates the Attended stream TE, the Global scene TE, and the time courses of MEG signals at a left temporal-lobe sensor for 0 dB SNR. From these traces, it is evident that cortical activity in the left temporal lobe followed better the Attended stream than the Global scene TE; this effect was seen in all speech-in-noise conditions (except at −10 dB).

The Coh_{global}-Coh_{att} contrast maps did not reveal any cortical area with significantly higher coherence level with the Global scene than with the Attended stream ($p > 0.9$ in all conditions).

Effect of SNR in speech-in-noise conditions

As the coherence was at maximum when computed with the Attended stream (Coh_{att}) compared with the Global scene (Coh_{global}), the repeated-measures ANOVA assessing the effect of the SNR and hemispheric lateralization on coherence level was performed only for Coh_{att}.

In the δ band, the analysis revealed a main effect associated with the SNR ($F_{(4,76)} = 30.4$, $p < 0.0001$) as well as an interaction between the SNR and hemispheric lateralization ($F_{(4,76)} = 2.86$, $p = 0.03$). Coh_{att} values indeed decreased with increasing level of Multitalker background (corresponding to decreasing SNR) at both left and right STG (Table 3; Fig. 4). The interaction between SNR and hemispheric lateralization likely reflects the more drastic decrease of coherence level with decreasing SNR in the right than the left STG (Table 3; Fig. 4). Of notice, the interaction was more significant for sensor-level than source-space coherence ($F_{(4,76)} = 4.64$, $p = 0.002$ at sensor level).

In the θ band, coherence was also affected by the Multitalker background ($F_{(4,76)} = 25.49$, $p < 0.0001$), decreasing as the Multitalker background increased. Furthermore and regardless of the SNR condition, we observed right hemisphere-lateralized coherence that further corroborated the hemispheric dominance ($F_{(4,76)} = 9.8$, $p = 0.0054$).

Thus, the hemispheric dominance of the coherence computed in speech-in-noise conditions was different for δ (i.e., left-lateralized) and θ (i.e., right-lateralized) bands.

Correlation between intelligibility of the attended stream and the coherence levels

The intelligibility of the Attended stream and the coherence levels were statistically significantly correlated across the different SNR conditions and in the two frequency bands of interest (Pearson's correlation, $p < 0.0001$). However, this relationship did not appear within each SNR condition (all corrected $p > 0.05$), likely because of the relative homogeneity of VAS scores across subjects within the same SNR condition.

Discussion

We used an ecological continuous speech-in-noise listening task, where the attended speech stream was embedded within a Multitalker background of different intensity levels. We found (1) left hemisphere-dominant coupling between the ~0.5 Hz modulation of the Attended stream and the STG activity, (2) a preserved right hemisphere dominant coupling at 4–8 Hz within the supratemporal AC, but (3) no specific cortical coupling with the Multitalker background (nor its individual talkers). Furthermore, no coupling was found between the Global scene (comprising all sounds) and cortical activity. We also found that the coupling between the slow (~0.5 Hz and 4–8 Hz) modulations of the Attended stream and the auditory cortex (STG at ~0.5 Hz, AC at 4–8 Hz) activity significantly decreased as the level of the Multitalker background progressively increased.

Corticovocal coherence in the absence of multitalker background noise

In line with previous studies (Bourguignon et al., 2013; Clumeck et al., 2014), in the absence of Multitalker background, Coh_{att} peaked at ~0.5 Hz in the STG bilaterally. This coupling reflects the common fluctuations between the time courses of the Attended stream and the listener's STG activity occurring at the phrasal and the sentence levels (Bourguignon et al., 2013; Clumeck et al., 2014).

