Neural oscillations in human auditory cortex revealed by fast fMRI during auditory perception

Sascha Frühholz, Wiebke Trost, Didier Grandjean, Pascal Belin

NeuroImage 207 (2020) 116401

Contents lists available at ScienceDirect

NeuroImage

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

Neural oscillations in human auditory cortex revealed by fast fMRI during auditory perception

Sascha Frühholz [a,b,c,*], Wiebke Trost [a], Didier Grandjean [d,e], Pascal Belin [f]

[a] Department of Psychology, University of Zurich, Zurich, 8050, Switzerland
[b] Neuroscience Center Zurich, University of Zurich and ETH Zurich, Zurich, 8057, Switzerland
[c] Center for Integrative Human Physiology (ZIHP), University of Zurich, 8057, Switzerland
[d] Swiss Center for Affective Sciences, University of Geneva, 1202, Geneva, Switzerland
[e] Department of Psychology, University of Geneva, 1205, Geneva, Switzerland
[f] Institut de Neuroscience de la Timone, Aix-Marseille Université, 13005, Marseille, France

ARTICLE INFO

Keywords:
Voice
Auditory cortex
Oscillations
fMRI
Social signals

ABSTRACT

Previous work pointed to the neural and functional significance of infraslow neural oscillations below 1 Hz that can be detected and precisely located with fast functional magnetic resonance imaging (fMRI). While previous work demonstrated this significance for brain dynamics during very low-level sensory stimulation, we here provide the first evidence for the detectability and functional significance of infraslow oscillatory blood oxygenation level-dependent (BOLD) responses to auditory stimulation by the sociobiological relevant and more complex category of voices. Previous work pointed to a specific area of the mammalian auditory cortex (AC) that is sensitive to vocal signals as quantified by activation levels. Here we show, by using fast fMRI, that the human voice-sensitive AC prioritizes vocal signals not only in terms of activity level but also in terms of specific infraslow BOLD oscillations. We found unique sustained and transient oscillatory BOLD patterns in the AC for vocal signals. For transient oscillatory patterns, vocal signals showed faster peak oscillatory responses across all AC regions. Furthermore, we identified an exclusive sustained oscillatory component for vocal signals in the primary AC. Fast fMRI thus demonstrates the significance and richness of infraslow BOLD oscillations for neurocognitive mechanisms in social cognition as demonstrated here for the sociobiological relevance of voice processing.

1. Introduction

Oscillatory brain responses are a rich and differential measure of stimulus- and task-related neural activity as well as of neural activity during task-free resting-state brain recordings (Buzsáki and Draguhn, 2004). Oscillatory brain signals are usually quantified based on neural recordings with high temporal resolution in the millisecond range, such as electro- (EEG) and magnetoencephalography (MEG), or intracranial local-field potentials (LFP). These methods allow one to quantify brain signals from low to high oscillatory frequencies, but with the shortcoming of being either unable to precisely localize oscillatory brain signals (EEG, MEG) or of being spatially inflexible and constrained (LFP). A recent study (Lewis et al., 2016) reported the quantification of oscillatory blood oxygenation level-dependent (BOLD) responses resulting from the fast functional magnetic resonance imaging (fMRI) acquisition of brain data in the subsecond range.

Given the improved temporal resolution of fast fMRI protocols (Feinberg et al., 2010), BOLD oscillations can now be quantified with precise and relatively unconstrained spatial localization. However, this quantification is limited to the infraslow oscillatory range (<1 Hz). A previous report by Lewis and colleagues (Lewis et al., 2016) demonstrated the functional significance of these infraslow BOLD oscillations but was restricted to very low-level sensory stimulation in the form of visual flickering checkerboards. Specifically, the authors used the combined acquisition of EEG and fMRI brain signals elected by oscillatory visual stimuli at a specific frequency rate, and this oscillatory visual stimulation was directly represented by corresponding BOLD oscillations in primary visual cortex, and not by differences in the amplitude level of neural activity as measured by the EEG signal. Besides these distinct BOLD oscillations in the visual cortex to visual stimulation, there is also evidence from resting-state fMRI that there are significant endogenous BOLD oscillations even in the absence of external stimulation above 0.1

* Corresponding author. University of Zurich, Department of Psychology, Binzmuhlestrasse 14 (Box 18), 8050, Zürich, Switzerland.
E-mail address: sascha.fruehholz@uzh.ch (S. Frühholz).

https://doi.org/10.1016/j.neuroimage.2019.116401
Received 26 September 2019; Accepted 24 November 2019
Available online 27 November 2019

Hz (Bandettini and Bullmore, 2008; Chen and Glover, 2015). Thus, the human brain shows infraslow BOLD oscillations both in the resting-state as well as in response to low-level sensory stimulations. Additionally, these BOLD oscillations seem to have a significant meaning for sensory and cognitive processing.

However, little information is available on the significance of task-related and stimulation-induced infraslow BOLD oscillations during the neurocognitive processing of more complex social stimuli. Furthermore, compared to visual stimuli, both auditory signals and auditory processing evolve over time. Hence, a pending question is if BOLD oscillations, first, respond to auditory stimulation and, second, if BOLD oscillations show a dynamic pattern over the temporal course of auditory recognition. Motivated by these questions, we investigated BOLD oscillations in response to temporally extended vocal signal processing. Compared with the processing of simple sounds, sound features, and non-object-like sound patterns, vocal signals include a social dimension. Therefore, they trigger not only basic acoustic processing but also some evaluative processes regarding the social significance of the signal. For example, typical AC activity in response to vocal signals is associated with socially more relevant decisions within a voice/non-voice task compared to a simple acoustic task (Bestelmeyer et al., 2011). Also, vocal signals attract more attention than other sound objects due to their social nature (Lin et al., 2016).

Given that vocal signal recognition might be linked to some forms of social cognition, we therefore aimed at investigating the functional significance of infraslow BOLD oscillations during the processing of socially important vocal signals, which receive prioritized processing in subfields of the auditory cortex (AC) (Andics et al., 2014; Pascal Belin et al., 2000; Petkov et al., 2008). These AC subfields are mainly located in higher-level AC an area that is assumed to be highly sensitive to vocal signals. It also seems that compared to the left AC, the right AC shows an increased, but non-exclusive response and peak activity differentiation to vocal signals (P Belin et al., 2000; C R Pernet et al., 2015). Another aim of the present study is to investigate whether BOLD oscillations show a laterality difference and lateralized prioritization between the left and right AC in response to vocal and nonvocal sounds. With the term "prioritization" we refer to the ability of vocal sounds, compared to other sounds, to elicit enhanced and potentially faster, but not necessarily exclusive, brain activity using various measures of neural activity (Charest et al., 2009; Cyril R. Pernet et al., 2015). Across a broad variety of species, vocal signaling is an essential element of close and distant communication, and it is fundamental to any social interaction with both conspecifics (Belin et al., 2011; Frühholz and Belin, 2019) and heterospecifics (Andics et al., 2014; Andics and Farago, 2019). Across many species, voice processing therefore may be more important compared to the perception of other types of sounds and thus might receive prioritized neural processing (Frühholz and Belin, 2019).

The description of this voice-sensitive auditory cortex (AC) (termed the "temporal voice area", TVA (Pascal Belin et al., 2000)) has until recently been limited to the level of neural activity when stimulated with vocal compared with other non-vocal sounds. The area of increased neural activity usually extends over the bilateral cortical regions of primary (i.e. Heschl's gyrus (HG)), secondary (i.e. planum polare, planum temporale (PTe)), and higher-level auditory regions in the superior temporal cortex (STC, composed of the superior temporal gyrus (STG) and sulcus (STS)). Although the discovery of the TVA in terms of higher-level neural activity for vocal sounds remains an important and influential finding (Pascal Belin et al., 2000; Cyril R. Pernet et al., 2015), a more detailed functional description has remained relatively scarce given the limits of traditional neuroimaging techniques in humans.

Here we took advantage of fast neuroimaging techniques (Feinberg et al., 2010) in humans to determine novel infraslow BOLD oscillatory measures of brain responses during stimulation-related human voice processing. These measures enable investigators to precisely link the temporal aspects of neural and mental processing, especially at the level of the neural auditory system (Seifritz et al., 2002) together with precise spatial localization of these oscillatory signals. Concerning the auditory system, previous work suggests that there are a few different response modes within the AC's response to sound (Bathellier et al., 2012). In humans, a major distinction of the raw BOLD signal suggests both a transient and a more sustained response mode in the AC (Seifritz et al., 2002). A study by Seifritz and colleagues (Seifritz et al., 2002) specifically demonstrated that the AC showed different response modes to extended auditory stimulation. But they only investigated these response modes using the scanner noise as the auditory stimulation. Here we aimed to also quantify AC response modes to extended auditory stimulation but with vocal signals as socially relevant stimulus and by quantifying BOLD oscillations as a potentially richer indicator of neural activity compared to the raw BOLD signal.

Infraslow BOLD oscillations are usually below the commonly investigated oscillatory properties (i.e. in terms of the frequency level) of the AC in response to sound stimulation (Giraud and Poeppel, 2012), but they carry important information about the dynamics of sensory processing (Lewis et al., 2016) and about mental processes in response to sensory input (Buzsáki and Draguhn, 2004). Furthermore, BOLD oscillations are associated with neural activity in a widespread network of task-related brain systems beyond traditional brain locations that are endogenously active during the resting-state (Fox and Raichle, 2007). Thus, BOLD oscillations do not only resemble endogenous BOLD oscillations in the resting brain and are not restricted to the default brain network, but can be directly detected in brain areas that are responsible for sensory and task-related cognitive processing.

This indicates the neurobiological significance of BOLD oscillations as a detectable marker of normal cognitive processing in BOLD brain signals as well as of abnormal processing in neuropsychiatric populations (Greicius, 2009). Infraslow BOLD oscillations are periodic high-amplitude brain signals, which, unlike random neural signal fluctuations (Lewis et al., 2016), are driven by underlying oscillations in neural activity at the level of nerve cells, especially in the grey matter tissue (Zuo et al., 2010). Infraslow BOLD oscillations are thus directly representing neural activity either as the low-frequency component of such neural oscillations (Lewis et al., 2016) that are similarly captured with traditional electrophysiological methods (He et al., 2008) or as power modulations (i.e. the envelope) of high-frequency neural oscillations (>1 Hz) (Pan et al., 2013). Furthermore, infraslow BOLD oscillations are also driven by neural activity in a brain network that is similar to that of high-frequency neural oscillations (Lewis et al., 2016), indicating that neural oscillations at very different frequency levels can originate from, and superimpose in, the same brain areas.

