HHS Public Access

Author manuscript

Eur J Neurosci. Author manuscript; available in PMC 2021 July 01.

Published in final edited form as:

Eur J Neurosci. 2020 July ; 52(2): 2889–2904. doi:10.1111/ejn.14707.

Modulation change detection in human auditory cortex: evidence for asymmetric, nonlinear edge detection

Seung-Goo Kim[1,*], **David Poeppel**[2,3,4], **Tobias Overath**[1,5,6,*]

[1]Department of Psychology and Neuroscience, Duke University, Durham, North Carolina, United States

[2]Department of Psychology, New York University, New York, United States

[3]Center for Neural Science, New York University, New York, United States

[4]Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany

[5]Duke Institute for Brain Sciences, Duke University, Durham, North Carolina, United States

[6]Center for Cognitive Neuroscience, Duke University, Durham, North Carolina, United States

Abstract

Changes in modulation rate are important cues for parsing acoustic signals, such as speech. We parametrically controlled modulation rate via the correlation coefficient (*r*) of amplitude spectra across fixed frequency channels between adjacent time frames: broadband modulation spectra are biased towards slow modulate rates with increasing *r*, and vice versa. By concatenating segments with different *r*, acoustic changes of various directions (e.g. changes from low to high correlation coefficients, that is, random-to-correlated, or vice versa) and sizes (e.g. changes from low to high, or from medium to high correlation coefficients) can be obtained. Participants listened to sound blocks and detected changes in correlation while MEG was recorded. *Evoked responses* to changes in correlation demonstrated 1) an asymmetric representation of change direction: random-to-correlated changes produced a prominent evoked field around 180 ms, while correlated-to-random changes evoked an earlier response with peaks at around 70 ms and 120 ms, whose topographies resemble those of the canonical P50m and N100m responses, respectively; and 2) a highly non-linear representation of correlation structure, whereby even small changes involving segments with a high correlation coefficient were much more salient than relatively large changes that did not involve segments with high correlation coefficients. *Induced responses* revealed phase tracking in the delta and theta frequency bands for the high correlation stimuli. The results confirm a high sensitivity for low modulation rates in human auditory cortex, both in terms of their representation and their segregation from other modulation rates.

[*]**Corresponding authors:**Dr. Seung-Goo Kim, 308 Research Drive, Duke Box 90999 Duke University, Durham, North Carolina 27708, USA, +1 919 660 0719, solleo@gmail.com, Dr. Tobias Overath, 308 Research Drive, Duke Box 90999 Duke University, Durham, North Carolina 27708, USA, +1 919 684 6146, t.overath@duke.edu.

Author Contributions

DP and TO conceived the experiment; TO created stimuli and collected data; SGK and TO analyzed data; SGK, DP, and TO wrote the manuscript.

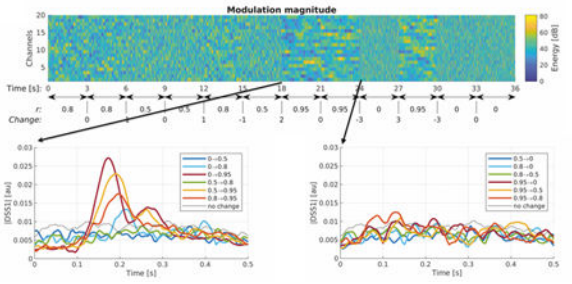
Data Accessibility

The data that support the findings of this study are available from the corresponding authors (SGK and TO) upon request.

Competing Interest The authors declare that there is no conflict of interest.

Graphical Abstract

Changes in modulation rate are important cues for parsing acoustic signals, such as speech. We parametrically controlled broadband modulation rates and recorded neural responses to changes in broadband modulation rate using MEG. We find asymmetric and non-linear evoked responses to various modulation rate changes, with particularly high sensitivity for changes that involve low modulation rates.