Similar coupling in the auditory cortex has been previously disclosed in the θ band (4–8 Hz) in subjects listening to normal and noise-vocoded sentences (Gross et al., 2013; Peelle et al., 2013). Speech envelopes reflect speech rhythmicity on which both acoustic and linguistic features are tightly coupled, and the envelope rhythmicity may thus be a part of a hierarchical rhythmic structure of speech (Peelle and Davis, 2012). Thus, although the speech envelope, as such, does not carry linguistic content, it may facilitate prediction of the forthcoming speech patterns/elements and thereby support speech comprehension as well as coordination of turn-taking behavior during conversation. Our results corroborate the findings by Gross et al. (2013) regarding the consistent right-lateralized θ band coupling that is related to the neural processing of speech content at syllabic level (Luo and Poeppel, 2007; Giraud and Poeppel, 2012; Gross et al., 2013; Peelle et al., 2013).

Auditory cortices are coupled to attended speech stream in a multitalker background

Our first key finding is that the coupling between the envelopes of the heard sounds and the MEG signals was stronger for the Attended stream than for the Global scene, even when the Attended stream was drowned within the Multitalker background (e.g.,

Table 3. Coherence in source space

	Attended stream				Multitalker background				Global scene			
Condition	Coherence	*x*	*y*	*z*	Coherence	*x*	*y*	*z*	Coherence	*x*	*y*	*z*
δ band												
No noise	0.131*	64	−32	6	—	—	—	—	—	—	—	—
	0.109*	−62	−18	0								
5 dB	0.068*	64	−24	5	0.009	61	−27	19	0.060*	64	−24	4
	0.081*	−62	−16	6	0.008	−32	50	61	0.071*	−61	−15	6
0 dB	0.084*	64	−20	1	0.009	60	−18	−2	0.064*	64	−20	1
	0.095*	−60	−17	7	0.009	−59	−9	−16	0.073*	−60	−16	7
−5 dB	0.032	62	−21	5	0.011	62	−12	−9	0.016	63	−32	−8
	0.062*	−59	−20	3	0.010	−56	−51	8	0.025*	−58	−21	8
−10 dB	0.015	63	−23	2	0.011	58	−15	1	0.013	59	−17	11
	0.013	−62	−29	−1	0.011	−53	−50	27	0.011	−53	−53	26
θ band												
No noise	0.029*	61	−14	5	—	—	—	—	—	—	—	—
	0.019	−61	−16	3								
5 dB	0.041*	61	−12	5	0.013*	61	−11	7	0.037*	61	−12	5
	0.030	−61	−19	7	0.011	−49	−19	4	0.028	−61	−19	7
0 dB	0.032*	61	−12	6	0.011*	60	−14	5	0.027*	61	−12	6
	0.025	−61	−18	6	0.011	−61	−21	4	0.021	−60	−17	8
−5 dB	0.019*	61	−12	7	0.012*	59	−12	9	0.016*	61	−10	6
	0.016	−60	−20	6	0.011	−60	−17	0	0.012	−61	−22	2
−10 dB	0.012*	61	−14	10	0.011*	61	−14	1	0.012*	61	−14	2
	0.010	−59	−16	6	0.011	−61	−21	5	0.012	−62	−20	2

Maximum corticovocal coherence values and the corresponding group-level source location expressed in MNI coordinates (*x*, *y*, *z*) for audio signal at various signal-to-noise ratios within δ (~0.5 Hz) and θ (4–8 Hz) bands. *$p < 0.05$.

SNR −5 dB) but was still intelligible. The preferential tracking of the slow (~0.5 Hz and 4–8 Hz) modulations of voice signals by the auditory cortices likely reflects the ability of these brain areas to extract the Attended stream from the Multitalker background. This finding brings further support for an object-based neural coding of the slow TE modulation of the attended speech stream in the auditory cortex. According to Simon (2014), such object-based neural coding corresponds to the neural representation of a specific auditory stream (i.e., the auditory object) isolated from the whole auditory scene, without encoding other elements of the scene. As further support for this object-based representation, the number of subjects showing significant coherence with at least one of the individual auditory components composing the Multitalker background ($Coh_{talker1\text{-}6}$) was compatible with what is expected by chance. This finding therefore demonstrates that, in a cocktail party auditory scene, not all the background talkers are represented in the listener's auditory cortex activity, and that there is no specific background talker that is especially represented across all subjects. Nevertheless, as the experimental paradigm used in this study did not allow precise monitoring of whether subjects were indeed attending to the designated speaker, we cannot totally rule out neural representation of some of these individual auditory streams related to possible transient switches of attention from one speech stream to another.