These previous reports together demonstrate and validate the neurobiological distinctiveness of infraslow BOLD oscillations that are meaningfully related to neural activity and that directly capture specific low-frequency properties of this neural activity (Mitra et al., 2018). Simple and complex auditory signals elicit neural oscillation across a wide range of frequency components (Giraud and Poeppel, 2012; Lakatos, 2005; O'Connell et al., 2015), so we accordingly aimed at using this possibility of infraslow BOLD oscillations to capture low-frequency neural oscillation to auditory vocal signals. As mentioned above, auditory signals have a generic temporal dimension, so instead of only quantifying BOLD oscillations on the mean level of their oscillatory power by using simple FFT analyses, we aimed at determining the dynamic nature of BOLD oscillations in their relation to the temporal dynamic vocal signals. Thus, beyond using traditional measures of neural activity (i.e. "static" BOLD activation maps), we specifically describe a novel and informative measure of BOLD activity that quantifies the neural oscillation of brain responses (i.e. "dynamic" BOLD oscillation maps) during the neural processing of complex auditory stimuli, such as vocal signals. BOLD oscillation measures combine the advantages of relatively precise localization of brain oscillatory activation and the ability to dynamically and temporally track neural activity even on slowly changing oscillatory components.

We hypothesized that the new measure of infraslow BOLD oscillations

would allow use to determine a difference between vocal and nonvocal sound processing beyond the traditional measures of BOLD activity differences. Similar to previous observation that vocal and nonvocal sounds generally elicit activity in similar subregions of the auditory cortex (Pascal Belin et al., 2000; Cyril R. Pernet et al., 2015), but with important activity level differences indicated by higher activity for vocal sounds, we also expected that vocal and nonvocal sounds might elicit a similar general pattern of BOLD oscillations, but with significant differences in the power and the temporal pattern of such oscillation patterns. These differences in BOLD oscillations might be an additional, more detailed, and dynamic neural measure of the prioritization of vocal sounds in the primate neural system. The present study thus had three major aims: first, we aimed at generally identifying the dynamic BOLD oscillatory pattern for a sound signal that evolve over time; second, we aimed at identifying the BOLD oscillatory differences to vocal compared to nonvocal sounds given that vocal signals are socially more relevant than other categories of sounds; and, third, we aimed at identifying BOLD oscillatory differences of the left and right AC as previous research indicated some lateralization differences in processing vocal and nonvocal sounds.

2. Materials and methods

2.1. Human participants

Experiments 1. Thirteen healthy volunteers (eight males; mean age 26.62 years, SD = 5.98, age range 21–41) took part in experiment 1. The participants had normal hearing abilities and normal or corrected-to-normal vision. No participant presented a neurological or psychiatric history. All participants gave informed and written consent for their participation in accordance with the ethical and data security guidelines of the University of Geneva. The experiments were approved by the ethics committee of the University of Geneva.

Experiment 2. Eighteen healthy volunteers (eight males; mean age 23.13 years, SD = 4.54, age range 18–33) took part in experiment 2. The participants had normal hearing abilities and normal or corrected-to-normal vision. No participant presented a neurological or psychiatric history. All participants gave informed and written consent for their participation in accordance with the ethical and data security guidelines of the University of Geneva. The experiment was approved by the ethics committee of the University of Geneva. Data is available on reasonable request in an anonymous format.

2.2. Experimental settings

Experiment 1. We used both an event-related and block design to define the consistent cortical area of the TVA that is common to any common functional TVA localizer scan. This double procedure to define the TVA was to get a firm and consistent definition of the voice-sensitive auditory cortex that responds both to short vocal events and to vocal utterances of longer duration. First, since we were interested in both sustained and transient BOLD oscillations, we also aimed to define the TVA based on short (transient) and longer vocal and nonvocal stimulation (sustained). Second, previous studies on the TVA used either short (Dricu et al., 2017; Fruhholz et al., 2016a) or extended vocal and nonvocal stimuli (Pascal Belin et al., 2000; Fruhholz et al., 2016b; C R Pernet et al., 2015), which led to common but also differential neural activations. Block-design voice localizer scans seem to elicit stronger activity in higher-level auditory cortex, while event-related design elicit more local peaks in the primary and secondary AC.

For the event-related TVA localizer run, the stimulus material consisted of 500 ms sound clips, 70 human speech and non-speech vocalizations (vocal sounds) and 70 non-human vocalizations and sounds (non-vocal sounds: animal vocalizations, artificial sounds, natural sounds). The 500 ms clips were sound snippets from the set of 20s sound clips used in the block-design run (see below). The same stimuli were used in a previous study (Capilla et al., 2013). Each sound clip was preceded by a 500 ms fixation cross and followed by a jittered blank 3550–5000 ms gap before the onset of the next stimulus.

For the block-design TVA localizer run, the stimulus material was identical to the event-related design but instead consisted of 20s sound clips including 21 human speech and non-speech vocalizations (vocal sounds) and of 21 non-human vocalizations and sounds (non-vocal sounds: i.e. animal vocalizations, artificial sounds, natural sounds). The same stimuli were used as those reported in the earlier original paper on voice-sensitive regions in the auditory cortex (Pascal Belin et al., 2000). Each sound clip was preceded by a 10s blank period without auditory stimulation. The experiment also ended with a blank period of 10s with no auditory stimulation.

All stimuli were normalized to have the same mean RMS and were presented at 70 dB SPL during the experiment. Each sound clip was presented one time during the event-related and the block-design, and presentation was randomized for each participant.

Experiment 2. This was a standard resting-state functional brain acquisition scan of about 7min duration and with a total of 650 vol acquired. Participants were asked to remain still in the scanner without thinking about anything specific and to keep their eyes open. The rationale for including a resting-state fMRI experiment was based on the observation that even the resting brain shows some endogenous BOLD oscillations (Bandettini and Bullmore, 2008). To separate task-related BOLD oscillations from endogenous BOLD oscillations, we also needed to obtain resting-state brain data.

2.3. Image acquisition

For all experiments, we recorded functional imaging data on a 3T SIEMENS Tim Trio System (Siemens, Erlangen, Germany), using a T2*-weighted gradient multiband echo-planar imaging (M-EPI) pulse sequence with an acceleration factor of four, 3-mm isotropic resolution, 36 slices in a 64 × 64 matrix, 20% distance factor, TR/TE = 650/30 ms, and FA = 50°. Structural images had 1-mm isotropic resolution (192 contiguous 1-mm slices, TR/TE/TI = 1900/2.27/900 ms, FoV = 296 mm, in-plane resolution of 1 × 1mm). Additional physiological data (heart rate and respiration) were recorded with the MP150 BIOPAC acquisition system (Santa Barbara, CA).

2.4. Basic data analysis

Data preprocessing. Preprocessing and statistical analyses of functional images were performed with Statistical Parametric Mapping software (SPM12, Welcome Department of Cognitive Neurology, London (fil.ion.ucl.ac.uk/spm/); Matlab (mathworks.com/products/matlab.html). Functional data were first manually realigned to the AC-PC axis, followed by motion correction of the functional images. Each participant's structural image was co-registered to the mean functional image and then segmented to allow estimation of normalization parameters. Using the resulting parameters, we spatially normalized the anatomical and functional images to the Montreal Neurological Institute (MNI) stereotactic space. The functional images for the main experiment were resampled into 2 mm^3 voxels. All functional images were spatially smoothed with an 8 mm full-width half-maximum isotropic Gaussian kernel.

Single-subject and group analysis. For the first-level analysis of experiment 1 using the event-related design, we used a general linear model (GLM) and all trials were modeled with a stick function aligned to the onset of each stimulus, which was then convolved with a standard hemodynamic response function (HRF). We accounted for serial correlations in the fMRI time series by using the "FAST" option implemented in SPM12. Contrast images comparing vocal with non-vocal stimuli (and vice versa) were then taken to separate random-effects group-level analyses in order to determine voice-sensitive regions (and potentially regions that are sensitive to non-vocal sounds) in both hemispheres of the auditory cortex. This analysis resulted in the first type of cortical maps, which we refer to as "activation maps".

For the first-level analysis of experiment 1 including the block-design TVA run, all trials were modeled in a GLM with a boxcar function aligned to the onset of each stimulus, including a 20s duration, which was then convolved with a standard HRF. Contrast images were analyzed the same way as for the event-related design. The common area of the TVA activity for the event-related and the block design run was defined as voxels commonly active in both runs.

For the first-level analysis of experiment 2, we used a GLM in which we created two regressors that randomly resembled the block design as described for experiment 1. Since the total acquisition time in experiment 2 was shorter than in experiment 1, we could model only eight vocal and eight non-vocal trials in each participant. Taking the data from this resting-state brain scan we virtually defined vocal and non-vocal events in these data by randomly selecting trial onset times from participants from experiment 1 (block-design) for 8 vocal and 8 non-vocal events. The order of "vocal" and "non-vocal" trials was completely random for each participant's data, but each trial was modeled with a 20s duration and a 10s blank period between trials. Thus, we randomly cut these resting-state brain data into 20s events and randomly assigned the label "vocal" or "non-vocal" trial., We performed the same analysis as for the real vocal and non-vocal trials of experiment 1 (block design) on these 20s segments of "virtual" vocal and non-vocal trials. Experiment 2 thus served as a baseline experiment to estimate brain oscillation in the resting brain that might be specifically due to physiological noise in the data.

For each GLM of experiments 1–2, six motion correction parameters and four physiological parameters were also included as regressors of no interest to account for signal changes not related to the conditions of interest. Physiological parameters were analyzed by using Retrospective Image Correction (Glover et al., 2000), as implemented for Matlab (cbi.nyu.edu/software). All group results were thresholded at a combined voxel threshold of $p < 0.05$ (FDR corrected) and a cluster extent threshold of $k = 10$.

2.5. Time-frequency analysis

For experiment 1 including the block-design TVA run, for every voxel inside the area of common activation, we extracted the time course of the BOLD response for every participant across the entire experiment. We subjected every time course to a time-frequency analysis by using a Morlet wavelet filter with five cycles in the frequency range 0.05–0.75 Hz (step size 0.05 Hz). The lower end of this frequency range was determined by the duration of the stimulus (20s), and the upper end was limited by the sampling rate of 650 ms of the TR, considering the law of the Nyquist frequency. After this wavelet transformation, the entire time-frequency signal was epoched separately for vocal and non-vocal trials with a baseline of 5*TR before stimulus onset and 39*TR after stimulus onset. The time-frequency power signal was subtracted by the mean of the signal in the baseline period. The $39*TR = 25.35$ post-stimulus period considered the stimulus duration of 20s and the time a little bit beyond the delay of the BOLD response. In total, we thus obtained time-frequency signals for 13 subjects $\times$ 42 trials $\times$ 4′262 voxels, resulting in 2′327′052 time-frequency signals.

