NeuroImage 141 (2016) 1–9

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

A 7T fMRI study investigating the influence of oscillatory phase on syllable representations

S. Ten Oever [a,*], L. Hausfeld [a], J.M. Correia [a], N. Van Atteveldt [a,b], E. Formisano [a,c], &, A.T. Sack [a]

[a] Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands
[b] Department of Educational Neuroscience, Faculty of Psychology and Education and Institute Learn, VU University Amsterdam, The Netherlands
[c] Maastricht Center for Systems Biology (MaCSBio), Maastricht University, Maastricht, The Netherlands

ARTICLE INFO

Article history:
Received 15 April 2016
Accepted 6 July 2016
Available online 7 July 2016

Keywords:
Language
Oscillations
Phase
fMRI
Temporal statistics

ABSTRACT

Stimulus categorization is influenced by oscillations in the brain. For example, we have shown that ongoing oscillatory phase biases identification of an ambiguous syllable that can either be perceived as /da/ or /ga/. This suggests that phase is a cue for the brain to determine syllable identity and this cue could be an element of the representation of these syllables. If so, brain activation patterns for /da/ should be more unique when the syllable is presented at the /da/ biasing (i.e. its "preferred") phase. To test this hypothesis we presented non-ambiguous /da/ and /ga/ syllables at either their preferred or non-preferred phase (using sensory entrainment) while measuring 7T fMRI. Using multivariate pattern analysis in auditory regions we show that syllable decoding performance is higher when syllables are presented at their preferred compared to their non-preferred phase. These results suggest that phase information increases the distinctiveness of /da/ and /ga/ brain activation patterns.

Introduction

Neuronal oscillations reflect subthreshold fluctuations in membrane potentials of neuronal ensembles. During one oscillatory period neuronal ensembles move closer and further away from the threshold for an action potential and so does the excitability level of these neurons. The exact point (aka phase) of the oscillation is therefore a reflection of the excitability level of the neuron. The role of oscillatory phases for perception and cognition is becoming increasingly clear (Fell and Axmacher, 2011; Kayser et al., 2009). While many studies have focused on the role of phase for stimulus detection [e.g. (Fiebelkorn et al., 2013; Henry et al., 2014; Ten Oever et al., 2015)], oscillatory phase also influences the categorization of stimuli (Ten Oever and Sack, 2015; Watrous et al., 2015a).

It has been proposed that oscillatory phase can also influence speech perception (Giraud and Poeppel, 2012; Schwartze and Kotz, 2015; Schroeder et al., 2008). Specifically, oscillatory phase can bias categorical perception of syllables when these syllables differ in their temporal properties (Peelle and Davis, 2012; Ten Oever and Sack, 2015). For example, /da/ and /ga/ have varying delays between the speech articulation and the speech sound (Ten Oever et al., 2013). Ongoing oscillatory phase as measured with electroencephalography (EEG) from central electrodes determines whether an ambiguous syllable that can either be perceived as /da/ or /ga/ is identified as one or the other syllable (Ten Oever and Sack, 2015). In the same study, we entrained oscillatory patterns in the brain to rhythmically presented sounds after which the same ambiguous syllable was presented at different onset delays (Fig. 1A). We found that depending on the delay between the target syllable and the offset of the entrainment (and thus the underlying phase) participants more likely identified the syllable as /da/ or /ga/ (Fig. 1B). This suggests that oscillatory phase is a cue for syllable identification and each syllable has one "preferred" phase.

Patterns of activation in the (auditory) cortex have been shown to reflect distributed representations of speech (Formisano et al., 2008; Mesgarani and Chang, 2012; Staeren et al., 2009; Tsunada and Cohen, 2014). These representations likely reflect the collective activation of numerous groups of neurons activated by different speech features present in the external speech input. Generally, the more distinct two different types of input are (e.g. by having multiple features that differentiate the inputs), the more distinct their activation patterns [or representations (Hausfeld et al., 2014)]. As oscillatory phase is a cue for syllable identification (see (Ten Oever and Sack, 2015) for theoretical background), it might also enhance the distinctiveness of the representation of a syllable.

Multi-variate pattern analysis (MVPA) in functional magnetic resonance imaging (fMRI) has been used successfully to discriminate between distributed speech representations (Formisano et al., 2008; Kilian-Hütten et al., 2011), and is more sensitive than classical univariate approaches to dissociate these distributed patterns of activation (Haxby et al., 2001; Haynes and Rees, 2006). We used this method to investigate whether discrimination performance between /da/ and /ga/ of this classifier would be

* Corresponding author.
E-mail address: sanne.tenoever@maastrichtuniversity.nl (S. Ten Oever).

http://dx.doi.org/10.1016/j.neuroimage.2016.07.011
1053-8119/

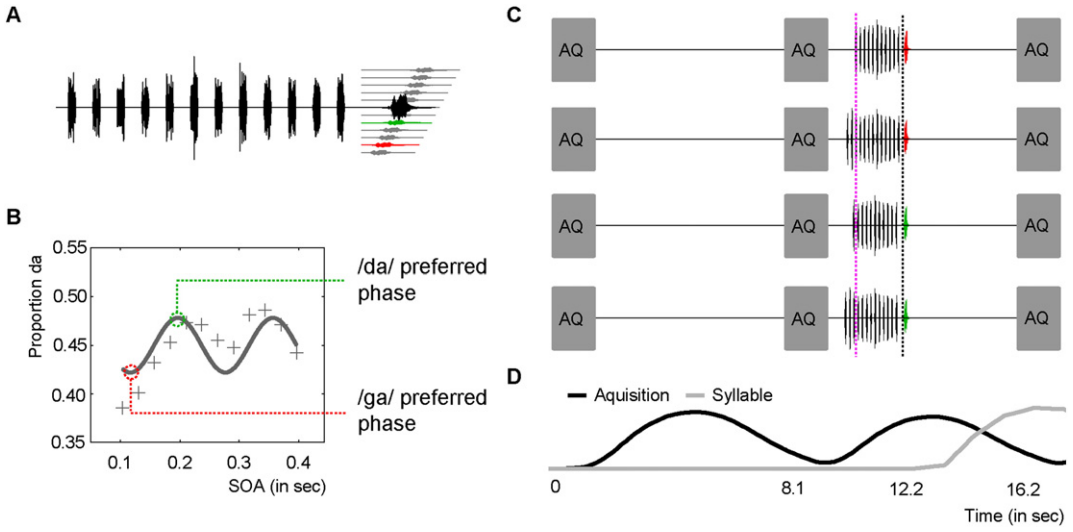

Fig. 1. Previous results and stimulation protocol. A) Entrainment stimulus after which a syllable is presented at different intervals. B) The results from the previous study [adapted with permission from Ten Oever and Sack (2015)]. The red and green SOAs represent the preferred phase for /ga/ and /da/ respectively. C) In the current fMRI design, the stimuli were always presented in the silent gap after image acquisition (AQ). Four different stimuli presentations are visualized in the figure; from top to bottom: at a stimulus onset asynchrony (SOA) of 120 ms with an entrainment train of n = 11, an SOA of 120 ms with n = 13, an SOA of 200 ms with n = 11, and at an SOA of 200 ms with n = 13. While the syllable types (red and green lines) are always presented at 12.2 (or 12.28) seconds after the first acquisition (see black dotted line) the entrainment trains (black) start at different time points dependent on condition (see pink dotted line). D) The predicted BOLD response to the acquisition noise (black) and the syllable (grey) is displayed. Due to the long TR the response to the acquisition noise is reduced while estimating the peak BOLD response of the syllable.

better when both syllables were presented at their preferred compared to non-preferred phase. This would support the notion that phase is a cue that increases the distinctiveness of the representation or activation patterns of /da/ and /ga/.