This result also extends findings from previous studies, which showed that, in a noisy background, the nonprimary auditory cortical activity does not reflect the global acoustic scene but instead follows the TE of the attended speech at frequencies >1 Hz (Ding and Simon, 2012b, 2013; Mesgarani and Chang, 2012; Zion Golumbic et al., 2013).

Interestingly, we did not find any evidence of cortical activity specifically related to the TE of the whole auditory scene (Global scene). This result challenges a recent hypothesis (Ding and Simon, 2014), according to which cortical entrainment to the low-frequency part of the TE (e.g., the δ band) is related to rhythmic acoustic stimulation in a non–speech-specific way. In contrast, our data suggest that a preferential speech-sensitive cortical entrainment (i.e., higher for the Attended stream compared with the Multitalker background or the Global scene) occurs for the very low (~0.5 Hz) TE frequency components.

Differential effects of the Multitalker background on the hemispheric dominance of speech-stream tracking

Our second major finding was that, in the presence of noise, the cortical coupling with the slow (~0.5 Hz) modulations of the Attended stream TE became left-hemisphere dominant in the δ band, whereas it was bilateral in the absence of background (No noise), and it remained right-hemisphere dominant for the θ band.

This differential effect of the Multitalker background on the hemispheric dominance of the δ and θ band couplings might reflect a difference in the functional relevance of these couplings. Indeed, some authors have suggested that the θ band coupling in AC might reflect a more automatic tracking of the physical properties of the speaker's voice (Hickok and Poeppel, 2007; Ding and Simon, 2013), whereas the left-lateralized δ band coupling in the STG could indicate an active cognitive process that promotes speech recognition (Schroeder et al., 2008; Schroeder and Lakatos, 2009).

Selective attention seems to control the coupling between cortical oscillations and the low-frequency rhythmic structure of attended acoustic stimuli (Lakatos et al., 2013). Interestingly, two simultaneous and continuous natural speech streams elicited a robust left-lateralized late attentional EEG effect, characterized by an increased scalp-positive evoked response at 190–230 ms in left hemisphere electrodes (Power et al., 2012). The authors suggested that the observed effect might represent the neural correlate of an attentional effect occurring at the semantic processing level. Based on this finding, the observed left hemisphere dominance of the δ band coupling in speech-in-noise conditions might be interpreted as an effect of selective attention on the low-frequency cortical tracking of the Attended stream in adverse auditory scenes. This hypothesis is consistent with an fMRI study in which contrasting hearing versus understanding speech-in-noise unraveled a left hemisphere-dominant temporal network

