Delta(but not theta)-band cortical entrainment involves speech-specific processing

Nicola Molinaro[1,2], Mikel Lizarazu[1,3]

[1]BCBL, Basque center on Cognition, Brain and Language, Donostia/San Sebastian, Spain

[2]Ikerbasque, Basque foundation for Science, Bilbao, Spain

[3]LSCP, Département d'études cognitives, ENS, EHESS, PSL Research University, CNRS, Paris France

Running title: Speech specific delta-band entrainment

Number of pages: 19; Number of Figures: 6; Number of equations: 3

Word count, manuscript: 4589; abstract: 234

Keywords: Speech processing; Coherence; Neuronal oscillations; Cortical entrainment; Inferior frontal cortex

Corresponding author
Nicola Molinaro
BCBL, Basque center on Cognition, Brain and Language
Paseo Mikeletegi, 69, 2º
20009, Donostia/San Sebastian
Spain
Phone: +34 943 309 300 ext. 206
Email: n.molinaro@bcbl.eu
ORCID: 0000-0002-7549-6042

ABSTRACT

Cortical oscillations phase-align to the quasi-rhythmic structure of the speech envelope. This speech-brain entrainment has been reported in two frequency bands, i.e., both in the theta band (4 – 8 Hz) and in the delta band (< 4 Hz). However, it is not clear if these two phenomena reflect passive synchronization of the auditory cortex to the acoustics of the speech input, or if they reflect higher processes involved in actively parsing speech information. Here we report two magnetoencephalography experiments in which we contrasted cortical entrainment to natural speech compared to qualitative different control conditions (Experiment 1: amplitude modulated white-noise; Experiment 2: spectrally rotated speech). We computed the coherence between the oscillatory brain activity and the envelope of the auditory stimuli. At the sensor level, we observed increased coherence for the delta and the theta band for all conditions in bilateral brain regions. However, only in the delta band (but not theta) speech entrainment was stronger than either of the control auditory inputs. Source reconstruction in the delta band showed that speech, compared to the control conditions, elicited larger coherence in the right superior temporal and left inferior frontal regions. In the theta band, no differential effects were observed for the speech compared to the control conditions. These results suggest that whereas theta entrainment mainly reflects perceptual processing of the auditory signal, delta entrainment involves additional higher order computations in the service of language processing.

INTRODUCTION

When people listen to speech, the phase of low-frequency (< 8 Hz) cortical oscillations entrain to slow temporal modulations in the acoustic envelope. This mechanism has been hypothesized to be crucial for speech recognition (Ghitza, 2011; Giraud & Poeppel, 2012). However, it is still under debate whether cortical entrainment to speech underlies pure auditory perceptual processing or if it additionally reflects higher-order processes involved in actively parsing speech information (for discussion Ding & Simon, 2014). The former view suggests that cortical oscillations in the auditory regions passively entrain to the temporal modulations of the external auditory signal. The information extracted from this "entrainment step" would then be available to higher-order cognitive processes that are not necessarily phase-synchronized with the external input. The latter view underscores that cortical speech entrainment can reflect the brain tendency to align a larger set of high-order cognitive processes to the speech rhythmicity (for a review: Meyer, in press). In this way, the neurocognitive network involved in speech processing would be selectively active in the time intervals in which relevant content information is conveyed by the speech input. In the present study, we provide evidence that these two hypotheses are differently reflected by the cortical entrainments observed at different frequency bands.

Cortical entrainment to the speech envelope has been mainly seen in delta (< 3 Hz) and theta (4 – 8 Hz) frequency bands. Theta entrainment correlates with the syllabic patterns of speech across languages (Ding et al., 2017) and would be involved in creating syllabic-level language representations. Given this empirical observation, it has been initially proposed that syllables constitute the basic unit for initial speech segmentation (Ghitza, 2013; Hickok, 2009). Syllabic units would then constrain the

access to segmental information (i.e., phonemes) through a cross-frequency coupling mechanism. The phase of theta has been shown to modulate the amplitude of gamma oscillations (> 30 Hz: cross-frequency coupling, Gross et al., 2013), and this mechanism would reflect the hierarchical clustering of individual phonemes (sampled at gamma) into syllables (Giraud & Poeppel, 2012). A large number of studies also reported strong speech-brain coupling in the delta band (e.g., Bourguignon et al., 2013). Temporal modulations in the delta band of speech reflect the intonational stress patterns for word and phrases (see Keitel et al., 2017). Delta entrainment thus shows the brain sensitivity to these prosodic cues (Bourguignon et al., 2013).

That said, it is still not clear whether delta and theta entrainment differ, if they reflect the same basic acoustics-related mechanism or if they involve higher-order mechanisms involved in speech processing. It has been observed that both frequency bands are related to speech intelligibility (for a discussion, Ding & Simon, 2014) across a variety of languages (English: Gross et al., 2013; French: Bourguignon et al., 2013; Spanish: Molinaro et al., 2016). Importantly, delta phase modulates the amplitude of theta band oscillations (Gross et al., 2013) suggesting a hierarchy "delta > theta > gamma" in which the delta oscillations guide the rhythmic activity at higher frequencies. Following this view, delta oscillations potentially reflect a high-order mechanism (sensitive to both attentional and linguistic influences) that top-down modulates entrainment in higher frequencies.