We induced brain oscillations by repeatedly presenting auditory stimuli at a 6.25 Hz rate (i.e. auditory entrainment), similar as in our previous study (Ten Oever and Sack, 2015). Non-ambiguous /da/ or /ga/ syllables were presented at differing delays after the entrainment finished, either corresponding to the syllable's preferred or non-preferred phase (see red and green lines in Fig. 1). Non-ambiguous stimuli were used to ensure that the classification discrimination does not reflect neuronal top-down processes based on stimulus identification (see (Kilian-Hütten et al., 2011)), but is a direct consequence of presenting the syllable at one specific phase. To ensure that the fMRI scanner noise did not influence the entrainment, we had a repetition time of the scanner of 8.1 s and presented the entrainment in the silent gap. With MVPA we calculated the accuracy of syllable identity discrimination from the resulting activation patterns. We found significantly better performance when both syllables were presented at their preferred phase compared to their non-preferred phase. These results show that syllable representations in auditory regions are optimally processed at a preferred oscillatory phase and indicate the potential of fMRI to map the spatial origin of these wide-spread oscillatory patterns.

Materials and methods

Participants

Ten healthy native Dutch speakers participated in the study (4 male, age range: 26–32, mean age: 29.1). One participant was left-handed. The study was approved by the local ethical committee at Maastricht University. Participants gave written informed consent prior to participation and filled out the safety screening from the Scannexus MRI facilities at Maastricht University. Participants received monetary compensation for participating. One participant was excluded from the analysis as the full fMRI session was not completed.

Stimuli and experimental procedures

In each trial first an entrainment sequence was presented, which consisted of band-passed noise-bursts (2.5 kHz–3.1 kHz, 50 ms) at a presentation rate of 6.25 Hz. The entrainment sequences were 11, 12, or 13 stimuli long to reduce temporal expectations of the arrival time of the syllable. After the train finished, the sound of either the syllable /da/ or /ga/ was presented. The original syllable used was a /da/ pronounced by a Dutch female speaker, lasting approximately 300 ms. This syllable was then morphed into a /ga/ by changing the third formant frequency from a mean frequency of 3.0 kHz to 2.6 kHz using Praat (Boersma and Weenink, 2013; Ten Oever and Sack, 2015). The syllables were presented after the entrainment sequence either 120 or 200 ms after the onset of the last noise burst (6 trials per run for each of these four conditions), which corresponded with the preferred phase of /ga/ and /da/ respectively (Ten Oever and Sack, 2015). In another condition the middle /da/-/ga/ morph was presented with reversed audio at either 120 or 200 ms (6 trials per run for each of these two conditions). In four additional trials per run (condition type randomly selected) the last stimulus in the noise sequence had a wider filtered broadband noise (2.2 kHz–3.4 kHz). In total there were 40 trials per run, 200 in total. Participants were required to press a button when they heard this stimulus sequence. These trials were not analyzed. All stimuli are online available at https://osf.io/n4mrf/.

Scanning parameters

MRI data was collected on a 7-tesla Siemens Magnetom scanner with a body gradient system with a whole brain coil at the Scannexus facilities, Maastricht, The Netherlands. Anatomical images were acquired via a T1 weighted MPRAGE sequence (TR = 3100 ms; TI = 1500 ms; TE = 2.25 ms; 0.6 mm isotropic) and a proton density (PD) weighted sequence with the same parameters (except the TR = 1440) not using the inversion module. This sequence was acquired to remove field inhomogeneities to improve image quality by dividing the T1 weighted image by the proton density weighted image (Van de Moortele et al., 2009). Five functional runs with 84

TRs were acquired for all participants. A blood oxygenation level-dependent (BOLD)-sensitive echo-imaging (EPI) sequence was used (matrix = 128 * 128; field of view = 192 * 192 mm^2; 66 slices; TR = 8100 ms; TE = 19 ms; acquisition time = 1.4 s resulting in a voxel size of 1.5 * 1.5 * 1.5 mm^3) with a GRAPPA acceleration factor of 2 (Griswold et al., 2002). Moreover, two slices were acquired simultaneously via an interleaved multiband sequence to improve the speed of acquisition (Moeller et al., 2010). To correct for the direction of acquisition 2 EPI sequences of 5 TRs were collected using both the anterior-to-posterior and posterior-to-anterior direction [main functional runs were acquired using the anterior-to-posterior acquisition direction (Andersson et al., 2003; Bowtell et al., 1994; Jezzard and Balaban, 1995)].

EPI sequences inherently generate loud acoustic noise, thereby challenging auditory research in the scanner. Even more troubling for the current paradigm is that the EPI sequence contains a strong rhythmic component as the separate images are acquired. To ensure that entrainment only occurs to our presented stream and not to the scanner noise we used a sparse sampling paradigm with a repetition time of 8100 ms. In this way we could position our stimuli in between two acquisitions such that our stimulus of interest (the syllable) would be presented 4 or 3.92 s before and 4.1 or 4.02 s after the start of the image acquisition, thereby collecting the data around the peak of the BOLD response while the signal related to scanner noise decreases (Fig. 1C and D). At the following acquisition interval no stimuli were presented to ensure that the signal would recover the baseline. As the syllable positioning was fixed at either 4000 or 3920 ms prior to acquisition, the onset of the entrainment train was slightly different, depending on the specific condition.

Data preprocessing

Data preprocessing was performed with BrainVoyager QX 2.8 (Brain Innovation, Maastricht, The Netherlands) and FSL5.0 (www.fmrib.ox.ac.uk). For anatomical data, the reconstructed MPRAGE T1 weighted images were divided by the images by the PD images to reduce inhomogeneities of the signal (Van de Moortele et al., 2009). An additional inhomogeneity correction was performed in BrainVoyager and the images were resampled to 0.5 mm isovoxel resolution and rotated to ACPC space. Then we performed automatic grey-white matter segmentation in FSL and manually adjusted the segmentation in BrainVoyager. A grey matter cortical mask of the temporal lobe was created to reduce the amount of voxels present in the multivariate pattern classification.