2.6. ICA decomposition

After calculating the time-frequency signals across all trials, subjects, and voxels, we aimed to determine the most typical time-frequency response as displayed in these data. For this purpose, we first demeaned all time-frequency (tf) signals for each participant separately (Norman-Haignere et al., 2015), and then used a 2D Gaussian filter (size 6, SD 2.61) to spatially smooth each tf signal. The tf signals were then subjected to an independent component analysis (ICA), which was identical to the ICA procedure described in a previous report (Norman-Haignere et al., 2015). This previously used non-parametric approach estimates the ICA components by minimizing the entropy of the weight distribution, using a rotation method to maximize the negative entropy (i.e. negentropy; histogram-based approach) summed across the components. The sign of each component was adaptively flipped to reveal "sparse" and comparable ICA components across all group-level and single-participant ICA decomposition analyses described. Changing the sign of an ICA component does not alter the general ICA decomposition solution.

Since the estimation of the number of reliable components in an ICA analysis cannot be directly determined, we estimated the number of possible components hidden in the time-frequency signals by using a method derived from the "ranking and averaging independent component analysis by reproducibility" (RAICAR) approach (Yang et al., 2008). By using several realizations of the ICA decomposition procedure, this method estimates the reproducibility of ICA components with a correlation measure (i.e. Pearson's correlation coefficient) to rank, threshold, and align components across realizations. For each estimation of the number of ICA components, we used 30 ICA realizations with a random initialization resampling method; the threshold for the correlation coefficient was set to 0.99 to estimate only the most reproducible ICA components. The estimated number of ICA components was chosen as the estimated number that was consistently above the half-maximum of the reproducibility index across the 30 ICA realizations. The reproducibility index for each component was calculated by summing the correlation coefficients for all aligned components across the realizations above the correlation threshold.

2.7. ICA component estimation based on experimental factors and brain lateralization

The above-described ICA decomposition was applied to the data in four different ways. Since an ICA analysis cannot be easily accomplished in a hierarchical statistical manner (including the differentiation of experimental factors) and given non-linearity effects when combining or splitting data from different conditions, we performed each of these analyses separately by splitting the data into relevant subsections. This leads to independent ICA estimations across these approaches with no statistical dependencies across these approaches and with no issues potentially relating to double-dipping. Also, the number of relevant ICA components for each factor can only be estimated in separate analyses (see below).

Here we fully describe the four different ways of the ICA in detail: first, we quantified BOLD oscillations to sound processing in general since we aimed at finding general dynamic BOLD oscillations to temporally extended sounds; second, we quantified differential BOLD oscillations for vocal and nonvocal trials since vocal signals have higher social relevance; third, we quantified BOLD oscillations separately in the left and right AC given some potential processing differences of sound; and, fourth, given that the human brain also shows endogenous BOLD oscillations in the absence of external stimulation, we quantified BOLD oscillations in the resting brain to compare them to task-related BOLD oscillations.

For the time-frequency signals derived from experiment 1, we first estimated the ICA components by taking both the vocal and the non-vocal trials into account in one single analysis. This analysis was done to estimate common time-frequency components for vocal and non-vocal trials. We additionally estimated whether the weight that each voxel contributed to each component was higher in the left or the right hemisphere. For each component, we averaged voxel weights in the left and the right hemisphere separately for vocal and non-vocal trials, and we calculated the difference between left and right voxel weights. We used a Wilcoxon rank sum test (WRST) to estimate whether there was a significant laterality difference in mean voxel weights for each component between vocal and non-vocal trials. We also estimated whether vocal or non-vocal trials contributed more to each component by determining the mean contribution of vocal and non-vocal trials for each component. We used a Wilcoxon signed rank test (WSRT) to estimate

whether there was a significant difference in mean voxel weights for each component for vocal and non-vocal trials. We also determined the mean weight that each of the 21 vocal and 21 non-vocal trials contributed to each component. The relative weight of trials was ordered from highest to lowest.

In a second analysis, we estimated the ICA components separately for vocal and non-vocal trials to find out whether this analysis would reveal different components for the two types of sounds while keeping all other analysis factors identical to the first analysis. For two of the ICA components (C4 and C5; see results section), we saw a latency difference in the higher frequency range when we estimated it on the vocal or the non-vocal trials. This latency difference was most pronounced on the frequency band centered on 0.65 Hz. To obtain a quantitative measure of this latency difference, we re-ran the ICA decomposition in each participant and determined the components that most resembled the original components from the general ICA analysis. From these components in each participant, we extracted the power signal time course in this frequency range centered on 0.65 Hz and determined the time point of the maximum peak power signal in each participant. We used a WSRT to estimate whether there was a significant difference in peak power latency between vocal and non-vocal trials. We also determined the mean weight that each of the 21 vocal and the 21 non-vocal trials contributed to each component. The relative weight of trials was ordered from highest to lowest.

Similarly, in a third analysis, we estimated the ICA components separately for the left and the right hemisphere voxels but combining vocal and non-vocal trials. This third analysis was based on the notion that the brain might show lateralized responses to vocal and non-vocal trials, including the potential hypothesis that the left and right hemispheres might be sensitive to different modulation rates in acoustic signals (Giraud and Poeppel, 2012) next to a similar gradient of processing slow vs. fast modulation rates in posterior-to-anterior STC (Hullett et al., 2016). All other analysis factors were identical to the first analysis.

The fourth analysis, finally, was applied to the time-frequency data from experiment 2. The time-frequency analysis data were identical to those in experiment 1. The ICA decomposition for these data should reveal the default oscillation activation in the auditory cortex during a resting-state condition. The time-frequency analysis and the ICA decomposition was applied to the same 4′262 voxels as part of the common area of activation determined from experiment 1 of the event-related and the block-design run. For each of these four different analyses, we determined the ICA components that best represent the single time-frequency signals. For each component, we determined the weights that each voxel contributed to each component.

2.8. Acoustic analysis

We determined the modulation spectrum of each sound stimulus (Fig. S1) by converting the amplitude waveforms of the vocalizations to their spectrogram with a Hamming window length of 1.81/BW (bandwidth (BW) 100 Hz, frequency range 0.05–8 kHz), which was then filtered with a 2D Gaussian filter (SD 1.3, size 3) and subjected to a 2D Fourier transform (zero-padded to match the length of the longest stimulus). From this transform, the power was computed, and the zero-frequency components were shifted to the center of the spectrum. The resulting modulation spectra for each stimulus were both averaged along the time axis and the frequency axis to give an average representation of the spectral and temporal modulation, respectively. For each stimulus, we computed the Shannon entropy of the modulation spectrum along the spectral and the temporal axis as an indicator of the respective modulation complexity.

We also performed an acoustic analysis of the stimuli used in experiment 1 (block-design) to assess if the time-course of the obtained time-frequency ICA components was influenced by the time course of the acoustic stimuli used. We, therefore, converted each of the 21 vocal and 21 nonvocal stimuli with a 20s duration to a cochleogram representation. The amplitude waveform of each stimulus was first subjected to a Gammatone filter-bank including 128 channels and a frequency range of 0.05–8 kHz. From this filtered signal, we estimated the cochleogram with a window length of 880 samples and then extracted the mean time course of the power signal in six non-overlapping frequency bands based on the center frequencies 0.2, 0.4, 0.8, 1.6, 3.2, and 6.4 kHz.

The time course of the power signal in these frequency bands, which was resampled to the sampling rate of the fMRI signal, was then correlated with the time course of each frequency band of the ICA component. To consider the delay of the hemodynamic response in the ICA components, we correlated the cochlear power signal with the part of the ICA components ranging from 6 to 37*TR (3.9–24.05s) after stimulus onset; we consider this period as the approximate 20s segment of the BOLD response corresponding to the 20s segment of auditory stimulation taking into account the delay of the BOLD response of about 4s. The significance of the correlation was tested using a non-parametric permutation test (10′000 iterations, $p < 0.05$) while shuffling the mapping between the cochleogram frequency band and the ICA component frequency band. Permutation tests do not require any corrections for multiple testing or adjustments to the alpha level.

3. Results

3.1. Auditory cortical activations for voice processing

In experiment 1, we calculated the classic activation map by comparing the activity for vocal and non-vocal sounds using both an event-related stimulation (i.e. 500 ms trials) (Capilla et al., 2013; Fruhholz et al., 2016a) and a block-design stimulation (i.e. 20s trials) (P Belin et al., 2000) (Fig. 1). For localizing the voice-sensitive region in the auditory cortex, previous studies used either an event-related (Dricu et al., 2017; Fruhholz et al., 2016a) or a block-design stimulation (Pascal Belin et al., 2000; Cyril R. Pernet et al., 2015) leading to slightly different cortical definitions of this area. To enable, first, a conservative estimation of the voice-sensitive auditory cortex that is common to event-related responses to short vocal signals and to sustained responses to extended vocal signals, we combined both experimental designs. Second, to validly determine that peak locations of important BOLD oscillations (see below) are largely separate from common peak location of BOLD activation level differences, we needed to determine the latter in both commonly used experimental approaches (i.e. event-related and block design).Figure 1

For the functional scan including the event-related design, we found bilateral and spatially extended activity in the low- and higher-level AC with peak activations in the bilateral anterior and posterior STC, as well as in the left secondary AC and the right primary AC (voxel threshold $p < 0.05$ (FDR corrected), cluster size threshold $k = 100$, $n = 13$) (Fig. 1a). Activations were specifically located in the left PTe [MNI xyz -52 -20 6, $z = 6.08$], aSTG [-60 -4 0, $z = 5.73$], and pSTG [-60 -32 10, $z = 6.30$], as well as in the right Te1.2 [56–8 2, $z = 5.86$], pSTG [66–18 8, $z = 5.89$], and pSTS [64–28 0, $z = 5.64$]. This cortical distribution of activity was similar to the described locations of the TVAs reported previously (P Belin et al., 2000; Cyril R. Pernet et al., 2015); no higher activity was found when we compared nonvocal with vocal sounds.

The same analysis was performed for the functional data of experiment 1 including the block design (Fig. 1b). Vocal trials showed peak BOLD activity in the left Te1.2 [-58 -20 6, $z = 5.34$], aSTG [-60 -10 0, $z = 6.42$], and mSTG [-66 -18 6, $z = 5.79$], as well as in the right Te1.2 [52–6 -2, $z = 5.91$], mSTG [62–14 -6, $z = 5.52$], and pSTG [62–26 8, $z = 5.75$]. No higher activity was found when we compared nonvocal with vocal sounds. The common areas of cortical activations for voice processing were defined by the area that spatially overlapped between the functional data from the event-related and the block design TVA run (Fig. 1, red dashed line).