Keywords

Amplitude modulation; spectrotemporal modulation; temporal edge detection; magnetoencephalography

Introduction

The detection of changes in the acoustic environment is one of the most important tasks of the auditory system. Temporal modulations are one of the most prominent features of natural sounds; as such, different modulation rates can be powerful segmentation cues for acoustic signals, enabling listeners to extract relevant parts of the acoustic information. For example, different aspects of speech signals can be attributed to specific modulation rates: the rate of syllables is about an order of magnitude slower (~4–7 Hz) than that of some phonemes (~30–50 Hz) (Pickett, 1999). However, little is known about the distinct neural mechanisms that code transitions between modulation rates at the level of cortex (e.g. their temporal, topographic, and anatomic characteristics). Furthermore, the representation of distinct modulation rates is likely sub-served by different mechanisms than those segregating between modulation rates. That is, while there is evidence for periodotopic organization in the midbrain (Baumann *et al.*, 2011) and primary auditory cortex (Baumann *et al.*, 2015), segregating different modulation rates seems to recruit comparative mechanisms beyond primary auditory cortex (Chait *et al.*, 2007; Chait *et al.*, 2008; Overath *et al.*, 2010; Barascud *et al.*, 2016; Teng *et al.*, 2017; Teng *et al.*, 2018). Here we investigate the temporal neural correlates of changes between modulation rates in human auditory cortex using magnetoencephalography (MEG).

Past studies investigating auditory change detection have drawn on the large body of work on the mismatch negativity (MMN), which is the M/EEG evoked response to rare sounds (deviants) that are acoustically different from other, frequently presented sounds (standards). The acoustic attributes driving mismatch responses range from local stimulus features (e.g. the loudness of a deviant compared to a standard stimulus) to relatively abstract patterns

(e.g. sequences of words or tones that can function as standards or deviants) (e.g.van Zuijen *et al.*, 2004; Winkler *et al.*, 2006; Näätänen *et al.*, 2007).

A slightly different approach builds on the assumption that the auditory system represents local stimulus statistics and distinguishes between the emergence or disappearance of 'objects' in an auditory stream. Previous studies have demonstrated distinct psychoacoustic processing asymmetries, both in terms of discrimination between acoustic tokens (Cusack *et al.*, 2004), as well as in terms of the neural correlates of acoustic changes (Chait *et al.*, 2007; Chait *et al.*, 2008). For example, previous studies (Chait *et al.*, 2007; Chait *et al.*, 2008; Barascud *et al.*, 2016; Southwell *et al.*, 2017) showed that order-to-disorder (i.e. disappearing pattern) and disorder-to-order (emerging pattern) transitions within rapid streams of brief tone pips have distinct temporal and topographic characteristics. More recently, it was found that repeated patterns of tone pips induce tonic (sustained) neural activity explaining asymmetrical transitional effects (Barascud *et al.*, 2016; Southwell *et al.*, 2017). Similarly, transitions in the coherence of acoustic textures, another form of a transient or acoustic edge, are more salient when the transition reflects a change to more coherence than vice versa (Overath *et al.*, 2010).

In addition to the above studies, several studies have demonstrated a compelling sensitivity to different modulation rates at the level of human auditory cortex across different techniques (Giraud *et al.*, 2000; Harms & Melcher, 2003; Boemio *et al.*, 2005; Harms *et al.*, 2005; Overath *et al.*, 2008; Overath *et al.*, 2012; Wang *et al.*, 2012; Teng *et al.*, 2017). This is similar in non-human species, where the preferred rate for which phase-locking or synchronized firing occurs decreases as one ascends the neuraxis (Lu *et al.*, 2001b; Bartlett & Wang, 2007; Bendor & Wang, 2007).

The analysis of induced neural responses, such as phase coherence (Luo & Poeppel, 2007; Howard & Poeppel, 2010), has proven a promising tool within the M/EEG literature to disambiguate the representation of individual acoustic signals at the neural level; this is beyond the scope of more classical evoked response analyses, which rely on the averaging of multiple presentations of a given stimulus or stimulus class (thereby neglecting phase information). By virtue of its nature, amplitude modulations are ideally suited for investigating neural phase tracking, or neural entrainment (Galambos *et al.*, 1981; Ross *et al.*, 2005; Wang *et al.*, 2012). Thus, phase coherence, or phase tracking, can be used as a stimulus-specific neural marker that distinguishes between signals of different modulation rates (Gross, 2014).