Functional images were motion corrected and temporal high-passed filtered using three cosine cycles and a linear trend regressor for each run separately. Slice acquisition timing was corrected with a sinc-weighted interpolation. In FSL we used the TOPUP function to correct for susceptibility induced distortions caused by the acquisition direction to improve alignment with the anatomical data. Then images were co-registered with the anatomical data. All following analyses are performed after these pre-processing steps.

Data analysis

Univariate analysis

Due to our long TR we only had 2 data points to model the BOLD response. Therefore, we estimated the activation patterns for each stimulus by calculating the proportion of signal change subtracting the activity of a single data point directly after the stimulus from the data point before the stimulus (baseline) and further dividing by this baseline. We repeated this calculation for all the stimuli, providing us with the features used for the MVPA analysis. To obtain a group map of activation we performed cortex based alignment of the surface maps (Goebel et al., 2006) and a random effect GLM using a step function as predictor for each sound condition and run (with conditions /da/ time point 120, /da/ time point 200, /ga/ time point 120, /ga/ time point 200, reverse, and control). The map was corrected for multiple comparisons using false-discovery rate (FDR) over the whole brain with a q of 0.05 (Benjamini and Yekutieli, 2001; Genovese et al., 2002). The unthresholded statistical map is available at https://osf.io/n4mrf/.

MVPA analysis

All MVPA analyses were performed in ACPC space. Support vector machines (SVM) were used to decode the multivariate activation patterns. As a first analysis we tested whether we could reliably decode syllable identity irrespective on which time point on the entrainment the syllables were presented. Training data consisted of randomly picking 96 out of the 120 trials (48 per syllable); the remaining trials were used for testing. Features consisted of the proportion of signal change as described above per voxel and per trial. Standardized z-scores were calculated over each run. Excessive amounts of features can harm classification performances and it is therefore important having an appropriate amount of features as input to the classifier (Norman et al., 2006). Therefore we repeated the classification using between 50 and 2500 most active voxels (overall activation of the single subject GLM collapsed over all sounds restricted to temporal areas) in 15 logarithmically spaced steps and extracted the best classification (Kilian-Hütten et al., 2011). Voxels with activation patterns stronger than 5 standard deviations of the mean were never included as this high activity pattern likely does not arise from neuronal activity, but more likely from bigger veins [see e.g. (Lee et al., 1995; Turner, 2002)]. This procedure (i.e. feature extraction, feature selection and classification) was repeated in ten cross-validations. To investigate whether the classification performance was above chance level we performed permutation tests. Syllable labels of both the training and testing were permuted 100 times and the exact same analysis was performed. Then we compared using a one-sided Wilcoxon signed rank test whether the original classification performance was higher than the average permuted labels for all participants. The same analysis was performed to classify the two time points irrespective of the identity of the syllable. Confidence intervals (CI) were determined by calculating bootstraps of all difference values (n = 10,000) and reporting the 5th percentile for one-sided tests and the 2.5th–97.5th percentile for two-sided tests.

Our main hypothesis was that decoding performance would increase when syllables are presented at their "preferred" phase. Therefore, we split the dataset in two parts to perform separate classifications: 1) syllables presented at their "preferred" (/ga/ at 120 ms and /da/ at 200 ms) and 2) syllables presented at their "non-preferred" phase (/da/ at 120 ms and /ga/ at 200 ms). The rest of the analysis was the same as above except that model training (testing) was based on 48 (12) trials. Additionally, to investigate whether one specific time point/phase would have a higher classification performance we repeated the analysis, but performed the /da/-/ga/ classification when both syllables were presented at 120 ms or when both were presented at 200 ms. Moreover, we investigated whether the time point classification was significantly different from the preferred phase classification. Since the time point classification has twice as many trials we performed permutation test by repeating the analysis 100 times and selected the same amount of trials as in the preferred-phase classification for each of the four conditions (i.e. in total 60 trials per permutation; 48 and 12 for training and testing, respectively; trials were pseudo-randomly selected). The average of all these classifications was compared to the preferred phase classification.

Finally, we wanted to perform a generalization analysis in which we trained the classifier to dissociate /da/ and /ga/ (independent of time point) and tested whether the classifier would identify the ambiguous syllable with reversed audio as a /da/ when it was presented at 120 ms and as /ga/ when it was presented at 200 ms. For this analysis we performed 6 cross-validations in which for each cross-validation 10 stimuli per condition were left out randomly for the training. This left us with 100 stimuli for the training and 60 stimuli for the testing per cross validation.

To investigate the spatial consistency of the voxels used for the classification we created group discriminative maps for the /da/-/ga/ classification when syllables were presented at their "preferred" phase.

These maps represent the shared cortical locations that contributed to the discrimination of the syllables. We created these maps in two different ways. First, we created a map using all the voxels that went into the final best classification level. This map shows the overlap over participants of all the voxels that were used for classification; however it does not dissociate which voxels contributed more to the discrimination. Moreover, the size of voxel sets varies over participants. Therefore, we created a second map in which only the 150 most discriminating voxels within the final best classification level were included. 150 voxels were chosen as it corresponded with the amount of voxels used for the classification of the participant with the least voxels in the feature selection. All maps were then transformed to the surface representation of one participant after cortex based alignment (Goebel et al., 2006).

Results

Univariate analysis

A random effect GLM of syllable presentation versus baseline showed bilateral regions of activation mainly in primary auditory and auditory association cortex. Additionally, parts of the right cingulate motor areas were active (Fig. 2). This indicates that our sparse sampling method was successful in eliciting reliable brain responses to spoken syllables. Other areas known to be activated by syllables or phonemes (Desai et al., 2008; Hickok and Poeppel, 2007; Liebenthal et al., 2005), such as inferior frontal cortex and insula, showed responses when using a more liberal threshold. Any direct contrast between /da/ or /ga/ or between the two time points did not result in any significant difference [as estimated using false discovery rate (Benjamini and Yekutieli, 2001)].

Multivariate analysis

/da/ versus /ga/

In a first step we wanted to replicate the finding that syllable identity can be decoded from fMRI BOLD patterns (Formisano et al., 2008). Moreover, this finding would show that our paradigm of sparse sampling can be successfully used to perform classification. Therefore, we trained a support vector machine classifier (see Materials and methods for details) to differentiate between /da/ and /ga/. We found a mean classification accuracy of 0.578 (Fig. 3, left panel). Statistical testing of this performance by permuting the labels of the syllables indicated that the accuracies of the original labels was higher than the empirical chance level ($Z = 1.836$; $p = 0.038$; 5th percentile $= 0.0087$). Note that the empirical chance level for classification is higher than 0.5 as we optimized the amount of voxels selected for the classification by using the classification with the best performance. However, permutation testing controls for this enhanced empirical chance level (Moeller et al., 2010).