In this magnetoencephalography (MEG) study, we better characterize the entrainment of brain activity to speech in delta and theta frequency bands. Specifically, we used coherence to measure the cortical entrainment to speech and to different non-linguistic auditory signals across two experiments. Coherence analysis was performed at the MEG sensor-level and in the source space (source-level). In the

first experiment, we compared the cortical entrainment to Speech with the cortical entrainment to AM white-noise at a fixed rate in delta and theta bands. In the second experiment, we compared cortical entrainment to Speech with cortical entrainment to Rotated Speech in the two frequency bands of interest. In the first experiment the control conditions provided a stable AM rate, while in the second experiment the variability and the complexity of the speech temporal modulations was preserved in the control condition. If delta and/or theta speech-brain coupling involves higher-order speech computations required for speech listening, the entrainment should extend beyond auditory regions in the temporal lobes, possibly involving more frontal lobe regions (Hickok & Poeppel, 2007; Park et al., 2015).

EXPERIMENT 1

METHODS

Participants

Thirty-five participants (17 females) took part in the first experiment (mean age=24 yrs., age range=16 – 48 yrs.). All participants were right-handed Spanish native speakers with no hearing impairments. The BCBL ethics committee approved the experiment (following the principles of the Declaration of Helsinki) and all participants signed an informed consent.

Stimuli and procedure

Participants performed two experimental blocks. In the first block, participants listened to natural Speech while sitting in the MEG facility (*Speech* condition). The Speech consisted of thirty meaningful sentences ranging in duration from 7.42 to 12.65 seconds (M=9.9; SD=1.13). Sentences were uttered by a Spanish native female speaker and digitized at 44.1 kHz using a digital recorder (Marantz PMD670). Each

trial began with a 1 sec long auditory tone (at 500 Hz tone) followed by a 2 secs-long silence before the sentence presentation. After the end of each sentence, a comprehension question about the content of the last stimulus was presented auditorily. Participants answered the question by pressing the corresponding button on the response pads (Yes/No, response side counterbalanced).

In the second block, participants listened to stimuli obtained by rhythmically modulating the amplitude (AM) of white-noise sound either in the delta or the theta band. The stimuli were generated at a sampling frequency of 44.1 kHz and modulated using Matlab R2010 functions. AMs were applied to different stimuli at the following frequencies: 2 (*2 Hz AM white-noise* condition) and 7 Hz (*7 Hz AM white-noise* condition) rates with 100% depth. The selection of these two frequency bands was based on previous pilot data on speech-brain coherence in which we observed these two coherence peaks. All stimuli lasted 10 secs and appeared 30 times throughout the experimental block. Participants were not instructed to pay attention to the auditory stimuli and watched an unrelated silent movie on the screen.

The presentation order of the blocks was balanced across participants. In both experimental blocks, the stimuli were delivered to both ears via plastic tubes. The sound level was fixed at 75 dB for all the participants. Participants were encouraged to take a break between blocks and were asked to avoid head movements and blinks during the auditory presentation. After these experimental blocks, resting state MEG was recorded for ~5 minutes with the participants' eyes being close.

Data acquisition and pre-processing

MEG data were acquired in a magnetically shielded room using the whole-scalp system (Elekta-Neuromag, Helsinki, Finland). Head position was continuously monitored using four Head Position Indicator (HPI) coils. The location of each coil

relative to the anatomical fiducials (nasion, left and right preauricular points) was defined with a 3D digitizer (Fastrak Polhemus, Colchester, VA, USA). Digitalization of the fiducials plus ~150 additional points evenly distributed over the scalp of the participant were used during subsequent data analysis to spatially align the MEG sensor coordinates with T1 magnetic resonance brain images acquired on a 3T MRI scan (Siemens Medical System, Erlangen, Germany). MEG recordings were acquired continuously with a bandpass filter at $0.01 - 330$ Hz and a sampling rate of 1 kHz. Eye-movements were monitored with two pairs of electrodes in a bipolar montage placed on the external chanti of each eye (horizontal electrooculography EOG) and above and below right eye (vertical EOG). Similarly, cardiac rhythm was monitored using two electrodes, placed on the right side of the participants' abdomen and below the left clavicle.

MEG data were pre-processed off-line using the Signal-Space-Separation (SSS) method (Taulu & Kajola, 2005) implemented in Maxfilter 2.1 (Elekta-Neuromag) to remove external magnetic noise from the MEG recordings. MEG data were also corrected for head movements and referenced to the initial head position. Bad channels detected during the acquisition were substituted using MaxFilter interpolation algorithms.

Subsequent analyses were performed using Matlab R2010 (Mathworks, Natick, MA, USA). Heart beat and EOG artifacts were detected using Independent Component Analysis (ICA) and linearly subtracted from the MEG data (Infomax algorithm implemented in Fieldtrip toolbox, Oostenveld et al., 2011). Finally, the continuous MEG data were segmented into epochs of 1 sec during the Speech perception. Epochs with large MEG peak-to-peak amplitude values (exceeding 4000 ft in magnetometers or 4000 ft/cm in gradiometers) were considered as artifact-

contaminated and excluded from the subsequent analyses. On average, the percentage of epochs retained in the final analyses were 87% (SD: 16%), 85% (12%) and 89% (11%) for the Speech, 2 Hz AM white-noise and the 7 Hz AM white-noise respectively.