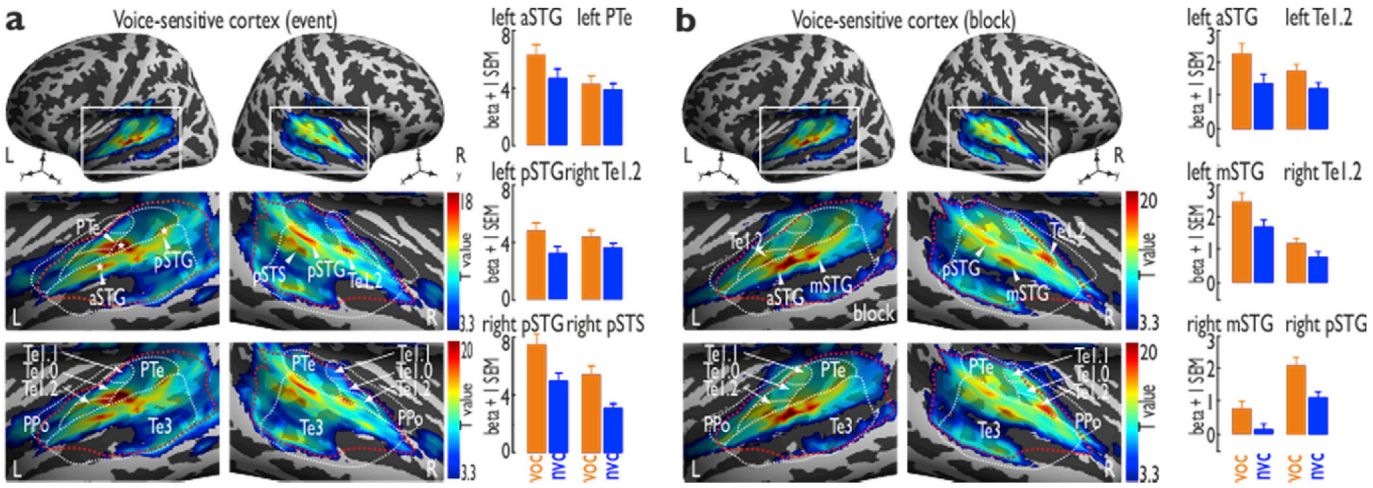

Fig. 1. BOLD activation for vocal and non-vocal trials.

(a) The voice-sensitive cortex, or temporal voice area (TVA), was determined by comparing vocal and non-vocal trials in experiment 1 (event-related design); middle panel is an enlarged view of the upper panel. The TVA overlaps with cortical boundaries (white dashed line) for primary AC (regions Te1.0, Te1.1, Te1.2 Scholtens et al., 2016), secondary AC (PPo and PTe), and higher-level AC (Te3) in the STC based on anatomical definitions (lower panel); red dashed outline defines the area of common activation for voice trials in the event-related and block-design (see (b)). Beta estimates shown for the peak cortical activations (right panel) found in experiment 1 for vocal (voc) and non-vocal trials (nvc).

(b) The TVA defined in experiment 1 using a block-design (upper panel); voxel threshold $p < 0.05$ (FDR corrected), cluster size threshold $k = 100$, $n = 13$. The middle panel shows an enlarged view of the upper panel of the voice-sensitive cortex.

3.2. Auditory cortical BOLD oscillations elicited by sounds

Having looked at voxel-wise activity differences, we next investigated whether this prioritization of vocal signals also appears in neural infra-slow BOLD oscillations in response to extended sound stimulation. For the functional data resulting from experiment 1 (block-design), oscillatory brain responses were quantified as measures of induced neural BOLD oscillations. The power of oscillations was baseline corrected to 3.25s before stimulus onset and quantified in the frequency range 0.05–0.75 Hz, which was defined by stimulus duration (20s) and sampling rate (1.5 Hz, corresponding to TR = 0.65s; Nyquist frequency 0.75 Hz) in its lower and upper limit, respectively. To this end, we subjected the time course of each cortical voxel that commonly (i.e. event-related and block design) showed higher activation in response to vocal sounds to a time-frequency (tf) analysis by using a continuous wavelet transform, resulting in a time-frequency-power decomposition of the BOLD signal at each voxel ($n = 2'327'052$ tf signals). Since the tf signal varies across voxels, we aimed at determining the most common tf response modes by using an independent component analysis (ICA) on all tf signals from each voxel, trial, and participant. The resulting IC weights of each voxel on each component was mapped on the AC, and we mainly identified, quantified, and spatially localized voxels with positive weights and their differential spatial distribution.

Previous work suggests that there are a few different response modes in the AC's response to sound (Bathellier et al., 2012), and in humans, a major distinction of the raw BOLD signal suggests both a transient and a more sustained response mode in the AC (Seifritz et al., 2002). Given this prior work, we assessed whether more diverse response modes could be identified in neural BOLD oscillations. We identified six major response modes for BOLD oscillations (ICA components C1-6, Fig. 2a and b; see Fig. S2 for the consistency of ICA components across participants). We estimated the number of reliable ICA components hidden in the signals by using a reproducibility index (Yang et al., 2008) across 30 unconstrained repetitions of the ICA analysis (30 iterations, half-max = 217.29, $n = 2'327'052$). Three of the six major response modes appeared as sustained oscillatory responses located in the lower frequency (LF; 0.05–0.30 Hz), middle frequency (MF; 0.30–0.55 Hz), and higher frequency range (HF; 0.55–0.75 Hz) relative to the frequency range investigated. Although C3 did not show continuous oscillation for the entire period compared to C1-2, we nevertheless classified it as a sustained component given that it extended over at least half of the post-onset period and did not show the temporally constrained appearance as the transient components C4-6. The other three modes appeared as transient responses in the HF range at an early (2.60s), mid (9.10s), and late peak latency (24.05) after stimulus onset (Fig. 2a). The three major frequency ranges were defined on the visual inspection of the BOLD oscillation pattern (Figs. 2–4) and the correlation analysis with acoustic temporal patterns of the sounds (see below and Fig. 5). The frequency ranges only have a descriptive value to better summarize the data.

For the sustained and transient oscillation components, we did not find an exclusive contribution of either vocal or non-vocal sounds, but rather a mixed contribution of both (Fig. 2e). Furthermore, we found no associations with the overall spectral and temporal sound features on the basis of the modulation spectrogram of sounds that scored the highest and lowest weight on the components, respectively (Fig. S1). The dynamic overall temporal and spectral profile of the sounds were unlikely to directly drive the BOLD oscillatory patterns.

The three sustained response modes, especially components C1–C2, showed oscillations for the entire duration of the sound and had positive voxel weights across the secondary AC and higher-level STC, with an additional contribution of the left primary AC for the HF oscillations. C1-2 might especially signify ongoing sound stimulation and analysis by continuous induced oscillations that are based on continuous neural firing (Galazyuk and Feng, 1997; Wang et al., 2005). Component C3 showed sustained LF oscillations starting at about 13s after stimulus onset, had large negative voxel weights in the posterior secondary AC, and had a resemblance to a resting-state brain oscillation rather than to stimulation-related oscillations (Fig. 3). This resemblance is demonstrated with data from experiment 2, which included a resting-state brain acquisition during a "task-free" recording of whole-brain data in an independent sample of participants. These resting-state brain data were analyzed similarly to the data of experiment 2 (block-design) by epoching the data into 20s segments of simulated and randomly assigned vocal and non-vocal events.

Although the sustained LF component C3 resulting from the data of experiment 1 might thus largely reflect default brain activity and physiological periodicities given its resemblance to the BOLD oscillation analysis on the resting-state brain data, the LF component C3 was

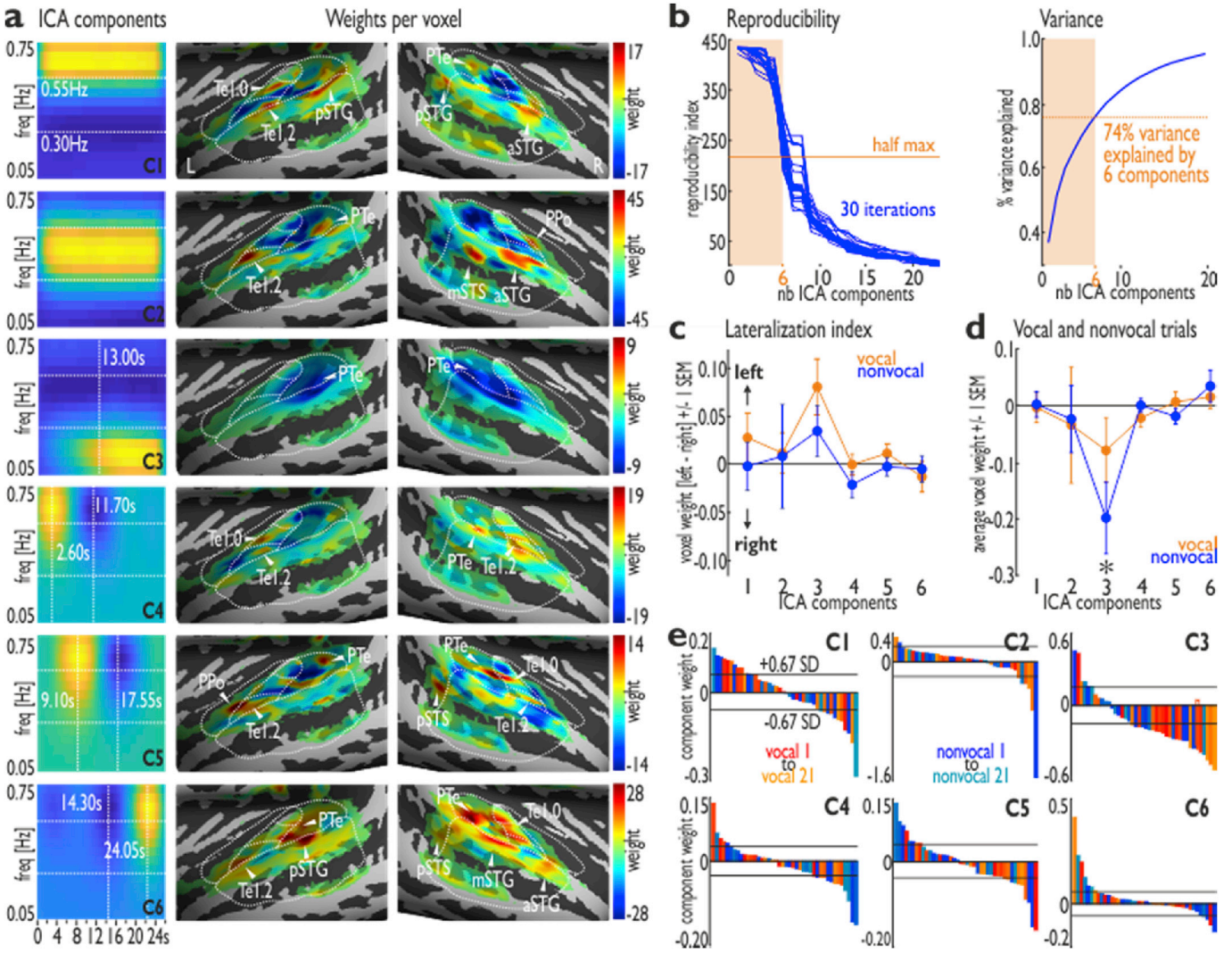

Fig. 2. Common auditory cortical oscillations in response to vocal and non-vocal sounds.