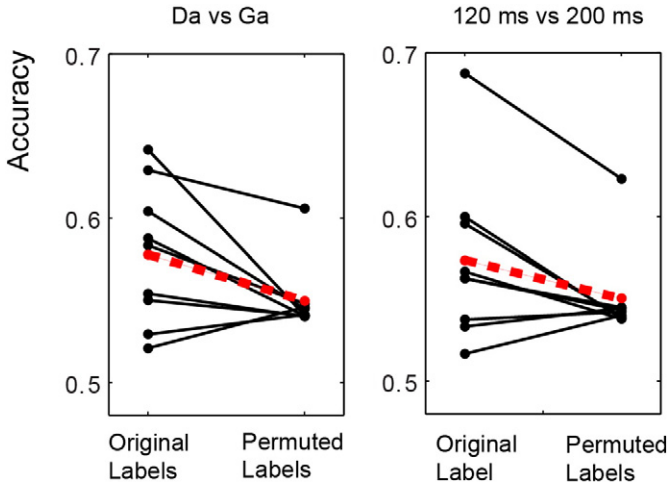

Fig. 3. Classification performance. Classification performance for each participant (black line) and the average (red dotted line) for the both original labels and permuted labels for the contrast /da/ versus /ga/ (left panel) and 120 versus 200 ms SOA (right panel).

Time point comparison

In a second step we classified the two time points (120 vs 200 ms) irrespective of the syllable identity. Overall classification performance was 0.574 which was higher compared to the empirical chance level ($Z = 1.659$; $p = 0.049$; 5th percentile $= 0.0063$). When we repeated the same analysis separately for the /da/ or /ga/ trials, we did not find any significant effect ($Z = -0.474$; $p = 0.682$ and $Z = 0.237$; $p = 0.406$ for /da/ and /ga/ respectively). The lower classification accuracy is likely due to the reduction in the amount of trials used in the analysis.

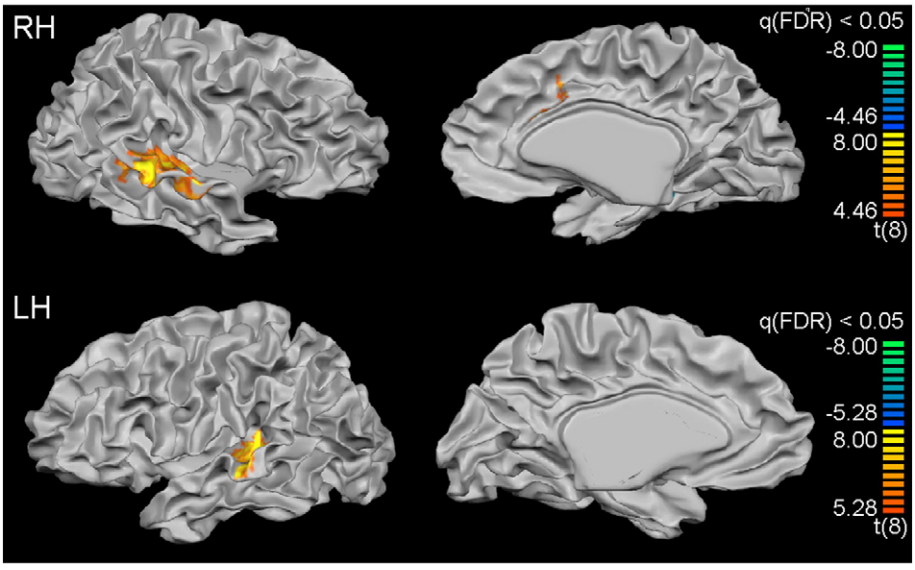

Fig. 2. Univariate results. Overall activation map of all presented syllables versus baseline as measured with a random-effect GLM. Results are presented on the surface of the brain of one participant after cortex-based alignment to this brain.

/da/ versus /ga/ at the preferred versus non-preferred phase

Our main expectation was that classification performance should be higher when syllables are presented at their preferred phase. This analysis is orthogonal to the previous two analyses as each syllable is presented in equal amounts at both phases. We split the data in two and repeated the /da/ versus /ga/ classification either when the syllables were both presented at their preferred or at their non-preferred phase. We found a higher /da/-/ga/ classification performance (0.605) for the preferred compared to the non-preferred phase (0.559; Fig. 4A, top panel; $Z = 2.140$; $p = 0.016$; 5th percentile = 1.898). Moreover, we found that only for the preferred phase the classification performance was higher compared to the empirical chance level ($Z = 1.778$; $p = 0.038$; 5th percentile = 0.0135 and $Z = -0.355$; $p = 0.639$; 5th percentile = −0.0257 for preferred and non-preferred phase, respectively). The amount of voxels that went into the final preferred phase analysis was: 618, 1891, 50, 153, 1891, 50, 267, 2500, and 116 for all nine participants respectively. Finally, the classification for the preferred phase was significantly higher compared to the pure time point classification ($Z = -2.369$; $p = 0.009$; 5th percentile = 0.0232).

/da/ versus /ga/ at 120 or 200 ms

To test whether there was one specific phase that increased classification accuracy, we repeated the previous analysis, but testing /da/-/ga/ classification when both syllables were presented at 120 ms or both presented at 200 ms. There was a classification performance of 0.578 and 0.550 when both syllables were presented at 120 ms or 200 ms, respectively. The classification performance did not significantly differ ($Z = 0.653$; $p = 0.514$; 95% CI = −0.0176–0.0889; two-sided). Additionally, both classifications were not significantly different from the empirical chance level ($Z = 0.71$; $p = 0.237$; 5th percentile = −0.0166 and $Z = 0.355$; $p = 0.361$; 5th percentile = −0.0308 for 120 and 200 ms, respectively).

Finally, we wanted to calculate the interaction between the factors /ga/ phase and /da/ phase. To do so with a non-parametric test we performed a signed rank test between two difference scores: 1) the difference in /da/-/ga/ classification performance between the preferred and non-preferred phase and 2) the difference in /da/-/ga/ classification performance between 120 and 200 ms. Initially, there was no significant effect ($Z = 0.355$; $p = 0.143$; 5th percentile = −0.0111). However, one participant had an extreme value in the 120–200 ms classification difference that was more than three standard deviations from the average (see Fig. 4B. One participant has a difference of almost 0.25). When removing this participant, the difference was significant ($Z = 1.890$; $p = 0.027$; 5th percentile = 0.0136 see Fig. 4C).