Coherence analysis

Sensor level analysis

The synchronization between neural oscillations and the envelope (Env) of auditory signals was calculated using coherence. Envelopes of the stimuli were computed by applying the Hilbert transform to the auditory signals. For each experimental condition, coherence between the artifact free epochs and the Env of the audio signals was calculated in the $1-10$ Hz frequency band with 1 Hz frequency resolution (as in Molinaro et al., 2016, see also Lizarazu et al., 2015). The data from pairs of gradiometers were linearly combined so as to maximize the coherence according to the following procedure. First, for each MEG sensor ($r\{1:102\}$), signals from gradiometer pairs were linearly combined to estimate the signal of a virtual gradiometer in the orientation θ $[0 \; ; \; \pi]$:

$$gr,\theta(t) = gr,2(t)\cos\theta + gr,2(t)\sin\theta \quad ,$$

Following Halliday et al (1995), coherence based on the Fourier transform of artifact-free epochs was then computed between Env and g_θ

$$Coh(r,f,\theta) = \frac{\left\| \langle Env(f) g_{r,\theta}^*(f) \rangle \right\|^2}{\langle |Env(f)|^2 \rangle \langle |g_{r,\theta}(f)|^2 \rangle} \quad ,$$

where * is the Hermitian conjugate and $\langle \bullet \rangle$ the mean across epochs. In practice, $Coh(r,f,\theta)$ was estimated from the cross-spectral density (CSD) matrix formed

with Env, $g_{r,1}$ and $g_{r,2}$ Env for θ spanning $[0 ; \pi]$ by steps of $\frac{\pi}{2}$. The optimum

coherence value was finally obtained as follows:

$$Coh_{op}(r) = \max_{\theta \epsilon [0;\pi]} \langle Coh(r,f,\theta) \rangle_{f \epsilon F} \quad ,$$

where $F = [0.5 - 10 \text{ Hz}]$ and $\langle \bullet \rangle$ is the arithmetic mean. Thus, a coherence value for

each (i) participant, (ii) MEG sensor (combination of gradiometer pairs) and (iii)

frequency bin below 10 Hz was obtained.

We used nonparametric permutation test (maximum statistic permutations,

Nichols & Holmes, 2002) to identified frequency bins that showed significant

coherence values at the sensor level (p<0.05). To do so, coherence values for each

frequency bin were contrasted with resting state coherence values at the same

frequency (coherence between the Env of the corresponding auditory signal and the

resting state data). The sampling distribution of the maximal difference of coherence

values (maximum taken across all sensors) was evaluated using the exhaustive

permutation test. Frequencies for which the non-permuted maximal difference

exceeded the 95th percentile of this permutation distribution were defined as

frequencies of interest, and the corresponding supra-threshold sensors were identified

for this frequency band. Significant frequencies were grouped in frequency bands of

interest for each condition. These frequency bands were selected to compute

coherence analysis in the source space.

We also used this nonparametric permutation test (Nichols & Holmes, 2002) to

identified sensors that showed significant coherence differences between conditions

(Speech vs. AM white-noise) at the frequency bands of interest.

Source level analyses

Coherence values were estimated at the source level for each participant, condition and frequency band of interest using beamforming techniques (Van Veen et al., 1997). MRI images were segmented into scalp, skull, and brain components using the segmentation algorithms implemented in Freesurfer (Martinos Center of Biomedical Imaging, MQ). The forward model was computed for three orthogonal tangential current dipoles, which were distributed on a 5-mm homogeneous grid source space covering the whole brain. For each source (three directions), the forward model was then reduced to its two principal components of highest power. The CSD matrix between the artifact free epochs and the envelope of the auditory signals was calculated at each frequency of interest. Based on the forward model and the CSD matrix, brain coherence maps were produced using DICS algorithm (Gross et al., 2001). Following this procedure, we also calculated brain coherence maps for the resting state (coherence between the amplitude of the corresponding auditory signal and the resting state data).

A non-linear transformation from individual MRIs to the standard Montreal Neurological Institute (MNI) brain was first computed using the spatial normalization algorithm implemented in SPM8 (Wellcome Department of Cognitive Neurology, London, UK). This was then applied to every individual brain coherence map. We used a nonparametric permutation test to identify brain sources that showed significant (p<0.05) coherence values (Nichols & Holmes, 2002) for each condition. Coherence values in the frequency band of interest were contrasted with resting state coherence values at the corresponding frequency band (Molinaro et al., 2016). This nonparametric permutation test was used to identify brain sources that showed significant coherence difference between conditions (Speech vs. AM white-noise) at the frequency band of interest.

RESULTS

Behavioral data

Participants answered correctly to the comprehension questions following the Speech stimuli with an overall accuracy of 96.43% (Standard Deviation: 2.68, min=89; max=100).

Sensor-level

First, we evaluated the frequency bins at which coherence was higher at the sensor-level for each condition (Speech, 2 Hz AM white-noise and 7 Hz AM white-noise) (Fig. 1). In the Speech condition, two bands of interest were identified in which coherence values were significantly ($p < 0.05$) higher compared to the resting state. The first frequency band fell in the delta (1 – 2 Hz) range and the second band in the theta (6 – 7 Hz) range. In the 2 Hz AM white-noise condition, coherence values were significantly higher ($p < 0.05$) at 2 Hz compared to resting state. Finally, in the 7 Hz AM white-noise condition, coherence values were significantly higher at 7 Hz compared to resting state. Significant coherence values emerged bilaterally in the sensor space for all conditions and frequency bands of interest (Fig. 1). Importantly, when contrasting the coherence values across sensors for the delta band in the Speech condition with the 2 Hz AM white-noise condition, coherence values showed to be higher for Speech in the right temporal sensors and in the left hemisphere frontal-temporal sensors (Fig. 3). The comparison between the Speech theta band coherence values and the 7 Hz AM white-noise condition on the opposite did not highlight any significant effect. Both the topographical distribution of the effect and its magnitude in coherence were highly similar across conditions for this latter contrast.