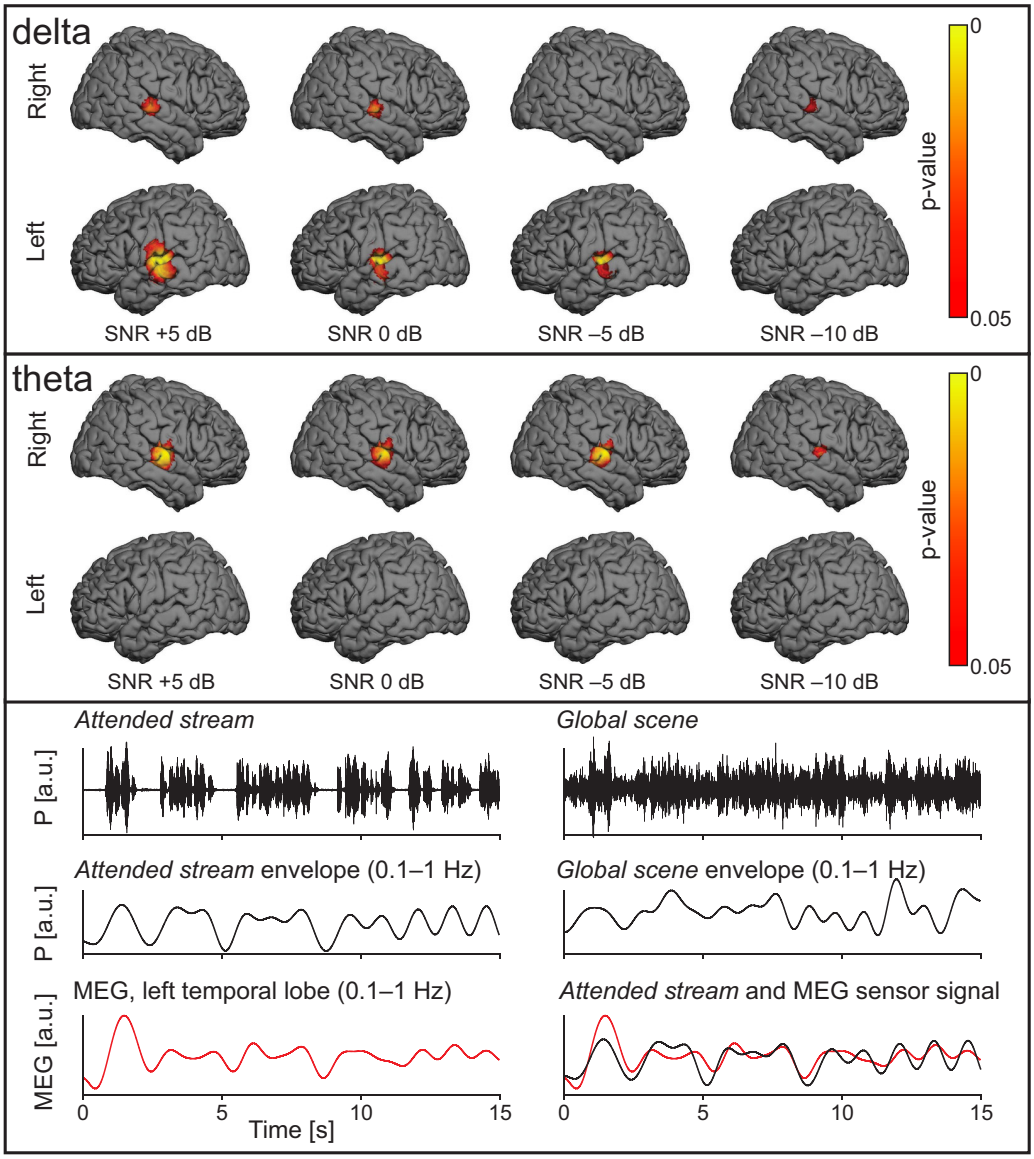

Figure 5. Top, Cortical areas sensitive to the Attended stream in speech-in-noise conditions in the δ band (~0.5 Hz). The *p* valued maps represent the contrast Coh_{att}-Coh_{global} with a threshold at statistical significance level ($p < 0.05$). Specific coupling between the Attended stream and MEG signals occurs at superior temporal gyri, with a left hemisphere dominance. Middle, Same illustration but for the θ band (4–8 Hz). Specific coupling between the Attended stream and MEG signals was observed at the right supratemporal auditory cortex, but also at the left superior temporal gyrus in the 0 and −5 dB condition. (The coherence at left auditory cortices was not significant in Coh_{att} and Coh_{global} both at sensor and source levels.) Bottom, Comparison between sound time courses and left temporal-lobe MEG signals in a typical subject. Top, Left, Time course of the Attended stream. Top, Right, The same sample of voice signal but merged with the Multitalker background at an SNR of 0 dB (Global scene). Middle, Same audio signals as above but bandpass filtered through 0.1–1 Hz. Bottom, Time course of a left temporal MEG sensor showing that the coupling with the slow temporal fluctuations was stronger with the Attended stream than with the Global scene (same MEG sensor signal displayed on the left and the right).