Generalization of ambiguous /daga/ with reversed audio

The generalization of the ambiguous stimulus with reversed audio did not show any significant effect ($Z = 0$; $p = 0.500$; 5th percentile = −0.0178). The absence of an effect is likely due the alignment of the onsets of the audio of the reversed syllable and the regular syllable. While a regular syllable has a very strong initial response, the reversed audio of a syllable has slow onset and its main energy is at the end of the stimulus. Subsequently, the brain most likely has a different readout of these two sounds.

Discriminative maps

The spatial consistency of the voxels used for the classification was investigated by creating group discriminative maps for the /da/-/ga/ classification when syllables were presented at their "preferred" phase. Fig. 5A shows the overlap of all the voxels used for the final classification,

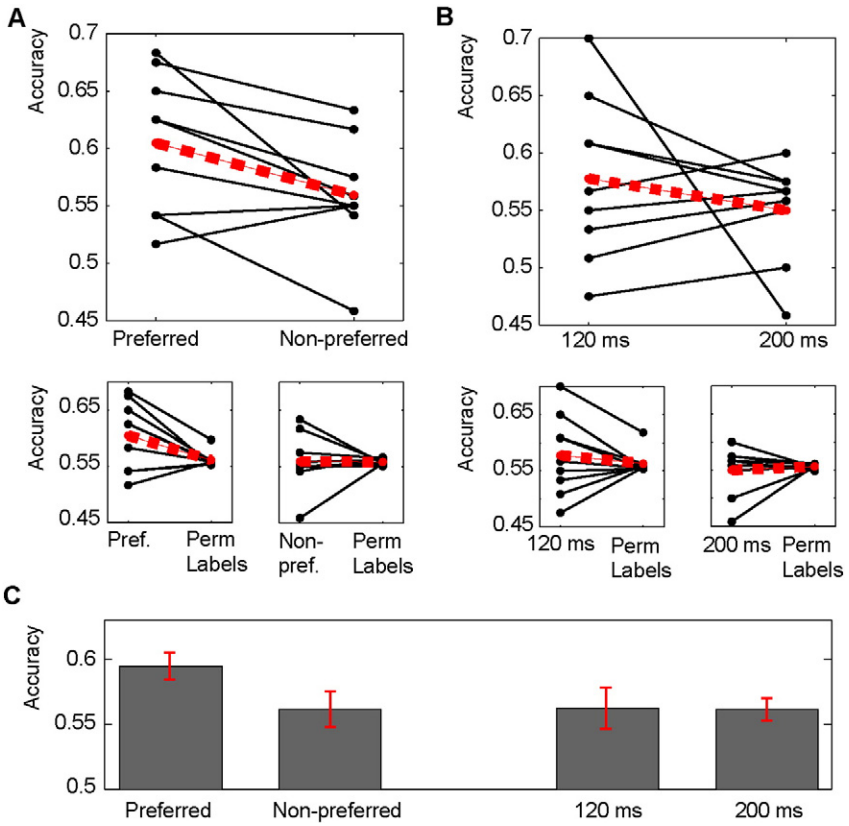

Fig. 4. Classification performance main contrasts. A) /da/-/ga/ classification performance for each participant (black line) and the average (red dotted line) for the preferred and non-preferred phase (top panel). Bottom two panels reflect the comparisons with the actual labels of the preferred phases (left) and non-preferred phases (right) and their respective permuted labels. B). /da/-/ga/ classification performance when syllables where presented at the early or late time point. C) The average classification performance excluding one outlier participant. Error bars represent the within subject standard error of the mean.

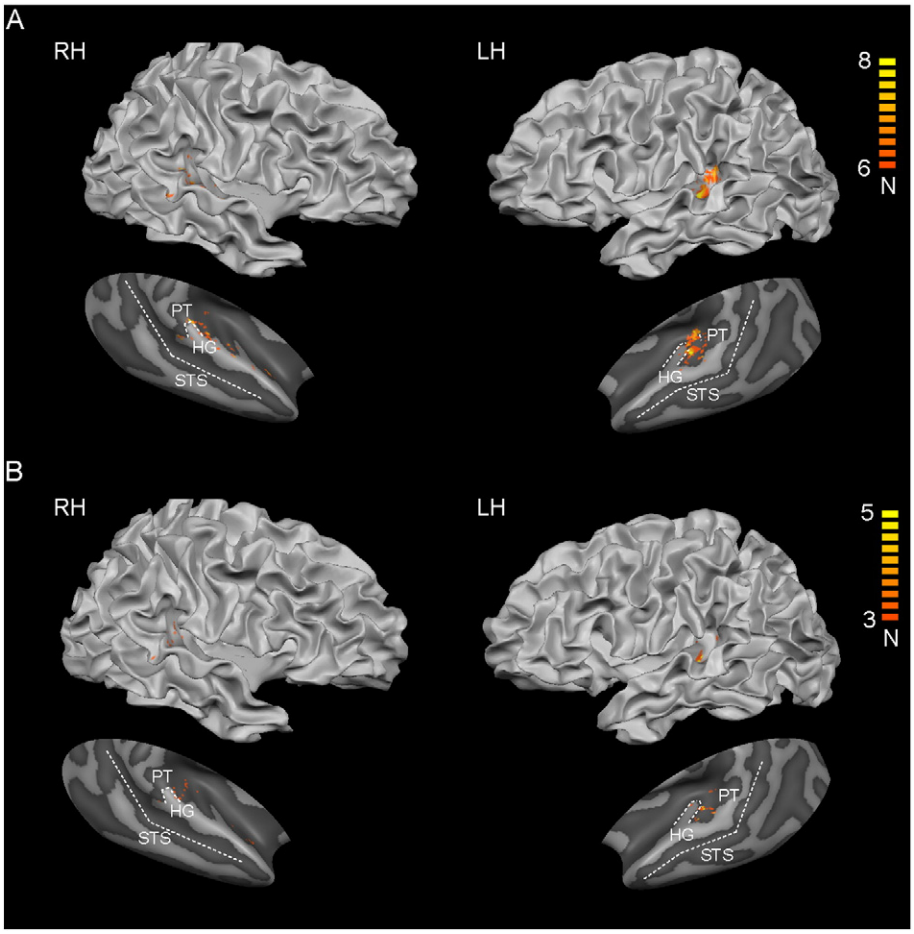

Fig. 5. Discriminative maps of the /da/ vs /ga/ at their "preferred" phase. A) The spatial overlap over participants of all voxels used in the optimized classification. Colour indicates the amount of participants. B) The spatial overlap over participants when only using the 150 most discriminative voxels. The overlap is highly reduced.

so the voxels having the highest percent signal change. The amount of voxels was dependent on the participant as it was individually tailored to optimize the classification. It is clear that only a small proportion of voxels had overlap for more than 6 out of 9 participants, mostly overlapping around the left Heschl's sulcus (HS) and in the right hemisphere anterior of Heschl's gyrus (HG) at the first transverse sulcus (FTS). Then, we investigated the overlap only using the 150 most discriminative voxels in the final classification (Fig. 5B). Voxels only showed overlap for 3 to 5 participants. Fig. 6 shows for each individual the 150 most discriminative voxels that resulted in the highest classification performance. Although most participants have their most discriminative voxels around the main auditory regions including bilateral superior temporal gyrus (STG), FTS, planum temporale (PT), HG, and HS, the exact distribution varied across participants.