-- please insert Figure 1 around here --

Source-level

The frequency bands of interest identified by the sensor-level analyses were further investigated with source reconstruction (Fig. 2). In the Speech condition, significant (p<0.05) delta (1 – 2 Hz) coherence emerged in bilateral superior temporal and left frontal regions, whereas significant (p<0.05) theta coherence was located in bilateral superior temporal regions. In the 2 Hz and 7 Hz AM white-noise conditions, coherence values were significantly (p<0.05) stronger at the corresponding frequency in bilateral superior temporal regions.

-- please insert Figure 2 around here --

Importantly, the coherence values obtained for the delta band in Speech were significantly higher (p<0.05) than for the 2 Hz AM white-noise in right superior temporal and left inferior frontal regions (Fig. 3). No significant differences (all p-values>0.05) emerged between the coherence maps obtained for the theta Speech rhythms and the 7 Hz AM white-noise.

-- please insert Figure 3 around here --

EXERIMENT 2

METHODS

Participants

Thirty-five participants (19 females) took part in the second experiment (mean age=27 yrs., age range=18 – 51 yrs.). Participants in Experiment 1 and 2 matched by age (t(68) = 1.01, p > .31). All participants were right-handed Spanish native speakers with no hearing impairments. The BCBL ethics committee approved the experiment (following the principles of the Declaration of Helsinki) and all participants signed an informed consent.

Stimuli and procedure

Participants listened to natural Speech and Rotated Speech in the MEG. The Speech stimuli consisted of the same sentences used in the first block of the first experiment. The Rotated Speech involved spectral inversion of the original sentence. Spectrally-Rotated Speech was produced by flipping the spectrum of the original sentences around a center frequency of 1.5 kHz by applying a custom digital implementation of the original algorithm (Blesser, 1972). The Rotated Speech has very similar temporal and spectral complexity to ordinary Speech, but it is not intelligible. Each trial began with a 1 sec long auditory tone (at 500 Hz tone) followed by a 2 sec-long silence before the sentence (Speech and Rotated Speech) presentation. After each sentence, participants listened to a word segment of the speech stimuli and they had to decide if the stimulus appeared in the previous speech or not. In 50% of the cases, the word segment was contained in the sentence. Participants were instructed to press the "yes" button on the response pads if the word was present in the previously presented sentence and "no" if it was not (response side counterbalanced).

Stimulus presentation parameters were the same of Experiment 1. Order of the block was counterbalanced across participants. Participants were encouraged to take a break between blocks and were asked to avoid head movements and blinks during the auditory presentation. After these experimental blocks, resting state MEG was recorded for ~5 min with the participants' eyes being close.

Data acquisition and analyses

The analysis procedures employed for Experiment 2 were the same as the ones described for Experiment 1. On average, the percentage of epochs retained in the final analyses was 88% (SD: 12%) and 87% (11%) for the Speech and the Rotated Speech respectively.

RESULTS

Behavioral data

Participants recognized the word segments at the end of the Speech stimuli with an accuracy of 96.54% (SD=2.45, min=89; max=100). The accuracy for the segments in the Rotated Speech was lower 55.37% (SD=4.70, min=50; max=68) but significantly higher than chance level (p<0.05).

Sensor-level

Both in the Speech and the Rotated Speech conditions, the coherence values were significantly (p<0.05) higher compared to the resting state in delta (1 – 2 Hz) and theta (6 – 7 Hz) bands (Fig. 4). Significant coherence values emerged in bilateral sensors in all conditions and frequency bands of interest (i.e., in both the delta and the theta band). The statistical comparison between Speech and Rotated Speech in the delta band showed stronger entrainment for Speech both in the right temporal and left frontal-temporal sensors (Fig. 6). The statistical comparison between these two conditions in the theta band did not show any reliable statistical difference.

-- please insert Figure 4 around here --

Source-level

In the Speech condition, significant (p<0.05) delta (1 – 2 Hz) coherence emerged in bilateral superior temporal and left frontal regions, whereas significant (p<0.05) theta coherence (6 – 7 Hz) was evident in bilateral superior temporal regions (compared to resting state, Fig. 5). In the Rotated Speech condition, coherence values were significantly (p<0.05) stronger (compared to resting state) in bilateral superior temporal regions both in delta and theta frequency bands (Fig. 5).

-- please insert Figure 5 around here --

14

Importantly, the coherence values obtained for the delta band in the Speech were significantly higher (p<0.05) than for the delta band in the Rotated Speech in right superior temporal and left inferior frontal regions (Fig. 6). No significant differences (all p-values>0.05) were obtained between the Speech and the Rotated Speech in the theta band.

-- please insert Figure 5 around here --

DISCUSSION

The two experiments showed that cortical entrainment in the delta (but not in theta) band was stronger for the speech compared to the control conditions (AM white-noise and rotated speech). In both experiments, the effects emerged in the right temporal and the left inferior frontal cortex. In Experiment 1, the envelope of the 2 Hz white-noise signal was periodic, while the envelope of the Speech signal was quasi-rhythmic. Rhythmic variability made the Speech envelope more complex than the AM white-noise envelope signal, and this could have triggered the need for additional processing resources. However, this explanation cannot hold for Experiment 2, where the envelope of the Rotated Speech condition preserves the temporal modulation patterns (and thus the complexity) of the original Speech envelope. Another possible explanation of the delta coherence differences observed in Experiment 1 could be ascribed to differences in the task. Participants performed an active comprehension task during the Speech listening condition, but passively listened to the 2 Hz AM during the white-noise condition while watching a silent movie. In Experiment 2, however, we employed the same task for both the Speech and the Rotated Speech condition and reported the same effects observed in Experiment 1, thus ruling out the explanation that the differential delta effect across the two experiments was due to

differences between the tasks. Thus, to account for the effects observed in delta, we turn to a different explanation.