(Bishop and Miller, 2009). In addition, this left hemisphere dominance of δ band coupling in noisy conditions could also be attributed to the correct identification of the attended reader's voice. Alain et al. (2005) indeed demonstrated the existence of a left hemisphere thalamocortical network, including STG, associated with the correct identification of two concurrent auditory streams.

Further studies in which the level of attention would be manipulated are required to clarify the functional relevance of the observed change in the δ band coupling hemispheric dominance in noisy conditions.

Corticovocal coherence is sensitive to a Multitalker background noise

The last finding of the present study is that the coupling between bilateral STG and right AC activity and the slow (~0.5 Hz and 4–8 Hz) modulations of the Attended stream significantly decreased as the level of the Multitalker background progressively increased. This noise sensitivity was observed even for low Multitalker background level (5 dB; i.e., when the Attended stream was still easily intelligible). In our most adverse condition (−10 dB), significant coherence was only observed in the θ band.

These results partially challenge previous findings that suggested a relative noise insensitivity of neural synchronization with slow (<4 Hz) speech TE modulations even when intelligibility decreased, whereas the neural entrainment in the 4–8 Hz range appeared to be more sensitive to increasing noise levels (Ding and Simon, 2013). The present study therefore suggests that the neural coupling with speech TE modulations occurring at frequencies <1 Hz is also noise-sensitive.

In the θ band, the phase coupling between audio and MEG signals decreases when the intelligibility of the voice stimulus decreases (Peelle et al., 2013). What happens for frequencies <1 Hz might, however, be somewhat different. Indeed, we previously observed similar δ band coupling and similar cortical sources when native French-speaking subjects listened to comprehensible (French) versus incomprehensible (Finnish) speech (Bourguignon et al., 2013). The finding was interpreted to reflect neural coupling to prelinguistic, acoustic features of the speech sounds (Bourguignon et al., 2013), thereby questioning the functional relevance of the coupling for speech recognition (as defined by Hickok and Poeppel, 2007). We therefore decided not to thoroughly investigate the level of speech understanding in speech-in-noise conditions, but we rather used a subjective (i.e., VAS) speech-intelligibility score. Indeed, as speech intelligibility inevitably declines with decreasing SNR in speech-in-noise conditions, the observed correlations between the SNR, coherence levels, and speech intelligibility are trivial and do not allow to draw any conclusions about the functional relevance (i.e., role for speech recognition) of the observed couplings. Further studies are needed to clarify this major issue.

References

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc Natl Acad Sci U S A 98:13367–13372. CrossRef Medline

Alain C, Reinke K, McDonald KL, Chau W, Tam F, Pacurar A, Graham S (2005) Left thalamo-cortical network implicated in successful speech separation and identification. Neuroimage 26:592–599. CrossRef Medline

Ashburner J, Friston KJ (1999) Nonlinear spatial normalization using basis functions. Hum Brain Mapp 7:254–266. CrossRef Medline

Ashburner J, Neelin P, Collins DL, Evans A, Friston K (1997) Incorporating prior knowledge into image registration. Neuroimage 6:344–352. CrossRef Medline

Bishop CW, Miller LM (2009) A multisensory cortical network for understanding speech in noise. J Cogn Neurosci 21:1790–1805. CrossRef Medline

Bortel R, Sovka P (2007) Approximation of statistical distribution of magnitude squared coherence estimated with segment overlapping. Signal Process 87:1100–1117. CrossRef