(a) Six ICA components (C1-6, left panel) for the time-frequency (tf) signals estimated across vocal and non-vocal trials. Oscillations had three major frequency ranges: high >0.55 Hz (HF), mid 0.3–0.55 Hz (MF), and low <0.30 Hz (LF). C1 showed sustained HF oscillations with high voxel weights in the left (Te1.0, Te1.2, pSTG) and right regions (PTe, aSTG, pSTG); C2 involved sustained MF oscillations peaking in the left (Te1.2 PTe) and right regions (right PPo, aSTG, mSTS); and C3 had sustained LF oscillations starting at ~13 s and peaking mainly in the bilateral PTe.

(b) The RAICAR approach estimated six replicable oscillatory response components in the tf signals by using a half-maximum reproducibility threshold (74.08% variance explained).

(c) C3 seemed more lateralized to the left hemisphere, but only for vocal trials; vocal and non-vocal trials did not show a significant lateralization difference across components; data are mean±s.e.m.

(d) Vocal and non-vocal trials did not contribute significantly differently to the six ICA components, except for C3, where vocal trials showed more positive weights.

(e) Ordered mean weights of vocal (red) and non-vocal trials (blue) for C1-6; both contributed equally to C1-6. Dashed lines indicate the±0.67SD threshold (~25% highest and ~25% lowest scores given a normal distribution of the data).

differentially influenced by the stimulus type and thus does not only reflect resting-state oscillations or oscillations due to neurophysiological noise. This sustained LF component C3 had indeed increased negative weights for nonvocal compared to vocal trials (Wilcoxon rank sum test (WRST), two-tailed, $Z = 2.205$, $p = 0.027$, $n = 13$; Fig. 2d) and stronger left lateralization for vocal trials (WSRT, two-tailed, $Z = -1.223$, $p = 0.017$, $n = 13$). Although this LF component might partly reflect resting-state oscillations, these latter differential effects nevertheless suggest that vocal trials in particular more easily change the "task-free" resting-state brain mode into a more stimulation-related mode. Vocal and non-vocal trials did not show a significant lateralization difference across components (WRST, two-tailed, all Zs < 1.282, all ps > 0.200, $n = 13$; Fig. 2c).

Unlike the sustained responses, the transient HF response modes (centered on 0.65 Hz) had different peak latencies after stimulus onset (C4 - 2.60s, C2 - 9.10s, and C3 - 24.05s), and the location of high voxel weights shifted from the primary and secondary AC for the early HF oscillation component (C4) to also include the higher-level STC for the components with the mid (C5) and the late HF oscillations (C6). While the highest weights for C4 were found in the left (Te1.0, Te1.2) and right low-level regions (Te1.2, PTe), the maximum weights also extended more to the higher-level STC for C5-6.

3.3. *Vocal sounds elicit unique BOLD oscillation patterns*

Since it can be hypothesized that the BOLD oscillatory brain responses to vocal and non-vocal sounds might differ given their difference in social-biological and communicative significance (P Belin et al., 2000; Petkov et al., 2008), we estimated the major oscillatory response modes

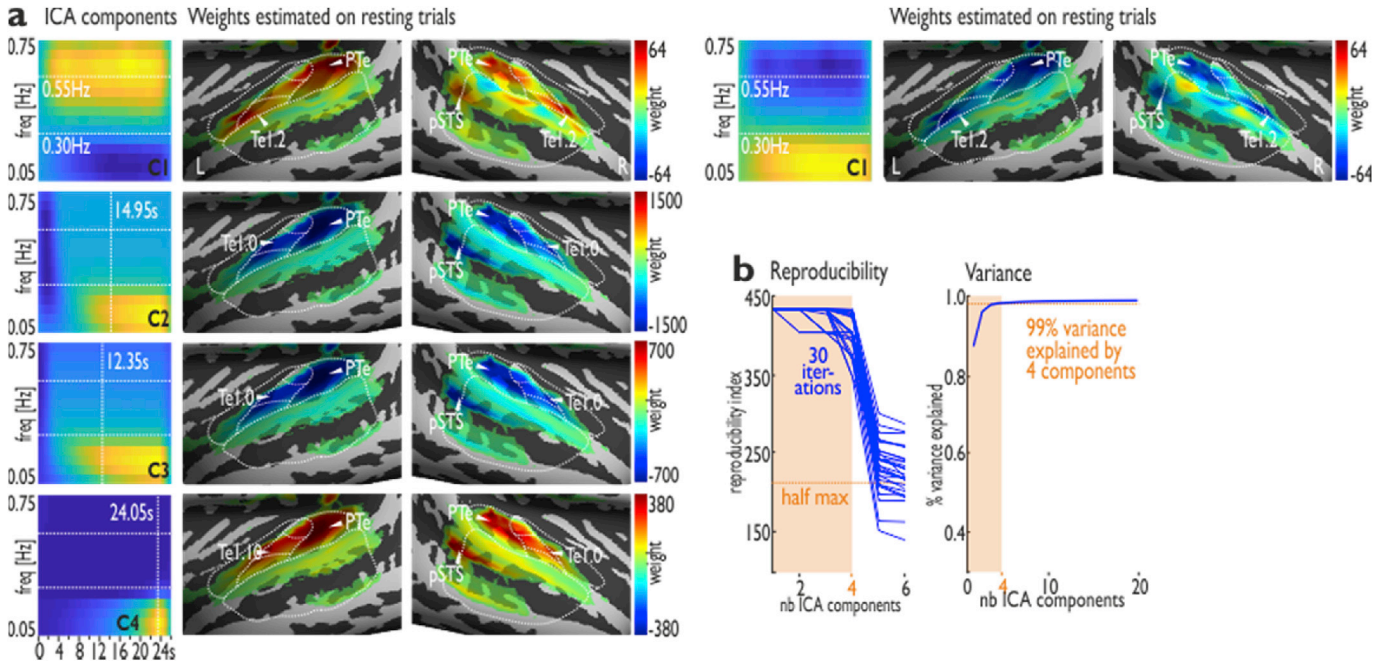

Fig. 3. Cortical oscillations estimated during a resting-state functional brain acquisition.
(a) Four ICA components (left panel, C1-4) were identified for the tf signals estimated on the functional brain data during resting-state acquisition (n = 1′159′264). C1 showed sustained HF oscillations with a positive power signal accompanied by a sustained low-frequency oscillation with a negative power signal. The left panel shows the same ICA profile for C1, but with a reversed sign. C2–C4 showed mainly LF oscillations, starting at different latencies after stimulus onset. All components showed a similar distribution of voxel weights with high negative/positive weights in the primary auditory cortex, PTe, and right pSTS.
(b) The RAICAR approach (30 iterations, half-max = 217.50, n = 1′159′264) estimated four components to be reproducible with the highest maximum of the reproducibility index. Although the RAICAR approach also indicated that five components would be consistently detected considering the half-maximum threshold, four components already explained 99.25% of the variance, and these four components were similar in terms of their ICA profile and their distribution of voxel weights in the auditory cortex.

separately for vocal (n = 1′163′526 tf signals) and non-vocal sounds (n = 1′163′526 tf signals) in the next step (Fig. 4a; see also Figs. S3–4). While five response components (C1–C5), as described above, could be replicated when estimated on non-vocal trials (RAICAR 30 iterations, half--max = 216.71, n = 1′163′526; 70.74% variance explained), all six components could be replicated for vocal trials (RAICAR 30 iterations, half-max = 217.27, n = 1′163′526). In addition, we found a sustained oscillatory component in the lower MF range (C7) with high voxel weights in the bilateral primary AC and right STC (Fig. 4a and b). It also seems that the components for the vocal trials were being more robustly reproduced, given the shape of the ICA reproducibility indices (Fig. 4b, left panel). The reproducibility indices for non-vocal trials steeply dropped to the threshold level for C1–C5, whereas these indices stayed on a relatively high level for C1–C7 estimated on vocal trials.

None of these transient components showed strong lateralization in the brain for vocal (WSRT two-tailed, all Zs < 1.292, all ps > 0.196, n = 13) and non-vocal trials (WSRT two-tailed, all Zs < 1.363, all ps > 0.173, n = 13), and there was no lateralization difference for C1-5 (WSRT two-tailed, all Zs < 0.564, all ps > 0.573, n = 13) (Fig. 4d). However, the early and mid-latency HF components showed a significantly faster peak latency for vocal rather than non-vocal sounds (Fig. 4c), such that C4 (WSRT two-tailed, Z = 2.788, p = 0.005) and C5 (WSRT two-tailed, Z = 2.735, p = 0.006) had significantly faster max peak latencies for vocal compared with non-vocal trials.

3.4. The right brain shows more diverse BOLD oscillations

We finally also estimated the major oscillatory response modes separately in the left and the right hemisphere but collapsed over vocal and nonvocal trials (Figs. S5–7). Five ICA components (C1-5; RAICAR 30 iterations, half-max = 217.28, n = 1′099′644) were identified for the tf signals estimated on left hemispheric voxels (n = 1′099′644), while six ICA components (C1-6; RAICAR 30 iterations, half-max = 217.29, n = 2′327′052) were identified for tf signals estimated on right hemispheric voxels (n = 1′227′408). Components C1-5 were similar for left and right voxels, but C6 was unique for right voxels. C6 showed oscillations in the higher frequency (HF) range with a maximum appearing 22.75s after stimulus onset and with the highest voxel weights in the Te1.0, PTe, and STG and STS. For left and right voxels, we also separately scored the contribution of vocal and non-vocal trials. For right voxels, we found a significant difference (WRST, two-tailed, Z = 2.00, p = 0.046, n = 13) for component C3, where vocal trials showed significantly higher voxel weights than for non-vocal trials.

Components C4 and C5 were similar in their ICA profile in terms of HF oscillations, but the peak maximum in the power signal differed when compared across the left and right brain with C4 (WRST, two-tailed, Z = 2.469, p = 0.014) and C5 (WRST, two-tailed, Z = 2.879, p = 0.004) showing faster peak latencies in the right brain.