Before doing so, however, we note the lack of differential coherence in the theta band. In both Experiments, we did not observe cortical entrainment differences (either at the sensor or at the source level) for the speech compared to the control conditions in the theta band. Even if this is a null effect, it should be noted that the brain coherence maps for the Speech coherence in theta, the 7 Hz AM white-noise condition in Experiment 1 (Figure 2) and to the Rotated Speech in Experiment 2 (Figure 5) are highly similar, showing a right lateralized effect mainly restricted to the auditory regions. Theta coherence was similar both in magnitude and in the brain distribution of the effect in both Experiments. The fact that there was no effect in Experiment 1 is even more striking given the largely different scenarios (active task for the Speech condition and passive listening for the 7 Hz AM white-noise) in which coherence was estimated within the same group of participants.

Based on the present findings, we can conclude that the cortical speech entrainment observed in the theta band likely reflects a passive synchronization of the auditory cortices with the acoustic regularities of speech. On the other hand, the systematic differences observed in the delta band for speech compared to different control conditions in different experimental scenarios suggest that a larger set of higher-order processes - beyond the basic acoustic analysis of the speech signal - contribute to the delta entrainment observed for the speech. Recent data from our lab suggest that this would be true also for the cortical entrainment observed in right temporal regions. In a recent study (Bourguignon et al., in press), we took advantage of a location comparison procedure and showed that whereas the right hemisphere peak of the theta coherence emerged close to the primary auditory cortex, the right-

hemisphere delta peak was located ~15 mm more posteriorly along the superior temporal gyrus, in a higher-order auditory processing region. This MEG location evidence supports the idea that theta entrainment reflects activity related to the acoustic analysis of the auditory input, whereas delta entrainment reflects an additional set of higher order speech processing mechanisms.

This observation is supported by a number of recent studies that highlight how cortical entrainment in delta contributes to additional processes that go beyond auditory parsing. Along these lines, the data provided by Ding et al. (2016) showed that low-frequency oscillatory activity can reflect tracking of syntactic structures in connected speech even in the absence of prosodic contours. In a similar vein, Meyer et al. (2017) observed that delta-band phase information reflects individual variability in syntactically grouping words into larger phrases, independent of the prosodic structure of the speech input. Delta-band oscillatory brain activity thus not only reflects the acoustic tracking of the speech envelope, but also entails higher order processes involved in the syntactic structuring of the input (Ding et al., 2016; Meyer et al., 2017).

Delta-band entrainment modulations (Vander Ghinst et al., 2016) have been observed for speech processing in increasingly noisy conditions, with the left temporal regions recruited more for stimuli requiring increased attentional control. Interestingly, no effect was observed for theta cortical entrainment (see also Giordano et al., 2017, for audio-visual speech integration). A possible mechanism for such attentional modulation has been reported by Park et al. (2015). They showed that left frontal oscillatory activity in the delta channel causally modulates the speech-entrained oscillations in the left auditory regions in temporal cortex. Importantly, the speech-brain coupling was enhanced by top-down control and the effect was related to

speech intelligibility. The frontal lobe regions discussed by these authors are very similar to the ones we report in this study, involving the pars opercularis of the inferior frontal cortex extending to more dorsal regions more involved in motor programming (see Figures 2 and 5). These controlled processes during speech perception have been suggested to support turn-taking in conversational settings (Friston & Frith, 2015; Levinson, 2016), where a listener has to plan what she/he is going to utter once the talker has concluded her/his message.

Finally, delta-band cortical entrainment has been associated with normal compared to impaired reading. In a recent study (Molinaro et al., 2016), we reported differential delta entrainment between dyslexic and control readers in the right temporal cortex and in the left inferior frontal gyrus (pars opercularis). Effective connectivity analyses indicated that the strength of the coupling between these two regions is positively related to the development of normal phonological skills. Again, no specific role for the theta entrainment was observed for reading acquisition.

The similar network recruited for the theta entrainment for both speech and control stimuli suggests that this latter phenomenon is more tightly linked to the acoustic analysis of the input independent from its linguistic nature. According to some authors the theta entrainment could be modelled as a train of evoked responses to the syllabic modulations of the speech (Ding & Simon, 2014, for a review). Given the observed delta (phase)-theta (amplitude) coupling during speech perception (Gross et al., 2013), it can be hypothesized that the "syllabic" theta rhythm is modulated by slower oscillatory brain activity taking advantage of the prosodic (more macroscopic) structure of the speech envelope. How this cross-frequency interaction develops and through which cortical networks must be better evaluated in the future.

We underscore that while theta cortical entrainment during auditory processing seems insensitive to the linguistic nature of the input, it plays a mediation role in audio-visual speech integration between the visual and the auditory modality (Park et al., 2016; Zion Golumbic et al., 2013). In fact, entrainment of the occipital cortex to lip movements is more prominent in the theta band (compared to speech-only conditions) and is correlated with successful speech understanding. Occipital theta band lip-movement entrainment can thus facilitate speech perception. A recent study (Giordano et al., 2017), however, still highlights a relevant modulation of the delta band activity in response to differential signal-to-noise ratios during audio-visual speech integration. Interesting, higher noise levels recruited the inferior frontal regions to a larger extent.