Discussion

In the current study, we investigated whether oscillatory phase information changes the distinctiveness of neural syllable representations as measured with fMRI. This study was based on our previous results showing that syllable identification of an ambiguous stimulus (either perceived as /da/ or /ga/) is biased when it is presented at a specific phase of ongoing oscillation (Ten Oever and Sack, 2015; Ten Oever et al., 2013). We used the same sensory entrainment paradigm and presented /da/ and /ga/ stimuli either at their "preferred" or "non-preferred" phase to investigate whether phase information would change the fMRI activation patterns to these syllables. As we hypothesized, we found that /da/-/ga/ classification from these activation patterns (with MVPA) was more accurate when both syllables were presented at their preferred compared to non-preferred phase. These results support that syllable processing is phase-dependent and show that this information can be extracted even with slow fluctuating BOLD responses.

Phase dependent syllable processing

Phase coding has been proposed as a mechanism to represent information in the brain (Fries, 2005; Jensen et al., 2014; Watrous et al., 2015a, 2015b). Different electrophysiological studies have shown that adding phase information to classifiers aid classification performance (Kayser and Logothetis, 2009; Lopour et al., 2013). Moreover, neuronal populations coding for similar representations seem to communicate with each other by synchronizing their firing rates to a specific phase (Fries, 2005; Lisman and Jensen, 2013; O'Keefe and Recce, 1993). Since specific syllable representations prefer specific oscillatory phases (Ten Oever and Sack, 2015) neuronal populations coding for one syllable might become active when another syllable is presented at their preferred phase. This could be reflected in, for example, /da/ sensitive neurons being active if a /ga/ syllable is presented at a /da/ preferred phase. Alternatively, /da/ sensitive neuronal populations might have more robust processing for specific phases. In either way, syllable representations are more distinctive from each other when syllables are presented at their preferred phase. This effect was independent from top-down influences of the identification process as the used syllables that were not ambiguous.

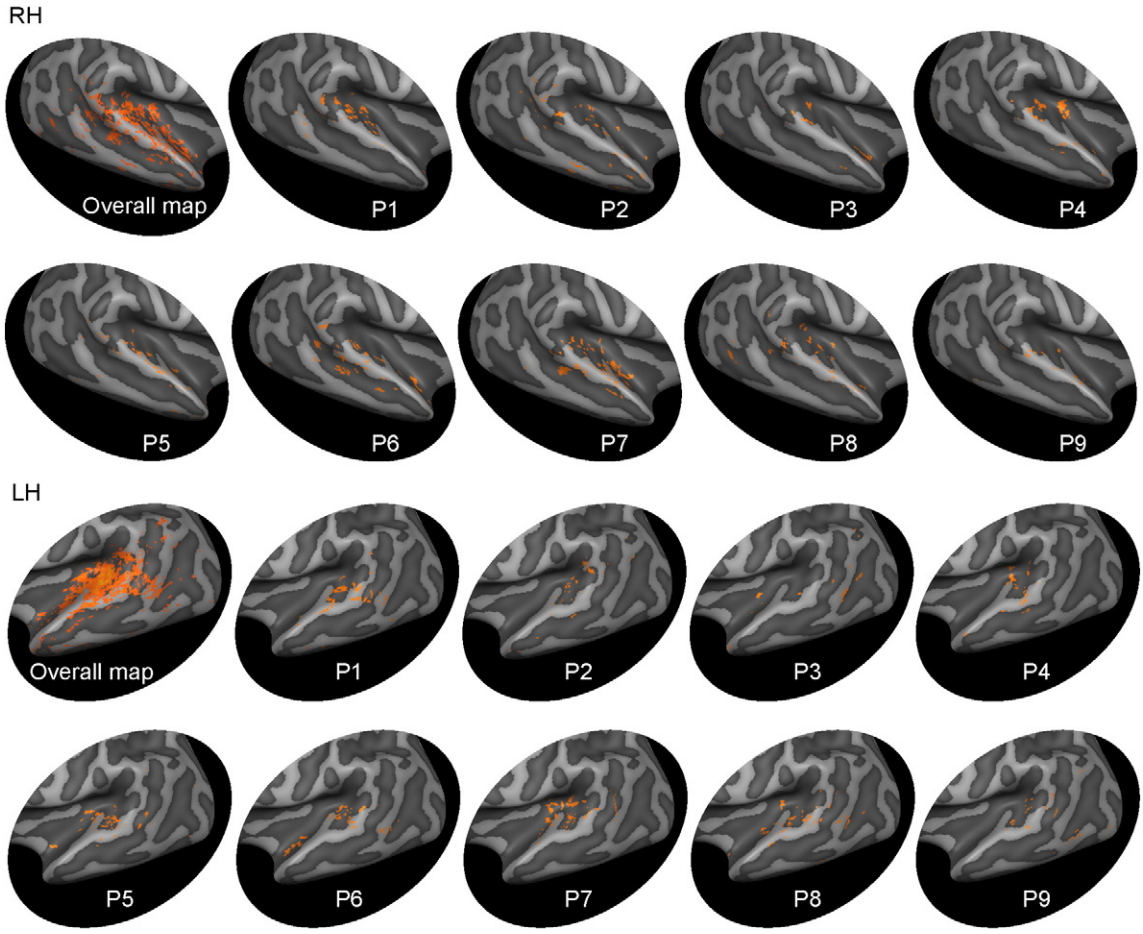

Fig. 6. Individual discriminative maps for the /da/ vs /ga/ classification at the "preferred" phase. The top 150 discriminative voxels for each participant are shown for the right and left hemisphere. The map labeled "overall map" shows all the top 150 discriminative voxels of all participants combined on one map. All voxels are plotted on the surface map of one representative participant after cortex based aligned.