3.5. Acoustic sound features and BOLD oscillations

Although we found no associations of BOLD oscillations with the overall spectral and temporal sound features on the basis of the modulation spectrogram of sounds (see above; Fig. S1), we computed a more detailed analysis to investigate the relationship between dynamic acoustic features and the BOLD oscillation patterns. We found that certain sound features in certain frequency bands correlated with the time course signal in the oscillatory signals when estimated on all sounds together. This was estimated by creating the cochleogram for each sound (Fig. 5), and then calculating correlation matrices for the cochleogram-oscillation relationship (COR) across the frequency bands of the cochleogram and the oscillatory signals. Most importantly, the sustained HF

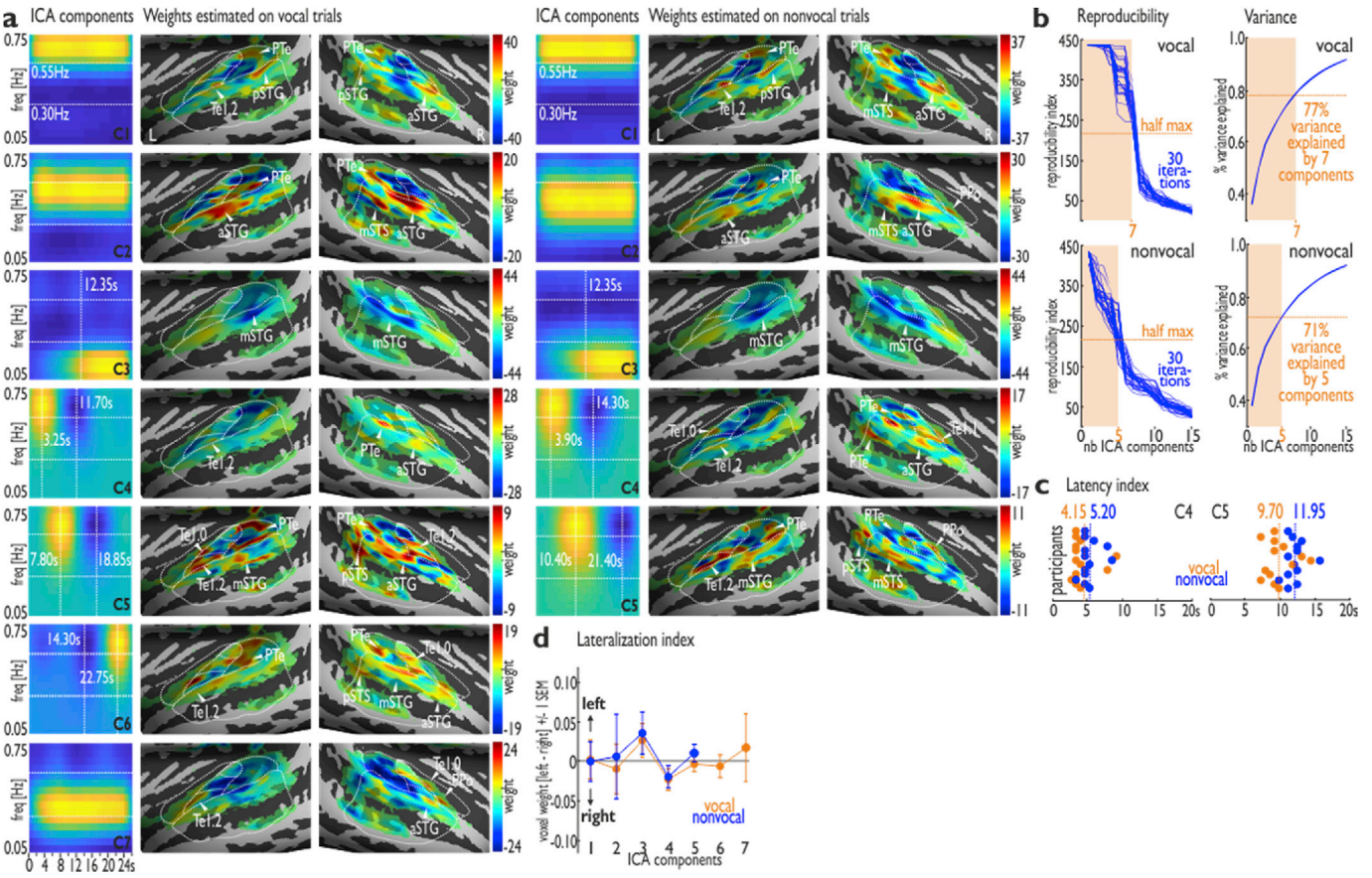

Fig. 4. Separate cortical oscillations estimated for vocal and non-vocal sounds.

(a) Seven ICA components (C1-7, left panel) for the tf signals estimated on vocal trials and five ICA components (C1-5, right panel) for non-vocal trials. C1-5 were similar for vocal and non-vocal trials, but C6-7 were unique to vocal trials. C6 showed HF oscillations (max peak 22.75s) with the highest weights in the left (Te1.2, PTe) and right regions (Te1.0, PTe, STG, STS), while C7 showed sustained lower MF oscillations with the highest weights in the left (Te1.2) and right regions (Te1.0, Te1.2, aSTG).

(b) The RAICAR approach estimated six replicable oscillatory response components in the tf signals for vocal trials (77.11% variance explained), while five replicable components were estimated for non-vocal trials (70.74% variance explained).

(c) Components C4 and C5 were similar in their ICA profile in terms of HF oscillations, but C4 and C5 had significantly faster max peak latencies for vocal and non-vocal trials; colored numbers indicate the mean latency for vocal and non-vocal trials.

(d) Vocal and non-vocal trials did not reveal a significant lateralization, and there was no lateralization difference for C1-5; data are mean±s.e.m.

component correlated positively and the LF oscillation correlated negatively with the low-range cochleogram, while sustained MF oscillation correlated positively with the high-range cochleogram (Fig. 5b). For the transient response components, only mid and late latency HF components correlated negatively and positively (respectively) with the low range cochleogram.

For the ICA estimation separately on vocal and nonvocal trials, we also found some oscillatory modes partly correlated with dynamic acoustic features of the cochleogram (Fig. 5c). While CORs for nonvocal trials largely followed the principles identified for the CORs across all sounds (Fig. 5b), we found a different pattern for the vocal trials, especially for the sustained components C1-2. These components showed largely a reversed pattern for vocal compared to nonvocal sounds, with the sustained HF oscillations being negatively correlated with the low-range cochleogram, and the sustained MF oscillations being positively correlated with the low-range cochleogram and negatively correlated with a high-range cochleogram (2.2–4.2 kHz). For vocal trials, some of the COR for the transient components showed some resemblance to the CORs for sustained components, such as C5 to C1, and C6 to C2.

4. Discussion

Besides the previously known prioritization of vocal signals in the higher-level STC according to the level of neural activity (i.e. activations maps) (P Belin et al., 2000; Petkov et al., 2008), the neural prioritization of voice signals compared to other sounds starts significantly earlier in the neural auditory system as shown here by the new measure of stimulation-related BOLD oscillations (i.e. BOLD oscillation maps) (Lewis et al., 2016). Here we critically extend previous findings of infraslow BOLD oscillations to low-level sensory stimulation (Lewis et al., 2016) to the domain of BOLD oscillations in response to complex auditory stimuli. The neural measure of infraslow BOLD oscillation seems to provide a valid, novel, rich, and differential measure into the neurocognitive mechanisms of social cognition of conspecific voices. Concerning the differential nature of this measure, of critical note is that the spatial distribution of ICA weights (Figs. 2 and 4, Fig. S5; see the labeled location of ICA weight maxima) only marginally overlapped with peak activations found in classical activations maps when comparing neural activity for vocal against non-vocal sounds (Fig. 1; see the labeled peak location of activations). Thus, the ICA analysis seems to quantify other neural processes during voice processing rather than simply being driven by larger neural responses to vocal sounds in specific AC subregions. The latter would result in a high similarity in the spatial distribution of the activations maps and the oscillation maps, which was not evident in this study.

We specifically determined the most dominant BOLD oscillations in

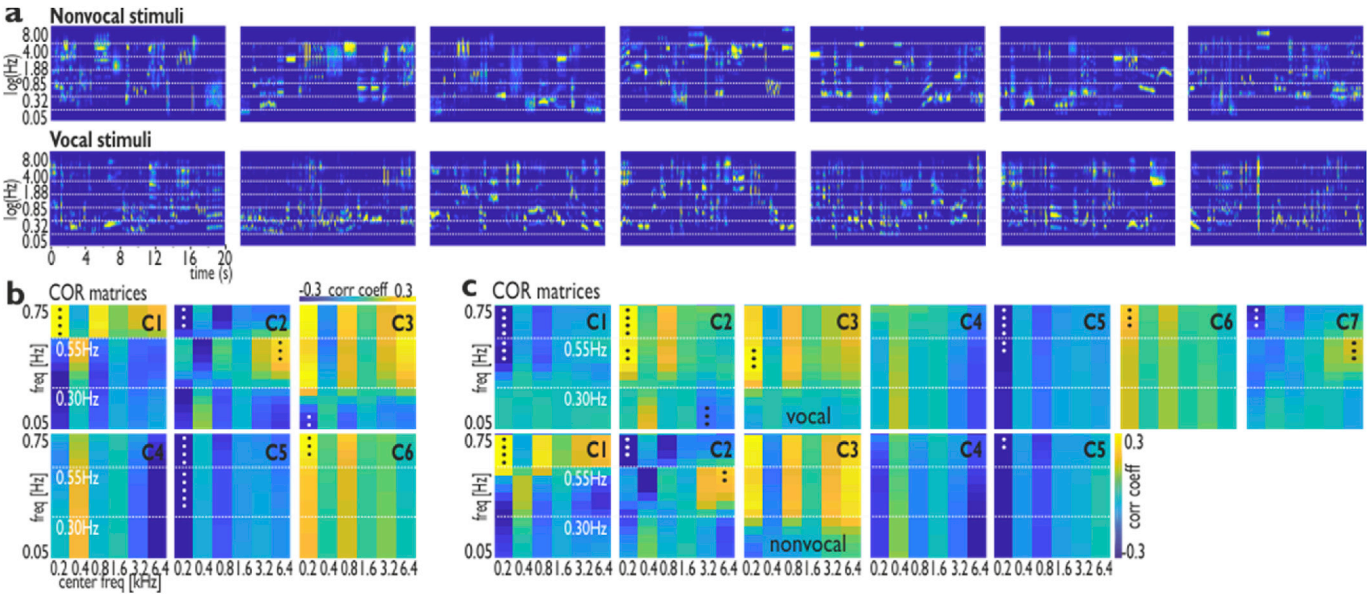

Fig. 5. Cochleograms of all 21 nonvocal stimuli (upper panel) and of all 21 vocal stimuli (lower panel) and cochleogram-oscillation-relationship.
(a) Cochleograms of example stimuli. The frequency axis on is in log(Hz) units, and the white dashed lines indicate the center frequency of non-overlapping bands that were used for the correlation analysis (COR) of the tf-signal of the ICA components.
(b) Correlation matrices of the cochleogram-oscillation-relationship (COR, 15 × 6 matrices, Pearson's correlation coefficient) for the relation between the power time course in six log-spaced frequency bands of the cochleogram of the sounds and the power time course (3.90–24.05s post-stimulus onset) in the 15 frequency bands of C1-6 (see Fig. 3a); black/white dots mark significance (permutation test, 10′000 iterations, $p < 0.05$). C1, C3, and C5 showed significant positive or negative correlations with the low-range cochleogram (0.05–0.26 kHz) of the stimulus; C2 showed a correlation with the high-range cochleogram (4.80–8.00 kHz).
(c) CORs for C1-7 for vocal trials (upper panel) and for C1-5 non-vocal trials (lower panel) (see Fig. 4a); permutation test, 10′000 iterations, $p < 0.05$; significance marked with a black or white dot.

the AC in response to sound stimulation and found a major distinction between sustained and transient BOLD oscillation patterns. This distinction of sustained and transient response modes resembles a previous observation concerning the raw BOLD time course (Seifritz et al., 2002), but our findings critically extend this observation to the oscillatory properties of the BOLD response. Our data show a diversity of sustained and transient modes, such that transient responses not only appear at stimulus onset and are localized to the primary AC as previously noted (Seifritz et al., 2002), but they also appear throughout the time course of the stimulus and in any subregion of the AC/STC. Concerning these early, mid-latency, and late transient oscillation patterns, the early transient component might indicate the onset of the sound (Lehmann et al., 2007), while the mid-latency transient HF component might indicate cognitive engagement with, and significance evaluation of, the stimulus after several seconds of sound decoding (Mesgarani and Chang, 2012). This mid-latency HF component resembles neural binding processes in the retention phase of working memory (Pina et al., 2018) as well as continuous information integration during working memory updating (Rac-Lubashevsky and Kessler, 2018) at a critical time-point after stimulus onset. Finally, the late transient HF component might be an offset indicator of the sound (Baba et al., 2016).