Experimental designs that allow both analysis techniques (evoked and induced response analyses) enable the investigation of both the neural representation and segregation of specific aspects of acoustic signals, such as modulation rate. To this end, the current MEG study employed a design similar to that of Overath *et al.* (2010), whereby changes in modulation rate (assessed via complex correlation structure, as in Overath et al., 2008) can be investigated via evoked responses, while the representation of individual modulation rates can be elucidated via response analyses such as phase coherence.

We hypothesized — based on the assumption that auditory cortex is sensitive to ongoing stimulus statistics — that we would observe similar processing asymmetries as described previously (Chait *et al.*, 2007; Chait *et al.*, 2008; Overath *et al.*, 2010; Barascud *et al.*, 2016), such that transitions to lower correlation coefficients (correlated-to-random) would reveal different response characteristics from transitions to higher correlation coefficients (random-to-correlated). Furthermore, we expected to observe a dissociation with respect to phase coherence, such that sounds with high correlation coefficients, i.e. slow temporal modulations, would reveal strong phase coherence in low frequency bands (e.g. delta and theta bands), while the processing of sounds with low correlation coefficients, i.e. fast temporal modulations, would be characterized by phase coherence in higher frequency bands (e.g. the low gamma band).

Materials & Methods

Participants

Sixteen participants (2 left-handed, mean age: 23, range: 18–33, 8 females) took part in the study. Data from three participants had to be excluded because of excessive data artifacts or chance task performance during the MEG recording, leaving a total of 13 participants (2 left-handed, mean age: 24, range: 18–33, 5 females) for the main analysis. Participants provided written informed consent in accordance with the New York University Committee on Activities involving Human Subjects.

Stimuli

The stimuli were based on Overath et al. (Overath *et al.*, 2008), and a visual example is depicted in Figure 1. All stimuli were created digitally using MATLAB (RRID:SCR_001622; http://www.mathworks.com) software at a sampling rate of 44.1 kHz and 16-bit resolution. Each sound consisted of 20 sinusoids pseudorandomly chosen from a pool of 101 logarithmically spaced frequencies between 246–4435 Hz. The particular parameters were chosen so as to approximate respective features in naturally occurring sounds, which typically have complex spectra with multiple frequencies present. The passband (246–4435 Hz) covers the acoustic range of maximal acoustic sensitivity in the human auditory system, and the number and spacing of frequencies within this pool are a result of this range.

The amplitude spectrum was defined for each 20 ms frame such that the correlation from one frame to the next was operationalized as the Pearson correlation r:

$$r(\mathrm{x,y}) = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \bar{x}) - (y_i - \bar{y})}{S_x S_y}$$

where x is a vector with the amplitude values (in dB) of the 20 frequency components in a frame, y is a vector with amplitude values (in dB) of the 20 frequency components in the consecutive frame, n is the number of frequencies, s_x and s_y represent the standard deviations of x and y, and $\bar{x}$ and $\bar{y}$ are the arithmetic means of x and y, respectively. Thus, the amplitude spectrum of a given sound varied with a specified correlation coefficient r

between the 20 ms segments. Importantly, the mean amplitude (65 dB) and standard deviation (SD = 15) were identical for each frequency component in a given sound and across correlation levels. Linear spline interpolation amplitude transitions were applied between frames in the time domain so that sounds were continuous and did not have any sudden amplitude transients. This was applied in order to render the sounds more similar to most ethological sounds; however, some speech sounds like plosives or consonant-vowel transitions do display amplitude discontinuities (Rogers, 2000).