Methodological considerations

The noisy scanner environment makes any type of auditory experiment difficult (Cho et al., 1997; Griswold et al., 2002). In the current design we chose to overcome this problem by having a very long repetition time and only trying to sample the peak of the BOLD response to the syllables. This has the drawback that the data points are limited. However, we were still able to find normal activation patterns to auditory stimuli and above chance level classification performance. The increased signal-to-noise ratio of 7T MRI might have helped to increase the overall activation levels. Moreover, the BOLD response to the scanner noise that normally accompanies the BOLD response to the auditory target stimuli might be significantly reduced (Bandettini et al., 1998; Talavage et al., 1999). This shows the added value of high field fMRI and the feasibility of this type of silent paradigms [see also (Amaro et al., 2002; Zaehle et al., 2007)]. For most experiments this type of sampling is not necessary, but if the rhythmic auditory patterns of the scanner noise are too intrusive for the specific experimental set-up, the proposed acquisition scheme represents one option to overcome this limitation of fMRI.

Spatial overlap

Our discriminative maps (Figs. 5 and 6) indicated a limited spatial overlap of the voxels used for the syllable classification. The only area that showed some overlap when using the 150 most discriminative voxels was left Helschl's sulcus bordering the Planum Temporale (PT). Left Helschl's sulcus is involved in the primary auditory analysis and largely part of the belt area (Moerel et al., 2014). It is normally sensitive to a broader tuning width of sounds (Hackett et al., 1998; Moerel et al., 2013; Rauschecker et al., 1995) and the most lateral part of the Helschl's sulcus also seems speech/voice sensitive (Belin et al., 2000; Moerel et al., 2014). In contrast, PT is viewed as a computational hub in which complex spectrotemporal inputs are matched to stored memories of auditory objects (Griffiths and Warren, 2002). It has been shown that this area plays an important role discriminating the perceived identity of an ambiguous syllable (Kilian-Hütten et al., 2011). In sum, it seems that the most discriminate areas in our study include areas that perform a higher order acoustic transformation linking the acoustic input to stored auditory categories (Obleser and Eisner, 2009). On a critical note, it could be that the high activation of these broadly tuned areas is partly induced by the broadband noise used in the entrainment. Moreover, individual discriminative maps are much more diverse and include more widespread areas, covering almost all auditory and auditory association areas.

Generalization of the effect

Previous studies have reported differential phase responses dependent on stimulus type (Kayser et al., 2009; Watrous et al., 2015b). Therefore, not only these reported syllables, but also other stimulus types might have a preferred oscillatory phase. Hence, it is relevant to investigate the contribution of phase information for stimulus identification. The phase information effect we report (a mean percentage classification difference of 4.5% between preferred and non-preferred

phase presentation) is comparable to other fMRI MVPA studies investigating for example attentional effects [significant changes from superior temporal cortex ranging from ±3–8% (Bonte et al., 2014)], expectancy effects (Kok et al., 2012 reported a ±3% classification change), or learning effects (a classification change of 6.5% was reported in Ley et al., 2012). The currently presented results, together with previously reported behavioral effect (Ten Oever and Sack, 2015), indicate that phase information might be an important determinant for the brain to categorize sensory information.

Conclusion

In this study, we showed that oscillatory phase contributes to the distinctiveness of the representation of /da/ and /ga/. These results add to a growing literature showing the role of oscillatory phase in perception and cognition (Cravo et al., 2011; Jensen et al., 2014; Lakatos et al., 2008; Peelle and Davis, 2012). Furthermore, it indicates that oscillatory properties in the brain might be an essential part of the representation of categorical information. FMRI has thus far not been used to investigate influences of oscillatory phase. We are one of the first to demonstrate that slow fluctuating BOLD patterns are indeed sensitive to this type temporal manipulation, if combined with a sophisticated stimulation design. This opens a new way to investigate phase, using the high spatial resolution that fMRI provides. This is important, as the phase coding mechanism that we demonstrate might be a unique strategy of the brain to memorize and organize perceptual input and future studies should aim to unravel the principles of this mechanism.

Acknowledgements

This study was supported by a grant from the Dutch Organization for Scientific Research (NWO; grant number 406-11-068) and funding from the Maastricht Brain Imaging Center. The authors declare no competing financial interests.