Such neurophysiological onset and offset as well as sustained responses have been described by previous EEG and MEG studies on auditory processing reflected by the N1–P2 complex and the sustained field of the event-related potential (ERP) (Gutschalk et al., 2002; Pantev et al., 1996, 1994). These amplitude related ERP effects have been localized to the supra-temporal plane of the auditory cortex. Compared to our current study, these previous studies used relatively short auditory sounds of maximum 2s duration and only quantified the amplitude level of the EEG/MEG signal. We quantified infraslow BOLD oscillations to temporally more extended sound objects with social characters and could identify similar onset and offset as well as sustained responses as quantified by the oscillatory feature of the BOLD signal. Thus, these on-/offset responses seem to be a characteristic feature of brain signals at many levels of signal quantification. However, these physiological responses are not only common to general sound processing in the current study but partly specific to processing vocal sounds.

Besides these more common sustained and transient BOLD oscillatory patterns found across all sounds, there seem to be distinct and faster neural oscillatory responses unique to vocal signal processing. These earlier onsets of oscillatory responses to voice sounds were found as a faster transient HF onset response in the primary AC and STC, as a faster mid-latency transient HF response, and as a dedicated offset response in all auditory regions. Furthermore, a distinct oscillatory response was found as a sustained low MF oscillatory response mode, especially in the primary AC. This finding demonstrates the significant neural preference of auditory cortical regions for the appearance (i.e. faster onset response), presence (i.e. unique sustained response), and disappearance (i.e. dedicated offset response) of vocal signals, which can be specifically quantified by stimulation-related BOLD oscillations. Using EEG, a similar earlier onset and stronger oscillation desynchronization in the alpha and beta brain oscillations was found for vocal compared to nonvocal sounds of only 750 ms duration (Lévêque and Schön, 2013) as well as for newly learned compared to old voices in the beta range (Zaske et al., 2014).

Vocal signals have a unique and prominent role in mammalian and primate communication. Therefore, decoding vocal signals at any phase of their appearance and disappearance seems of higher importance than decoding other sounds within and across species (Jiang et al., 2018). Furthermore, vocal signals usually have a unique and complex temporal structure making temporal processing of voice information crucial to vocal communication. Detecting voice offsets and vocalization durations via auditory cortical synapses tracking tonal on and off responses (Scholl et al., 2010) is thus of critical importance for social and communicative adaptations (Baba et al., 2016).

Regarding the temporal aspects of vocalizations and given the dynamic properties of some of the BOLD oscillation patterns identified, we also found that some BOLD oscillatory patterns are associated with dynamic acoustic features in sounds. Similar to previous reports on brain

oscillations following the periodicity of speech (Giraud and Poeppel, 2012) and music cycles (Trost et al., 2014), we found that some BOLD oscillatory patterns track dynamics sound patterns especially in the low- and high-range patterns of the cochleogram of sounds (Fruhholz et al., 2016b; Pannese et al., 2016). First, for the analysis with the ICA estimation based on all trials (Fig. 2), low-range cochleogram associations were found for the sustained HF and LF oscillations and the transient mid and late HF components, while only MF oscillations correlated positively with the high-range cochleogram. The sustained components thus seem to partly track basic acoustic features across most parts of the sound stimulus, such as the pitch (i.e. low-frequencies of the cochleogram) and tonal quality features of the sounds (i.e. high-frequencies of the cochleogram). Pitch is mainly represented by the low-range cochleogram, while tonal qualities responsible for the timbre of the sound that aid in localization and intelligibility are represented by the high-range cochleogram (Monson et al., 2014). Second, for the separate ICA analyses of vocal and nonvocal sounds, we found that the sustained components C1-2 largely showed a reversed pattern for vocal compared to nonvocal sounds. This seems to indicate that different sustained oscillatory patterns track the differential dynamics of vocal and nonvocal sounds, which might be linked to differential dynamics in the most relevant spectral frequency range in both types of sounds (Gygi et al., 2004). The faster feature changing nature of nonvocal sounds are positively tracked by the HF sustained component, while the more slowing changing nature of feature dynamics in vocal sounds are positively tracked by the LF sustained component (Woolley et al., 2005). For vocal trials alone, some of the associations for the transient components showed some resemblance to the associations for sustained components, for example, a resemblance of C6 to C2. This finding might indicate that some of the sustained HF and MF oscillation components receive some temporally transient enhancements, as indicated by the transient enhancements driven by certain acoustic features of vocal signals compared to nonvocal sounds.

Here we add a final note on the infraslow brain oscillations that we investigated. Infraslow oscillations (Monto et al., 2008; Wang et al., 2014) are only a small and often neglected part of the full range of neural oscillations during neural processing (Buzsáki and Draguhn, 2004). The focus on infraslow oscillations is limited by the still non-optimal temporal sampling resolution that exists even in fast fMRI. However, infraslow oscillations can carry important information about mental processes (Buzsáki and Draguhn, 2004; Lewis et al., 2016), and seem to reflect underlying neural oscillations (Lewis et al., 2016). Most critically, they can provide differential and dynamic neural measures beyond traditional BOLD measures of social cognition using dynamic stimuli, such as vocal signals compared to other sounds.

In conclusion, our study provides first-ever evidence for the significance of stimulation-related BOLD oscillations during social cognition of vocal signals. Specifically, we found that the primary AC not only prioritizes the processing of vocal over other auditory signals in terms of the activity level but also in terms of BOLD oscillations already starting at the level of the primary AC and extending to higher-level AC regions. This points to an early prioritization of vocal signals beyond the previously known prioritization in the higher-level STC according to activity level (Andics et al., 2010; P Belin et al., 2000; Petkov et al., 2008). Given this fast prioritization even in infraslow oscillation signals (Buzsáki and Draguhn, 2004; Lewis et al., 2016), stimulation-relevant neural information is meaningful and significant despite previous observations of low-frequency oscillations during task-free resting-state brain activity (Zuo et al., 2010). In this low-frequency range, vocal signals show a specific oscillatory fingerprint beyond common oscillations for general sound processing.

Author contributions

S.F. and W.T. contributed to designing the experiments, data acquisition, data analysis, and writing the manuscript; D.G. contributed to data acquisition; P.B. contributed to designing the experiments, data analysis, and writing the manuscript.

Acknowledgments

This study was supported by the Swiss National Science Foundation (SNSF PP00P1_157409/1 and PP00P1_183711/1 to SF). We thank Caitlyn Trevor for helpful comments on the manuscript. The authors declare to have no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.116401.

References

Glover, G.H., Li, T.Q., Ress, D., 2000. Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magn. Reson. Med. 44, 162–167. https://doi.org/10.1002/1522-2594(200007)44:1<162::AID-MRM23>3.0.CO;2-E.

Lévêque, Y., Schön, D., 2013. Listening to the human voice alters sensorimotor brain rhythms. PLoS One 8. https://doi.org/10.1371/journal.pone.0080659.

Andics, A., Farago, T., 2019. Voice perception across species. In: The Oxford Handbook of Voice Perception. Oxford University Press, Oxford, UK, pp. 363–412.

Andics, A., McQueen, J.M., Petersson, K.M., Gál, V., Rudas, G., Vidnyánszky, Z., 2010. Neural mechanisms for voice recognition. Neuroimage 52, 1528–1540. https://doi.org/10.1016/j.neuroimage.2010.05.048.

Andics, A., Gácsi, M., Faragó, T., Kis, A., Miklósi, Á., 2014. Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. Curr. Biol. 24, 574–578. https://doi.org/10.1016/j.cub.2014.01.058.

Baba, H., Tsukano, H., Hishida, R., Takahashi, K., Horii, A., Takahashi, S., Shibuki, K., 2016. Auditory cortical field coding long-lasting tonal offsets in mice. Sci. Rep. 6, 34421. https://doi.org/10.1038/srep34421.

Bandettini, P.A., Bullmore, E., 2008. Endogenous oscillations and networks in functional magnetic resonance imaging. Hum. Brain Mapp. 29, 737–739. https://doi.org/10.1002/hbm.20607.

Bathellier, B., Ushakova, L., Rumpel, S., 2012. Discrete neocortical dynamics predict behavioral categorization of sounds. Neuron 76, 435–449. https://doi.org/10.1016/j.neuron.2012.07.008.

Belin, P., Bestelmeyer, P.E.G., Latinus, M., Watson, R., 2011. Understanding voice perception. Br. J. Psychol. 102, 711–725. https://doi.org/10.1111/j.2044-8295.2011.02041.x.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000a. Voice-selective areas in human auditory cortex. Nature 403, 309–312.

Bestelmeyer, P.E.G., Belin, P., Grosbras, M.H., 2011. Right temporal TMS impairs voice detection. Curr. Biol. 21, R838–R839. https://doi.org/10.1016/j.cub.2011.08.046.

Buzsáki, G., Draguhn, A., 2004. Neuronal oscillations in cortical networks. Science 304, 1926–1929. https://doi.org/10.1126/science.1099745.

Capilla, A., Belin, P., Gross, J., 2013. The early spatio-temporal correlates and task independence of cerebral voice processing studied with MEG. Cerebr. Cortex 23, 1388–1395. https://doi.org/10.1093/cercor/bhs119.

Charest, I., Pernet, C.R., Rousselet, G. a, Quiñones, I., Latinus, M., Fillion-Bilodeau, S., Chartrand, J.-P., Belin, P., 2009. Electrophysiological evidence for an early processing of human voices. BMC Neurosci. 10, 127. https://doi.org/10.1186/1471-2202-10-127.