Note that modulation rate here is not defined in its more common form (i.e. in terms of the rate of sinusoidal amplitude modulation of a single carrier), but with respect to the overall amplitude modulation across multiple carriers, as specified by the Pearson correlation of amplitude spectra between successive 20 ms frames. Therefore, the correlation coefficient is related to the spectra of the modulation rates (Figure 1D): the higher the correlation coefficient, the more the spectrum of the envelope is biased towards to the slower modulation rates, and vice versa.

We used four different values for $r = 0, 0.5, 0.8, 0.95$. These are based on their roughly linear relationship with respect to (i) Fisher's z-transform (Fisher, 1915) of the parameter values ($Z(r) = 0, 0.55, 1.1, 1.83$) and (ii) log-log slopes of their corresponding envelope modulation spectra (1–20 Hz; −0.07, 0.34, −0.60, −0.84). We predicted that this linear relationship (roughly equal step size between r-values) would result in roughly linearly spaced behavioral and neural response characteristics.

For a sound with a set correlation between adjacent time frames (e.g. $r = 0.95$), the correlation between a given frame and subsequent frames decays exponentially with the number of time frames (or lag). The window length of this decay process is defined as the duration (in ms) at which the correlation between any two time-frames reaches a minimum reference value ($r = 0.2$), which was determined for its indistinguishability from $r = 0$ as found in our previous study (Overath *et al.*, 2008), from its initial value r between adjacent time frames. The window length is relevant to a "temporal integration window" of non-primary auditory neurons proposed in Poeppel (2003): a neuron should only respond to sounds whose duration of correlation matches or surpasses its temporal integration window. The window length is calculated by the following equation:

$$w(0.2) = d\frac{\ln(0.2)}{\ln(|r|)},$$

where w is window length for a reference value of $r = 0.2$, d is frame duration, and ln is the natural logarithm to the base e. For the parameters used in the MEG study ($r = 0, 0.5, 0.8, 0.95$), the corresponding $w(0.2)$ are 0 ms, 46 ms, 144 ms, and 628 ms.

Audio files of exemplar stimuli are available at https://osf.io/frdqb/.

Experimental design

Prior to the MEG recordings, participants were familiarized with the stimuli and then performed two-interval two-alternative forced-choice (2I2AFC) psychophysics with $r = 0$ as

the reference sound and one of six target sounds (r = 0, 0.2, 0.4, 0.6, 0.8, 0.95). The target and reference sounds were sequentially presented in a random order. Participants were to indicate which of the two sounds was more correlated or had slower amplitude changes. Stimuli were 2 seconds long and were different exemplars from the ones subsequently used in the MEG experiment. Psychophysics ensured that participants were able to distinguish a highly correlated sound from the reference sound; participants needed to reach at least 90% correct performance for the strongest correlation (r = 0.95) to be included in the MEG study.

The MEG study used four levels of correlation: r = 0, 0.5, 0.8, 0.95. The four levels yielded four possible absolute change sizes between adjacent sound segments:

Change0: 0↔0, 0.5↔0.5, 0.8↔0.8, 0.95↔0.95

Change1: 0↔0.5, 0.5↔0.8, 0.8↔0.95

Change2: 0↔0.8, 0.5↔0.95

Change3: 0↔0.95

The bidirectional arrows indicate that, for example, a change from r = 0 to r = 0.95 has the same absolute change size as a change from r = 0.95 to r = 0.

So as to enable an investigation of the representation of correlation structure via phase coherence (Luo & Poeppel, 2007), each level of correlation was represented by 3 unique correlation structures (i.e. 3 unique exemplars for each of the 4 levels). This resulted in 12 (3 exemplars × 4 levels) unique correlation structures with 150 frames, where each corresponding sound segment was 3 s long (150 frames with 20 ms frame lengths). The 12 unique correlation structures were concatenated in a pseudo-randomized order to yield a 36 s long sound block (see Figure 1).

There were 40 sound blocks in total. In each of four sessions, 10 sound blocks were presented in a pseudo-randomized order. The frequency composition of each sound block was selected randomly but was unique and stayed fixed within a sound block. Thus, each correlation structure was presented 40 times overall (once in each sound block), but each time within the context of a different frequency composition.