The present study thus points to a large cortical network that shows delta entrainment. The network includes both associative regions of the auditory cortex and more frontal regions classically related to higher order processes involved in language comprehension and production. Future studies should better determine if this network takes advantage of the slow prosodic cues in the speech to top-down constrain the decoding of speech at faster rates for both syllabic and phonemic levels of processing. On the other hand, the theta entrainment emerged as a lower-level process tightly related to the acoustic features of the auditory input. This distinction seems to be supported by a number of studies in the literature (Keitel et al., 2017; for speech in noise: Vander Ghinst et al., 2016; in dyslexic populations: Molinaro et al., 2016). Future research should better determine which components of the delta entrainment are related to language processing and which are more related to top-down influences on sensory processing.

Authors contribution

NM and ML designed the experiments, collected and analysed the data and prepared the manuscript for submission.

Acknowledgements

Conflict of interest: The authors declare no potential sources of conflict of interest.

Data sharing: Data are available upon request to the authors

References

Bourguignon, M., De Tiege, X., de Beeck, M. O., 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 Map.*, **34**(2), 314-326.

Bourguignon, M., Molinaro, N. & Wens, V. (in press) Contrasting functional imaging parametric maps: the mislocation problem and alternative solutions. *Neuroimage*.

Blesser B. (1972) Speech perception under conditions of spectral transformation: I. Phonetic characteristics. *J. Speech Hear. Res.*, **15**, 5-41.

Ding, N., Melloni, L., Zhang, H., Tian, X. & Poeppel, D. (2016) Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neurosci.*, **19**, 158-164.

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C. & Poeppel, D. (2017) Temporal modulations in speech and music. *Neurosci. Biobehav. Rev.*, **81B**, 181-187.

Ding, N. & Simon, J. Z. (2014) Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.*, **8**, 311

Friston, K. J. & Frith, C. D. (2015) Active inference, communication and hermeneutics. *Cortex*, **68**, 129-143.

Ghitza, O. (2011) Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.*, **2**, 130.

Ghitza, O. (2013) The theta-syllable: a unit of speech information defined by cortical function. *Front. Psychol.*, **4**, 138.

Giordano, B. L., Ince, R. A., Gross, J., Schyns, P. G., Panzeri, S. & Kayser, C. (2017) Contributions of local speech encoding and functional connectivity to audio-visual speech perception. *eLife*, **6**, e24763.

Giraud, A. L. & Poeppel, D. (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neurosci.*, **15**(4), 511-517.

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**(12), e1001752.

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.*, **98**(2), 694-699.

Halliday, D. M., Rosenberg, J. R., Amjad, A. M., Breeze, P., Conway, B.A. & Farmer, S. F. (1995) A framework for the analysis of mixed time series/point process data. *Prog. Biophys. Mol. Biol.*, **64**, 237–278.

Hickok, G. (2009) The functional neuroanatomy of language. *Phys. Life Rev.*, **6**, 121-143.

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

Keitel, A., Gross, J. & Kayser, C. (2017) Speech tracking in auditory and motor regions reflects distinct linguistic features. *bioRxiv* 195941.

Levinson, S. C. (2016) Turn-taking in human communication–origins and implications for language processing. *Trends Cogn. Sci.*, **20**(1), 6-14.

Lizarazu, M., Lallier, M., Molinaro, N., Bourguignon, M., Paz‐Alonso, P. M., Lerma‐Usabiaga, G. & Carreiras, M. (2015) Developmental evaluation of atypical auditory sampling in dyslexia: Functional and structural evidence. *Hum. Brain Map.*, **36**(12), 4986-5002.

Meyer, L. (in press) The Neural Oscillations of Speech Processing and Language Comprehension: State of the Art and Emerging Mechanisms. *Eur. J. Neurosci.*

Meyer, L., Henry, M. J., Gaston, P., Schmuck, N. & Friederici, A. D. (2017) Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cereb. Cortex.*, **27**(9), 4293-4302

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M. & Carreiras, M. (2016) Out-of-synchrony speech entrainment in developmental dyslexia. *Hum. Brain Map.*, **37**(8), 2767-2783.

Nichols, T. E. & Holmes, A. P. (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Map.*, **15**(1), 1-25.

Park, H., Ince, R. A., Schyns, P. G., Thut, G. & Gross, J. (2015) Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.*, **25**(12), 1649-1653.

Park, H., Kayser, C., Thut, G. & Gross, J. (2016) Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *eLife*, **5**, e14521.

Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.*, **2011**: 156869.

Taulu, S. & Kajola, M. (2005) Presentation of electromagnetic multichannel data: the signal space separation method. *J. Appl. Phys.*, **97**(12), 124905.

Van Veen, B. D., Van Drongelen, W., Yuchtman, M. & Suzuki, A. (1997) Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng.*, **44**(9), 867-880.

Vander Ghinst, M., Bourguignon, M., de Beeck, M. O., Wens, V., Marty, B., Hassid, S., Choufani, G., Jousmaki, V., Hari, R., Van Bogaert, P., Goldman, S. & De Tiège, X. (2016) Left Superior Temporal Gyrus Is Coupled to Attended Speech in a Cocktail-Party Auditory Scene. *J. Neurosci.*, **36**(5), 1596-1606.

Zion Golumbic, E., Cogan, G. B., Schroeder, C. E. & Poeppel, D. (2013) Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party". *J. Neurosci.*, **33**(4), 1417-1426.