Chen, J.E., Glover, G.H., 2015. BOLD fractional contribution to resting-state functional connectivity above 0.1Hz. Neuroimage 107, 207–218. https://doi.org/10.1016/j.neuroimage.2014.12.012.

Dricu, M., Ceravolo, L., Grandjean, D., Frühholz, S., 2017. Biased and unbiased perceptual decision-making on vocal emotions. Sci. Rep. 7 https://doi.org/10.1038/s41598-017-16594-w.

Feinberg, D.A., Moeller, S., Smith, S.M., Auerbach, E., Ramanna, S., Glasser, M.F., Miller, K.L., Ugurbil, K., Yacoub, E., 2010. Multiplexed echo planar imaging for sub-second whole brain fmri and fast diffusion imaging. PLoS One 5, e15710. https://doi.org/10.1371/journal.pone.0015710.

Fox, M.D., Raichle, M.E., 2007. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat. Rev. Neurosci. https://doi.org/10.1038/nrn2201.

Frühholz, S., Belin, P., 2019. The science of voice perception. In: The Oxford Handbook of Voice Perception. Oxford University Press, Oxford, UK, pp. 3–14.

Fruhholz, S., Trost, W., Grandjean, D., 2016a. Whispering - the hidden side of auditory communication. Neuroimage 142, 602–612. https://doi.org/10.1016/j.neuroimage.2016.08.023.

Fruhholz, S., van der Zwaag, W., Saenz, M., Belin, P., Schobert, A.K., Vuilleumier, P., Grandjean, D., 2016b. Neural decoding of discriminative auditory object features depends on their socio-affective valence. Soc. Cogn. Affect. Neurosci. 11, 1638–1649. https://doi.org/10.1093/scan/nsw066.

Galazyuk, A.V., Feng, A.S., 1997. Encoding of sound duration by neurons in the auditory cortex of the little brown bat, Myotis lucifugus. J. Comp. Physiol. A Sens. Neural Behav. Physiol. 180, 301–311. https://doi.org/10.1007/s003590050050.

Giraud, A.-L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15, 511–517. https://doi.org/10.1038/nn.3063.

Greicius, M., 2009. Resting-state functional connectivity in neuropsychiatric disorders. Curr. Opin. Neurol. 24, 424–430. https://doi.org/10.1097/wco.0b013e328306f2c5.

Gutschalk, A., Patterson, R.D., Rupp, A., Uppenkamp, S., Scherg, M., 2002. Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. Neuroimage 15, 207–216. https://doi.org/10.1006/nimg.2001.0949.

Gygi, B., Kidd, G.R., Watson, C.S., 2004. Spectral-temporal factors in the identification of environmental sounds. J. Acoust. Soc. Am. https://doi.org/10.1121/1.1635840.

He, B.J., Snyder, A.Z., Zempel, J.M., Smyth, M.D., Raichle, M.E., 2008. Electrophysiological correlates of the brain's intrinsic large-scale functional architecture. Proc. Natl. Acad. Sci. 105, 16039–16044. https://doi.org/10.1073/pnas.0807010105.

Hullett, P.W., Hamilton, L.S., Mesgarani, N., Schreiner, C.E., Chang, E.F., 2016. Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. J. Neurosci. 36, 2014–2026. https://doi.org/10.1523/JNEUROSCI.1779-15.2016.

Jiang, X., Chevillet, M.A., Rauschecker, J.P., Riesenhuber, M., 2018. Training humans to categorize monkey calls: auditory feature- and category-selective neural tuning changes. Neuron 98, 405–416. https://doi.org/10.1016/j.neuron.2018.03.014 e4.

Lakatos, P., 2005. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J. Neurophysiol. 94, 1904–1911. https://doi.org/10.1152/jn.00263.2005.

Lehmann, C., Herdener, M., Schneider, P., Federspiel, A., Bach, D.R., Esposito, F., di Salle, F., Scheffler, K., Kretz, R., Dierks, T., Seifritz, E., 2007. Dissociated lateralization of transient and sustained blood oxygen level-dependent signal components in human primary auditory cortex. Neuroimage 34, 1637–1642. https://doi.org/10.1016/j.neuroimage.2006.11.011.

Lewis, L.D., Setsompop, K., Rosen, B.R., Polimeni, J.R., 2016. Fast fMRI can detect oscillatory neural activity in humans. Proc. Natl. Acad. Sci. 113, E6679–E6685. https://doi.org/10.1073/pnas.1608117113.

Lin, I.F., Agus, T.R., Suied, C., Pressnitzer, D., Yamada, T., Komine, Y., Kato, N., Kashino, M., 2016. Fast response to human voices in autism. Sci. Rep. 6 https://doi.org/10.1038/srep26336.

Mesgarani, N., Chang, E.F., 2012. Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485, 233–236. https://doi.org/10.1038/nature11020.

Mitra, A., Kraft, A., Wright, P., Acland, B., Snyder, A.Z., Rosenthal, Z., Czerniewski, L., Bauer, A., Snyder, L., Culver, J., Lee, J.M., Raichle, M.E., 2018. Spontaneous infra-slow brain activity has unique spatiotemporal dynamics and laminar structure. Neuron 98, 297–305. https://doi.org/10.1016/j.neuron.2018.03.015 e6.

Monson, B.B., Hunter, E.J., Lotto, A.J., Story, B.H., 2014. The perceptual significance of high-frequency energy in the human voice. Front. Psychol. 5 https://doi.org/10.3389/fpsyg.2014.00587.

Monto, S., Palva, S., Voipio, J., Palva, J.M., 2008. Very slow EEG fluctuations predict the dynamics of stimulus detection and oscillation amplitudes in humans. J. Neurosci. 28, 8268–8272. https://doi.org/10.1523/JNEUROSCI.1910-08.2008.

Norman-Haignere, S., Kanwisher, N.G., McDermott, J.H., 2015. Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. Neuron 88, 1281–1296. https://doi.org/10.1016/j.neuron.2015.11.035.

O'Connell, M.N., Ross, D., Lakatos, P., McGinnis, T., Barczak, A., Schroeder, C.E., 2015. Multi-Scale entrainment of coupled neuronal oscillations in primary auditory cortex. Front. Hum. Neurosci. 9 https://doi.org/10.3389/fnhum.2015.00655.

Pan, W.J., Thompson, G.J., Magnuson, M.E., Jaeger, D., Keilholz, S., 2013. Infraslow LFP correlates to resting-state fMRI BOLD signals. Neuroimage 74, 288–297. https://doi.org/10.1016/j.neuroimage.2013.02.035.

Pannese, A., Grandjean, D., Frühholz, S., 2016. Amygdala and auditory cortex exhibit distinct sensitivity to relevant acoustic features of auditory emotions. Cortex 85. https://doi.org/10.1016/j.cortex.2016.10.013.

Pantev, C., Eulitz, C., Elbert, T., Hoke, M., 1994. The auditory evoked sustained field: origin and frequency dependence. Electroencephalogr. Clin. Neurophysiol. 90, 82–90. https://doi.org/10.1016/0013-4694(94)90115-5.

Pantev, C., Eulitz, C., Hampson, S., Ross, B., Roberts, L.E., 1996. The auditory evoked "off" response: sources and comparison with the "on" and the "sustained" responses. Ear Hear. 17, 255–265. https://doi.org/10.1097/00003446-199606000-00008.

Pernet, C.R., McAleer, P., Latinus, M., Gorgolewski, K.J., Charest, I., Bestelmeyer, P.E., Watson, R.H., Fleming, D., Crabbe, F., Valdes-Sosa, M., Belin, P., 2015a. The human voice areas: spatial organization and inter-individual variability in temporal and extra-temporal cortices. Neuroimage 119, 164–174. https://doi.org/10.1016/j.neuroimage.2015.06.050.

Petkov, C.I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., Logothetis, N.K., 2008. A voice region in the monkey brain. Nat. Neurosci. 14, 211–212. https://doi.org/10.1038/nn2043.

Pina, J.E., Bodner, M., Ermentrout, B., 2018. Oscillations in working memory and neural binding: a mechanism for multiple memories and their interactions. PLoS Comput. Biol. 14 https://doi.org/10.1371/journal.pcbi.1006517.

Rac-Lubashevsky, R., Kessler, Y., 2018. Oscillatory correlates of control over working memory gating and updating: an EEG study using the reference-back paradigm. J. Cogn. Neurosci. 30, 1870–1882. https://doi.org/10.1162/jocn_a_01326.

Scholl, B., Gao, X., Wehr, M., 2010. Nonoverlapping sets of synapses drive on responses and off responses in auditory cortex. Neuron 65, 412–421. https://doi.org/10.1016/j.neuron.2010.01.020.

Scholtens, L.H., de Reus, M.A., de Lange, S.C., Schmidt, R., van den Heuvel, M.P., 2016. An MRI von economo - Koskinas atlas. Neuroimage.

Seifritz, E., Esposito, F., Hennel, F., Mustovic, H., Neuhoff, J.G., Bilecen, D., Tedeschi, G., Scheffler, K., Di Salle, F., 2002. Spatiotemporal pattern of neural processing in the human auditory cortex. Science 297, 1706–1708. https://doi.org/10.1126/science.1074355.

Trost, W., Fruhholz, S., Schon, D., Labbe, C., Pichon, S., Grandjean, D., Vuilleumier, P., 2014. Getting the beat: entrainment of brain activity by musical rhythm and pleasantness. Neuroimage 103, 55–64. https://doi.org/10.1016/j.neuroimage.2014.09.009.

Wang, X., Lu, T., Snider, R.K., Liang, L., 2005. Sustained firing in auditory cortex evoked by preferred stimuli. Nature 435, 341–346. https://doi.org/10.1038/nature03565.

Wang, Y.F., Liu, F., Long, Z.L., Duan, X.J., Cui, Q., Yan, J.H., Chen, H.F., 2014. Steady-state BOLD response modulates low frequency neural oscillations. Sci. Rep. https://doi.org/10.1038/srep07376.

Woolley, S.M.N., Fremouw, T.E., Hsu, A., Theunissen, F.E., 2005. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. Nat. Neurosci. 8, 1371–1379. https://doi.org/10.1038/nn1536.

Yang, Z., LaConte, S., Weng, X., Hu, X., 2008. Ranking and averaging independent component analysis by reproducibility (RAICAR). Hum. Brain Mapp. 29, 711–725. https://doi.org/10.1002/hbm.20432.

Zaske, R., Volberg, G., Kovacs, G., Schweinberger, S.R., 2014. Electrophysiological correlates of voice learning and recognition. J. Neurosci. 34, 10821–10831. https://doi.org/10.1523/JNEUROSCI.0581-14.2014.

Zuo, X.N., Di Martino, A., Kelly, C., Shehzad, Z.E., Gee, D.G., Klein, D.F., Castellanos, F.X., Biswal, B.B., Milham, M.P., 2010. The oscillating brain: complex and reliable. Neuroimage 49, 1432–1445. https://doi.org/10.1016/j.neuroimage.2009.09.037.