The task of participants was to press a response button whenever they perceived a change in correlation, irrespective of the direction of that change (i.e. whether it was from less correlated to more correlated, or vice versa). A time window for a valid response was defined as 100–1500 ms after transition. Furthermore, to minimize eye-related artifacts during trials, participants were asked to close their eyes during the sound blocks, open them during the 10-sec gaps between sound blocks, and close them in time for the beginning of the next sound block.

Data acquisition and analysis

During the MEG recording, the stimuli were delivered using Etymotic ER3A insert earphones calibrated to have a flat frequency response in the MEG set-up. The signals were

presented at a loudness level of approximately 75 dB. Data were acquired with a 160-channel whole-head MEG system (KIT, Kanazawa, Japan). Three reference channels recorded environmental noise, while the remaining 157 channels recorded neuromagnetic activity. The data were sampled at 1000 Hz and filtered online using a 1–200 Hz passband filter with a notch at 60 Hz.

Preprocessing

The external noise in the data was suppressed by removing principle components that are correlated with reference channels using the time-shift principal component analysis algorithm (TSPCA) (de Cheveigné & Simon, 2007). Subsequently, sensor-specific noise (i.e., uncorrelated with neighboring sensors) was suppressed using sensor noise suppression (SNS) with 10 neighbors defined (de Cheveigné & Simon, 2008).

To suppress physiological artifacts, independent component analysis (ICA) was applied to the whole data after an offline bandpass filtering of 1–60 Hz using a two-pass 4^{th}-order Butterworth infinite impulse response (IIR) filter and down-sampling at 200 Hz using FieldTrip (RRID: SCR_004849; http://www.fieldtriptoolbox.org/) (Oostenveld *et al.*, 2011). ICs with stereotypical topology, spectrum, and peak-aligned responses that are associated to cardiac artifacts (mostly two) and eye movement artifacts (mostly two to three) were manually labeled and the whole data was reconstructed without those ICs.

Evoked response analysis

To extract a multivariate component that is most reliably evoked by the stimuli, we used denoising source separation (DSS) (de Cheveigne & Simon, 2008). For each epoch we first subtracted the mean of the baseline period [−.2 0] sec. Then, only the post-stimulus epoch [0, 0.5] sec was fed into the DSS algorithm for each condition (e.g., 0 → 0, 0 → 0.5, 0 → 0.8, 0 → 0.95, …). Only the first DSS component (the most prominently evoked, “DSS1”) was used in the further analysis. Since the sign (or polarity) of DSS components is determined arbitrarily in conjunction with their spatial weights, the sign of a given DSS component needs to be adjusted for meaningful comparisons across conditions and participants. This is typically achieved by taking an averaged time-locked MEG response as a reference and flipping signs of DSS components to maximize correlation coefficients with the reference. However, certain transitions (e.g., 0 → 0.5) evoked no discernible peaks in some subjects, preventing the meaningful choice of a fixed sign. Thus, we instead took the absolute value of the DSS time course, which allows a comparison across conditions and participants. This is similar to root-mean-square (RMS) approaches for MEG timeseries data (Ding *et al.*, 2015).

Induced response analysis

We used two related but complementary measures of phase coherence across trials: cross-trial phase coherence (CTPC) and inter-trial correlation (ITC).

CTPC at time t and frequency f is given by:

$$\mathrm{CTPC}(t,f)=\left(\frac{1}{N}\sum_{n=1}^{N}\cos\theta(n,t,f)\right)^2+\left(\frac{1}{N}\sum_{n=1}^{N}\sin\theta(n,t,f)\right)^2$$

where $\theta(n,t,f)$is a phase angle at time t, frequency f, trial n and N is the total number of trials (Luo & Poeppel, 2007). When all phases are consistent across trials, the CTPC is one. When all phases are completely random, the CTPC is zero. To estimate phases, a Morlet wavelet transform was applied to epochs between [−1, 4] sec with a 50-ms step over 24 loglinear frequency bins from 1 to 60 Hz. The number of cycles (which controls Gaussian taper width) linearly increased from 3 to 11 to control temporal smoothness across frequency bins. We computed CTPC within each exemplar and averaged for each correlation level (i.e., within-CTPC). We also computed CTPC across different exemplars of a given correlation level (i.e., across-CTPC) to estimate chance level CTPC. To match the number of trials, an equal number of trials was drawn from each exemplar for 100 times then averaged.