Bourguignon M, Jousmäki V, Op de Beeck M, Van Bogaert P, Goldman S, De Tiège X (2012) Neuronal network coherent with hand kinematics during fast repetitive hand movements. Neuroimage 59:1684–1691. CrossRef Medline

Bourguignon M, De Tiège X, Op de Beeck M, Ligot N, Paquier P, Van Bogaert P, Goldman S, Hari R, Jousmäki V (2013) The pace of prosodic phrasing couples the listener's cortex to the reader's voice. Hum Brain Mapp 34:314–326. CrossRef Medline

Bourguignon M, Piitulainen H, De Tiège X, Jousmäki V, Hari R (2015) Corticokinematic coherence mainly reflects movement-induced proprioceptive feedback. Neuroimage 106:382–390. CrossRef Medline

Carrette E, Op de Beeck M, Bourguignon M, Boon P, Vonck K, Legros B, Goldman S, Van Bogaert P, De Tiège X (2011) Recording temporal lobe epileptic activity with MEG in a light-weight magnetic shield. Seizure 20:414–418. CrossRef Medline

Cherry EC (1953) Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am 25:975–979. CrossRef

Clumeck C, Suarez Garcia S, Bourguignon M, Wens V, Op de Beeck M, Marty B, Deconinck N, Soncarrieu MV, Goldman S, Jousmäki V, Van Bogaert P, De Tiège X (2014) Preserved coupling between the reader's voice and the listener's cortical activity in autism spectrum disorders. PLoS One 9:e92329. CrossRef Medline

Dale AM, Sereno MI (1993) Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. J Cogn Neurosci 5:162–176. CrossRef Medline

De Tiège X, Op de Beeck M, Funke M, Legros B, Parkkonen L, Goldman S, Van Bogaert P (2008) Recording epileptic activity with MEG in a light-weight magnetic shield. Epilepsy Res 82:227–231. CrossRef Medline

Ding N, Simon JZ (2012a) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J Neurophysiol 107:78–89. CrossRef Medline

Ding N, Simon JZ (2012b) Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci U S A 109:11854–11859. CrossRef Medline

Ding N, Simon JZ (2013) Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J Neurosci 33:5728–5735. CrossRef Medline

Ding N, Simon JZ (2014) Cortical entrainment to continuous speech: functional roles and interpretations. Front Hum Neurosci 8:311. CrossRef Medline

Drullman R, Festen JM, Plomp R (1994a) Effect of reducing slow temporal modulations on speech reception. J Acoust Soc Am 95:2670–2680. CrossRef

Drullman R, Festen JM, Plomp R (1994b) Effect of temporal envelope smearing on speech reception. J Acoust Soc Am 95:1053–1064. CrossRef Medline

Faes L, Pinna GD, Porta A, Maestri R, Nollo G (2004) Surrogate data analysis for assessing the significance of the coherence function. IEEE Trans Biomed Eng 51:1156–1166. CrossRef Medline

Fishman YI, Steinschneider M, Micheyl C (2014) Neural representation of concurrent harmonic sounds in monkey primary auditory cortex: implications for models of auditory scene analysis. J Neurosci 34:12425–12443. CrossRef Medline

Fullgrabe C, Stone MA, Moore BC (2009) Contribution of very low amplitude-modulation rates to intelligibility in a competing-speech task (L). J Acoust Soc Am 125:1277–1280. CrossRef Medline

Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci 15:511–517. CrossRef Medline

Greenberg S, Carvey H, Hitchcock L, Chang S (2003) Temporal properties of spontaneous speech: a syllable-centric perspective. J Phonetics 31:465–485. CrossRef

Gross J, Kujala J, Hämäläinen M, Timmermann L, Schnitzler A, Salmelin R (2001) Dynamic imaging of coherent sources: studying neural interactions in the human brain. Proc Natl Acad Sci U S A 98:694–699. CrossRef Medline

Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, Garrod S (2013) Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol 11:e1001752. CrossRef Medline

Halliday DM, Rosenberg JR, Amjad AM, Breeze P, Conway BA, Farmer SF (1995) A framework for the analysis of mixed time series/point process data-theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. Prog Biophys Mol Biol 64:237–278. CrossRef Medline

Hämäläinen ML, Mosher J (2010) Anatomically and functionally constrained minimum-norm estimates. In: MEG: an introduction to methods (Hansen P, Kringelbach M, Salmelin R, eds), pp 186–215. New York: Oxford UP

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402. CrossRef Medline

Hoen M, Meunier F, Grataloup CL, Pellegrino F, Grimault N, Perrin F, Perrot X, Collet L (2007) Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. Speech Commun 49:905–916. CrossRef

Koskinen M, Seppä M (2014) Uncovering cortical MEG responses to listened audiobook stories. Neuroimage 100:263–270. CrossRef Medline

Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013) The spectrotemporal filter mechanism of auditory selective attention. Neuron 77:750–761. CrossRef Medline

Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54:1001–1010. CrossRef Medline

McDermott JH (2009) The cocktail party problem. Curr Biol 19:R1024–R1027. CrossRef Medline

Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485:233–236. CrossRef Medline

Nichols TE, Holmes AP (2002) Nonparametric permutation tests for func-

tional neuroimaging: a primer with examples. Hum Brain Mapp 15:1–25. CrossRef Medline

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh Inventory. Neuropsychologia 9:97–113. CrossRef Medline

Peelle JE, Davis MH (2012) Neural oscillations carry speech rhythm through to comprehension. Front Psychol 3:320. CrossRef Medline

Peelle JE, Gross J, Davis MH (2013) Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb Cortex 23:1378–1387. CrossRef Medline

Perrin F, Grimaud N (2005) Fonds Sonores v-1.0. Available at: http://crnlgerland.univ-lyon1.fr/download/FondsSonores.html

Power AJ, Foxe JJ, Forde EJ, Reilly RB, Lalor EC (2012) At what time is the cocktail party? A late locus of selective attention to natural speech. Eur J Neurosci 35:1497–1503. CrossRef Medline

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 336:367–373. CrossRef Medline

Rosenberg JR, Amjad AM, Breeze P, Brillinger DR, Halliday DM (1989) The Fourier approach to the identification of functional coupling between neuronal spike trains. Prog Biophys Mol Biol 53:1–31. CrossRef Medline

Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci 32:9–18. CrossRef Medline

Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. Trends Cogn Sci 12:106–113. CrossRef Medline

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304. CrossRef Medline

Simon JZ (2014) The encoding of auditory objects in auditory cortex: insights from magnetoencephalography. Int J Psychophysiol 95:184–190. CrossRef Medline

Simpson SA, Cooke M (2005) Consonant identification in N-talker babble is a nonmonotonic function of N. J Acoust Soc Am 118:2775–2778. CrossRef Medline

Suppes P, Lu ZL, Han B (1997) Brain wave recognition of words. Proc Natl Acad Sci U S A 94:14965–14969. CrossRef Medline

Suppes P, Han B, Lu ZL (1998) Brain-wave recognition of sentences. Proc Natl Acad Sci U S A 95:15861–15866. CrossRef Medline

Suppes P, Han B, Epelboim J, Lu ZL (1999) Invariance between subjects of brain wave representations of language. Proc Natl Acad Sci U S A 96:12953–12958. CrossRef Medline

Taulu S, Simola J, Kajola M (2005) Applications of the signal space separation method. IEEE Trans Signal Process 53:3359–3372. CrossRef

Wang R, Perreau-Guimaraes M, Carvalhaes C, Suppes P (2012) Using phase to recognize English phonemes and their distinctive features in the brain. Proc Natl Acad Sci U S A 109:20685–20690. CrossRef Medline

Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013) Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." Neuron 77:980–991. CrossRef Medline