Figure captions

Figure 1. Spectra and topographic distribution of the coherence values for the Speech (Top), 2 Hz AM white-noise (Middle) and 7 Hz AM white-noise (Bottom) conditions in Experiment 1. For each condition, the coherence spectra (corrected by the coherence spectra in the resting state condition) was calculated in the $0 - 10$ Hz frequency range. The coherence spectra in representative sensors of the left and right hemisphere is represented (MEG0212/3, MEG0242/3, MEG1322/3, MEG1332/3, MEG2612/3). After the permutation test, the frequencies showing significant ($p<0.05$) coherence values during Speech condition compared to resting state are highlighted. Sensors showing significant ($p<0.05$) coherence values during each condition compared to resting state are also highlighted.

Figure 2. Source-level coherence analysis for the Speech (Top), 2 Hz AM white-noise (Bottom-Left) and 7 Hz AM white-noise (Bottom-Right) conditions in Experiment 1. Statistical map (p-values) showing significantly increased coherence for the Speech, the 2 Hz AM white-noise and the 7 Hz AM white-noise conditions compared to the resting state in the left (LH) and the right (RH) hemisphere.

Figure 3. Speech vs withe-noise AM at 2 Hz. Left: Topographic map of the coherence differences between Speech and 2 Hz AM white-noise. Sensors showing significant ($p<0.05$) higher coherence for the delta rhythms in Speech compared to the 2 Hz AM in the left (LH) and the right (RH) hemisphere are highlighted. Right: Brain map (p-values) showing significant ($p<0.05$) higher coherence for the delta rhythms in Speech compared to the 2 Hz AM in the LH and the right RH.

Figure 4. Spectra and topographic distribution of the coherence values for the Speech (Top) and Rotated Speech (Bottom) in Experiment 2. For each condition, the coherence spectra (corrected by the coherence spectra in the resting state condition) was calculated in the 0 – 10 Hz frequency range. The coherence spectra in representative sensors of the left and right hemisphere are selected (MEG0212/3, MEG0242/3, MEG1322/3, MEG1332/3, MEG2612/3). After the permutation test, the frequencies showing significant (p<0.05) coherence values during Speech condition compared to resting state are highlighted. Sensors showing significant (p<0.05) coherence values during each condition compared to resting state are also highlighted.

Figure 5. Source-level coherence analysis for the Speech (Top) and Rotated Speech (Bottom) in Experiment 2. Statistical map (p-values) showing significantly increased coherence for the Speech and Rotated Speech in delta and theta bands in the left (LH) and the right (RH) hemisphere.