ITC is given by the average Pearson correlation of filtered signals (Ding & Simon, 2013) as:

$$ITC(f)=\frac{1}{N(N-1)}\sum_{i\neq j}corr\left(y_{i,f},y_{j,f}\right)$$

where f is a frequency bin, N is the total number of trials, corr is Pearson product, and $y_{i,f}$ is the i-th trial MEG signal filtered at the frequency bin f. For five frequency bands (δ, 1–3 Hz; θ, 4–7 Hz; α, 8–12 Hz; β, 13–29 Hz; γ, 30–60 Hz), a sinc-windowed finite impulse response filter was applied to the whole data to avoid edge artifacts. Similar to CTPC, within-ITC was computed from trials of the same exemplar and across-ITC was computed from trials with different exemplars (but same correlation level). ITC measures phase coherence across trials similarly to CTPC, but since it does not use Fourier transform, it is computationally more efficient.

Statistical inference

For behavioral measures and evoked responses, we tested the effect of conditions and transitions using different two-way repeated-measures analysis of variance (ANOVA) models. Specifically, in the first model, which is faithful to the 4-by-4 design, we modeled the *Previous Level* and the *Current Level* as two within factors. The main effect of the Previous Level is an averaged difference in evoked responses across conditions with different previous correlation levels (i.e., differences between 0 → 0.95 vs. 0.5 → 0.95, vs. 0.8 → 0.95, …) whereas the main effect of the Current Level is an averaged difference across conditions with different current correlation levels (i.e., differences between 0 → 0 vs. 0 → 0.5, vs. 0 → 0.8, …).

In the second model, which is based on our assumption that the detection mechanism is sensitive to amount of change in r, we modeled the *Direction* (positive or negative) and *Absolute Step Size* (1, 2, or 3) of changes as two within factors. For this, behavioral and neuronal measures in the 4-by-4 cells were averaged according to the change direction and step size to fit into a 2-by-3 matrix. For example, a measure for the first cell (“Positive” x

"1") is an average of three change conditions (0 → 0.5, 0.5 → 0.8, and 0.8 → 0.95). Measures from Change0 conditions without changes (e.g., 0 → 0, 0.5 → 0.5) were not included in this model. It should be noted that averaging across conditions here is only intended to compare conditions diagonally (in the 4-by-4 design matrix), which is not possible in the first model. We will discuss differences between conditions with the same Direction and Absolute Step Size later.

For induced responses, we used one-sample T-tests to test if phase tracking (CTPC) is significant in the respective conditions and a one-way repeated-measures ANOVA with the Current Level as a within-subjects factor to test if phase tracking is different across correlation levels.

To perform statistical inference with multiple comparison correction based on cluster-based permutation (Maris & Oostenveld, 2007), we used Minimum Norm Estimation (MNE)-Python (RRID:SCR_005972; https://martinos.org/mne/) (Gramfort *et al.*, 2013). 100,000 randomizations (or full permutations, whichever was smaller) were performed to compute p-values. A cluster-forming threshold was determined based on a parametric, uncorrected p-value of 0.001 and extent-threshold of 10 ms. Family-wise significance level was set at $p < 0.05$, and Bonferroni correction was further applied where necessary.