References

Amaro, E., Williams, S.C., Shergill, S.S., Fu, C.H., MacSweeney, M., Picchioni, M.M., ... McGuire, P.K., 2002. Acoustic noise and functional magnetic resonance imaging: current strategies and future prospects. J. Magn. Reson. Imaging 16 (5), 497–510.
Andersson, J.L., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. NeuroImage 20 (2), 870–888.
Bandettini, P.A., Jesmanowicz, A., Van Kylen, J., Birn, R.M., Hyde, J.S., 1998. Functional MRI of brain activation induced by scanner acoustic noise. Magn. Reson. Med. 39 (3), 410–416.
Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403 (6767), 309–312.
Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 1165–1188.
Boersma, P., Weenink, D., 2013. Praat: a system for doing phonetics by computer (Version 5.3.56). (Retrieved from) http://www.praat.org/.
Bonte, M., Hausfeld, L., Scharke, W., Valente, G., Formisano, E., 2014. Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. J. Neurosci. 34 (13), 4548–4557.
Bowtell, R., McIntyre, D., Commandre, M., Glover, P., Mansfield, P., 1994. Correction of geometric distortion in echo planar images. Paper Presented at the Soc. Magn. Res. Abstr.
Cho, Z., Park, S., Kim, J., Chung, S., Chung, S., Chung, J., ... Wong, E., 1997. Analysis of acoustic noise in MRI. Magn. Reson. Imaging 15 (7), 815–822.
Cravo, A.M., Rohenkohl, G., Wyart, V., Nobre, A.C., 2011. Endogenous modulation of low frequency oscillations by temporal expectations. J. Neurophysiol. 106 (6), 2964–2972.
Desai, R., Liebenthal, E., Waldron, E., Binder, J.R., 2008. Left posterior temporal regions are sensitive to auditory categorization. J. Cogn. Neurosci. 20 (7), 1174–1188.
Fell, J., Axmacher, N., 2011. The role of phase synchronization in memory processes. Nat. Rev. Neurosci. 12 (2), 105–118.
Fiebelkorn, I.C., Snyder, A.C., Mercier, M.R., Butler, J.S., Molholm, S., Foxe, J.J., 2013. Cortical cross-frequency coupling predicts perceptual outcomes. NeuroImage 69, 126–137.
Formisano, E., De Martino, F., Bonte, M., Goebel, R., 2008. "Who" is saying "what"? Brain-based decoding of human voice and speech. Science 322 (5903), 970–973.
Fries, P., 2005. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn. Sci. 9 (10), 474–480.
Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15 (4), 870–878.
Giraud, A.L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15 (4), 511–517.
Goebel, R., Esposito, F., Formisano, E., 2006. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. Hum. Brain Mapp. 27 (5), 392–401.
Griffiths, T.D., Warren, J.D., 2002. The planum temporale as a computational hub. Trends Neurosci. 25 (7), 348–353.
Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., ... Haase, A., 2002. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn. Reson. Med. 47 (6), 1202–1210.
Hackett, T., Stepniewska, I., Kaas, J., 1998. Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. J. Comp. Neurol. 394 (4), 475–495.
Hausfeld, L., Valente, G., Formisano, E., 2014. Multiclass fMRI data decoding and visualization using supervised self-organizing maps. NeuroImage 96, 54–66.
Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293 (5539), 2425–2430.
Haynes, J.-D., Rees, G., 2006. Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. 7 (7), 523–534.
Henry, M.J., Herrmann, B., Obleser, J., 2014. Entrained neural oscillations in multiple frequency bands comodulate behavior. Proc. Natl. Acad. Sci. 111 (41), 14935–14940.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8 (5), 393–402.
Jensen, O., Gips, B., Bergmann, T.O., Bonnefond, M., 2014. Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. Trends Neurosci. 37 (7), 357–369.
Jezzard, P., Balaban, R.S., 1995. Correction for geometric distortion in echo planar images from B0 field variations. Magn. Reson. Med. 34 (1), 65–73.
Kayser, C., Logothetis, N.K., 2009. Directed interactions between auditory and superior temporal cortices and their role in sensory integration. Front. Integr. Neurosci. 3.
Kayser, C., Montemurro, M.A., Logothetis, N.K., Panzeri, S., 2009. Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. Neuron 61 (4), 597–608.
Kilian-Hütten, N., Valente, G., Vroomen, J., Formisano, E., 2011. Auditory cortex encodes the perceptual interpretation of ambiguous sound. J. Neurosci. 31 (5), 1715–1720.
Kok, P., Jehee, J.F., de Lange, F.P., 2012. Less is more: expectation sharpens representations in the primary visual cortex. Neuron 75 (2), 265–270.
Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., Schroeder, C.E., 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. Science 320 (5872), 110–113.
Lee, A.T., Glover, G.H., Meyer, C.H., 1995. Discrimination of large venous vessels in time-course spiral blood-oxygen-level-dependent magnetic-resonance functional neuroimaging. Magn. Reson. Med. 33 (6), 745–754.
Ley, A., Vroomen, J., Hausfeld, L., Valente, G., De Weerd, P., Formisano, E., 2012. Learning of new sound categories shapes neural response patterns in human auditory cortex. J. Neurosci. 32 (38), 13273–13280.
Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., Medler, D.A., 2005. Neural substrates of phonemic perception. Cereb. Cortex 15 (10), 1621–1631.
Lisman, J.E., Jensen, O., 2013. The theta-gamma neural code. Neuron 77 (6), 1002–1016.
Lopour, B.A., Tavassoli, A., Fried, I., Ringach, D.L., 2013. Coding of information in the phase of local field potentials within human medial temporal lobe. Neuron 79 (3), 594–606.
Mesgarani, N., Chang, E.F., 2012. Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485 (7397), 233–236.
Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Uğurbil, K., 2010. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn. Reson. Med. 63 (5), 1144–1153.
Moerel, M., De Martino, F., Santoro, R., Ugurbil, K., Goebel, R., Yacoub, E., Formisano, E., 2013. Processing of natural sounds: characterization of multipeak spectral tuning in human auditory cortex. J. Neurosci. 33 (29), 11888–11898.
Moerel, M., De Martino, F., Formisano, E., 2014. An anatomical and functional topography of human auditory cortical areas. Front. Neurosci. 8.
Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. 10 (9), 424–430.
Obleser, J., Eisner, F., 2009. Pre-lexical abstraction of speech in the auditory cortex. Trends Cogn. Sci. 13 (1), 14–19.
O'Keefe, J., Recce, M.L., 1993. Phase relationship between hippocampal place units and the EEG theta rhythm. Hippocampus 3 (3), 317–330.
Peelle, J.E., Davis, M.H., 2012. Neural oscillations carry speech rhythm through to comprehension. Front. Psychol. 3.
Rauschecker, J.P., Tian, B., Hauser, M., 1995. Processing of complex sounds in the macaque nonprimary auditory cortex. Science 268 (5207), 111–114.
Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S., Puce, A., 2008. Neuronal oscillations and visual amplification of speech. Trends Cogn. Sci. 12 (3), 106–113.
Schwartze, M., Kotz, S.A., 2015. Contributions of cerebellar event-based temporal processing and preparatory function to speech perception. Brain Lang.
Staeren, N., Renvall, H., De Martino, F., Goebel, R., Formisano, E., 2009. Sound categories are represented as distributed patterns in the human auditory cortex. Curr. Biol. 19 (6), 498–502.
Talavage, T.M., Edmister, W.B., Ledden, P.J., Weisskoff, R.M., 1999. Quantitative assessment of auditory cortex responses induced by imager acoustic noise. Hum. Brain Mapp. 7 (2), 79–88.

Ten Oever, S., Sack, A.T., 2015. Oscillatory phase shapes syllable perception. Proc. Natl. Acad. Sci. 112 (52), 15833–15837.
Ten Oever, S., Sack, A.T., Wheat, K.L., Bien, N., Van Atteveldt, N., 2013. Audio-visual onset differences are used to determine syllable identity for ambiguous audio-visual stimulus pairs. Front. Psychol. 4.
Ten Oever, S., Van Atteveldt, N., Sack, A.T., 2015. Increased stimulus expectancy triggers low-frequency phase reset during restricted vigilance. J. Cogn. Neurosci. 27 (9), 1811–1822. http://dx.doi.org/10.1162/jocn_a_00820.
Tsunada, J., Cohen, Y.E., 2014. Neural mechanisms of auditory categorization: from across brain areas to within local microcircuits. Front. Neurosci. 8.
Turner, R., 2002. How much cortex can a vein drain? Downstream dilution of activation-related cerebral blood oxygenation changes. NeuroImage 16 (4), 1062–1067.
Van de Moortele, P.-F., Auerbach, E.J., Olman, C., Yacoub, E., Uğurbil, K., Moeller, S., 2009. T 1 weighted brain images at 7 tesla unbiased for proton density, T 2* contrast and RF coil receive B 1 sensitivity with simultaneous vessel visualization. NeuroImage 46 (2), 432–446.
Watrous, A.J., Fell, J., Ekstrom, A.D., Axmacher, N., 2015a. More than spikes: common oscillatory mechanisms for content specific neural representations during perception and memory. Curr. Opin. Neurobiol. 31, 33–39.
Watrous, A.J., Deuker, L., Fell, J., Axmacher, N., 2015b. Phase-amplitude coupling supports phase coding in human ECoG. eLife 4, e07886.
Zaehle, T., Schmidt, C.F., Meyer, M., Baumann, S., Baltes, C., Boesiger, P., Jancke, L., 2007. Comparison of "silent" clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks. NeuroImage 37 (4), 1195–1204.