Figure 6. Speech vs Rotated Speech in the delta band. Left: Topographic map of the coherence differences between Speech and Rotated Speech. Sensors showing significant (p<0.05) higher coherence for the delta rhythms in Speech compared to the Rotated Speech in the left (LH) and the right (RH) hemisphere are highlighted. Right: Brain map (p-values) showing significant (p<0.05) higher coherence for the delta rhythms in Speech compared to the Rotated Speech in the LH and the right RH.
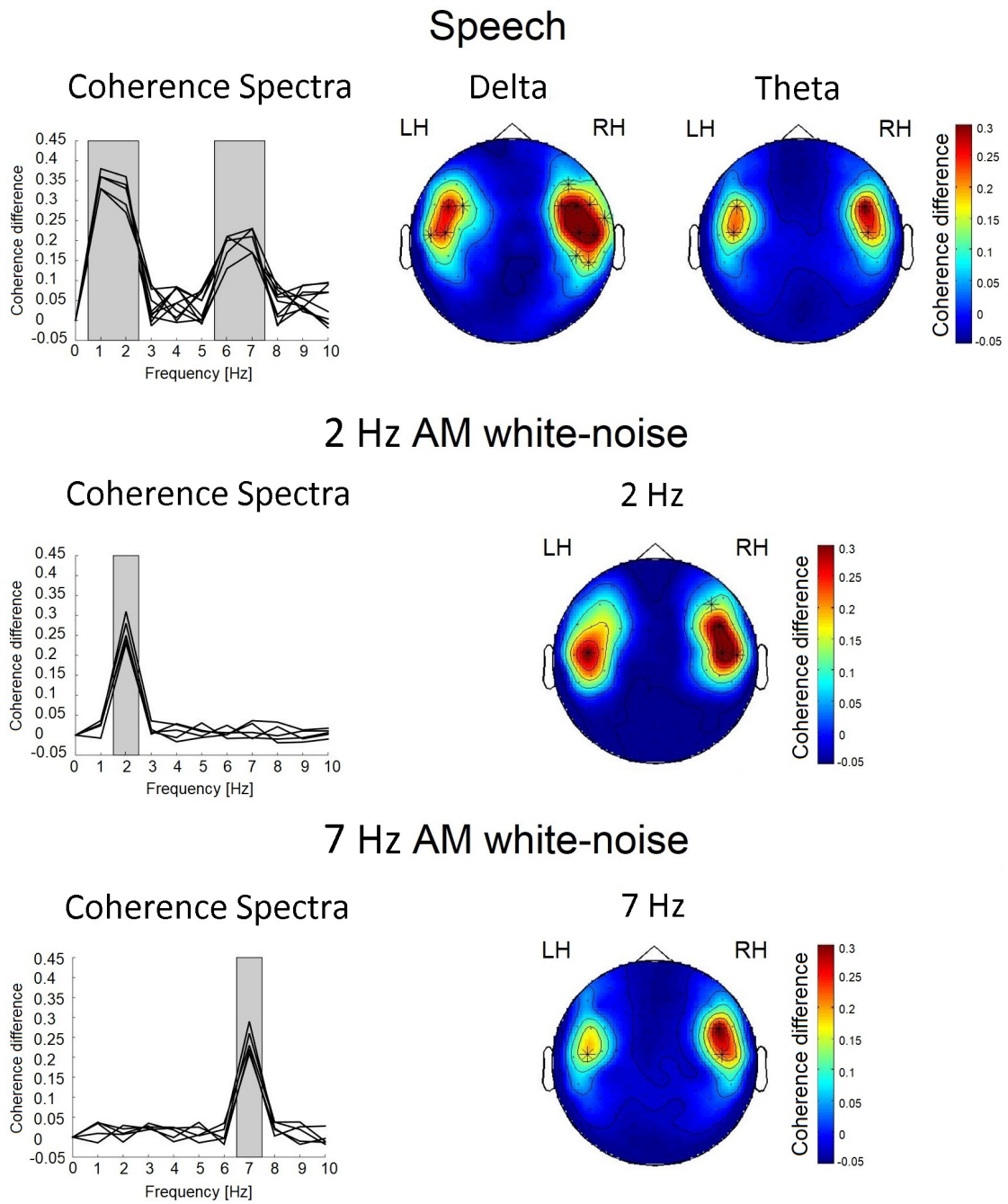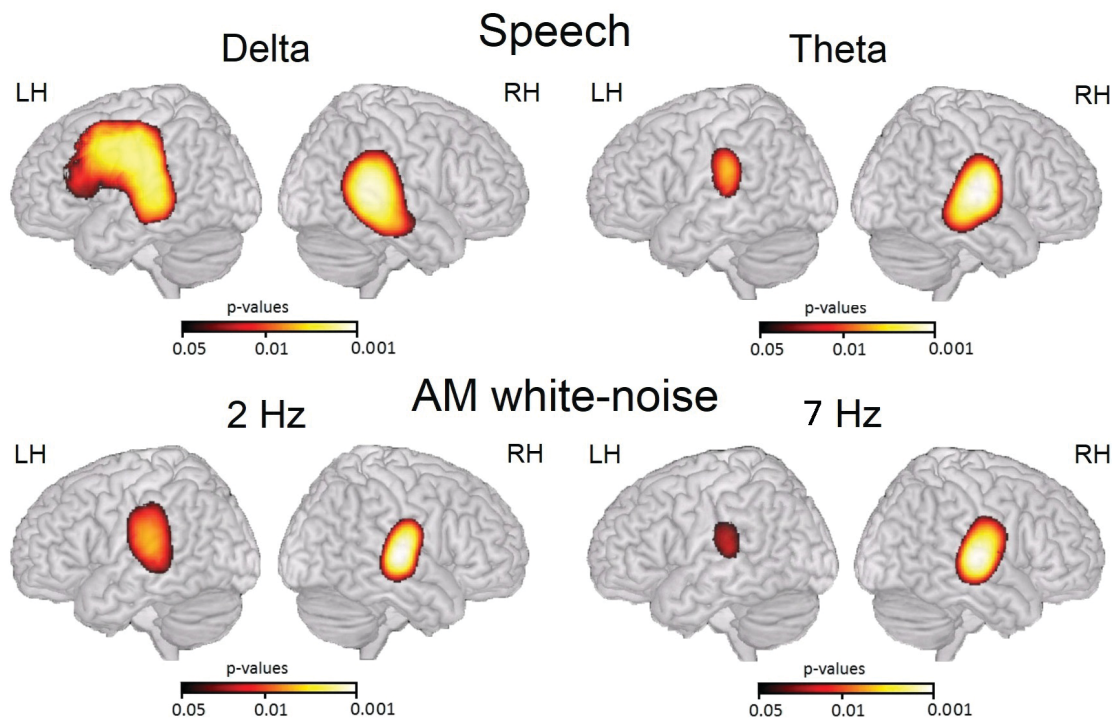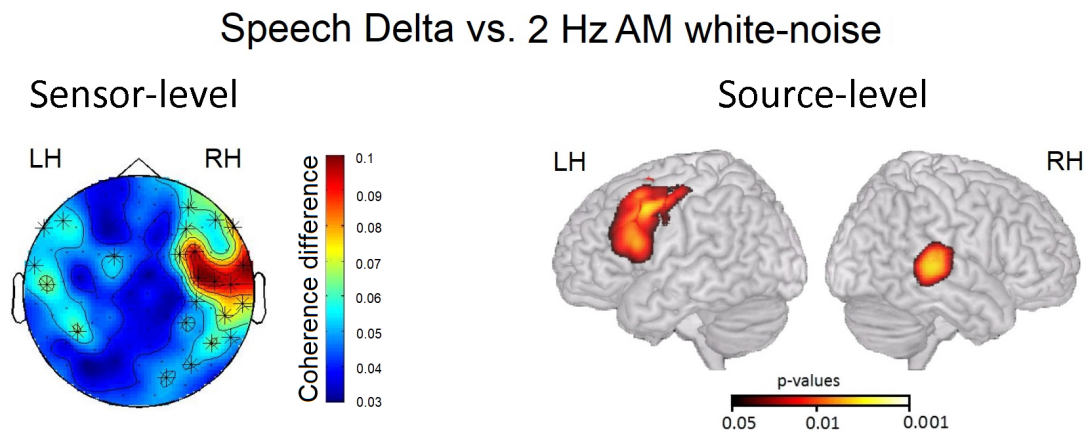
Figure 1

Figure 2

Figure 3

Speech Delta vs. 2 Hz AM white-noise

Sensor-level

Source-level

Figure 4

Figure 5

Figure 6

Speech Delta vs. Rotated Speech Delta

Sensor-level   Source-level