Results

Behavioral results

The behavioral data in the MEG study were analyzed in terms of d' performance (Macmillan & Kaplan, 1985), where responses to Change0 transitions served as the false alarm rate. The data revealed that changes involving the highest correlation coefficient ($r = 0.95$) were most salient (Figure 2). This was confirmed by a repeated-measures ANOVA model on 12 measures (off-diagonal elements only) with factors of Previous Level and Current Level. Main effects of Previous Level ($F[1.06,\ 12.70] = 67.82,\ \eta_p^2 = 0.849,\ p < 10^{-14}$) and Current Levels ($F[1.06,\ 12.70] = 117.54,\ \eta_p^2 = 0.907,\ p < 10^{-17}$) were highly significant after Greenhouse-Geisser correction, and they were driven by $r = 0.95$ (Bonferroni-corrected $p < 10^{-5}$ for all cases of post-hoc pairwise comparison for the main effects). The interaction between Previous Level and Current Level was also significant ($F[3.17,\ 38.12] = 283.76,\ \eta_p^2 = 0.959,\ p < 10^{-70}$), which was due to differences between conditions with the same Current Level but different Previous Levels, or vice versa.

Furthermore, the average d' in the 4-by-4 matrix was symmetric to the diagonal, suggesting little effect of the direction of changes. This was tested by a repeated measures ANOVA model with factors Direction (positive, negative) and Absolute Step Size (1, 2, 3). There was no main effect of Direction ($F[0.42,\ 5.05] = 2.93,\ \eta_p^2 = 0.196,\ p < 0.113$), but a significant main effect of Absolute Step Size ($F[0.84,\ 10.09] = 586.60,\ \eta_p^2 = 0.979,\ p < 10^{-12}$) and interaction ($F[0.84,\ 10.09] = 8.96,\ \eta_p^2 = 0.427,\ p = 0.002$). Post-hoc comparisons revealed that this was driven by step sizes involving $r = 0.95$. Thus, contrary to our prediction, transitions with the same nominal step size (e.g., 0 → 0.5 vs. 0.8 → 0.95) did not have the same

behavioral salience. One-way repeated-measures ANOVA models that were tested on each step size revealed that there were significant differences within each set of conditions with the same nominal step size ($F[1.40,\ 21.10] = 96.99,\ \eta_p^2 = 0.866,\ p < 10^{-13}$ for Step size = −1; $F[1.54,\ 23.23] = 60.48,\ \eta_p^2 = 0.930,\ p < 10^{-13}$ for Step size = +1; $F[1,\ 15] = 30.33,\ \eta_p^2 = 0.803,\ p < 10^{-5}$ for Step size = −2; $F[1,\ 15] = 61.36,\ \eta_p^2 = 0.804,\ p < 10^{-5}$ for Step size = +2; all Bonferroni corrected).

Evoked MEG responses

Figure 3 shows the absolute time course of the DSS1 component for conditions with positive and negative changes in correlation, respectively. Similar to the behavioral results, only those positive changes involving $r = 0.95$ showed distinctive effects. However, in contrast to the behavioral results, negative changes involving $r = 0.95$ evoked less prominent and qualitatively different responses.

Separate two-way repeated measures ANOVAs modeling either the correlation levels (Previous Level x Current Level) or the transitions (Direction x Absolute Step) found significant main effects and interactions (Figure 4 and 5). The significant effect at around 150–200 ms was attributed to the prominent response to positive changes to $r = 0.95$, as revealed by post-hoc pairwise comparisons. This effect drove multiple significant effects including the main effect of Current Level, its interaction with the Previous Level, the positive main effect of Direction (the only possible change direction to $r = 0.95$ is positive), and the main effect of Absolute Step, and their interaction. The averaged projection of DSS1 components showed a topography that is similar to an M100 (Figure 4B and 5B). Also, a later effect at around 260 ms (Figure 4C), which was also driven by positive changes to $r = 0.95$ (especially by 0→0.95 and 0.5→0.95), showed a topography that is similar to an M200 (reversed polarity of an M100, suggesting opposite source orientations).

The other significant effects at around 70 ms and 125 ms were related to a negative change from $r = 0.95$ (main effect of Previous Level, negative main effect of Direction). The averaged projection of DSS1 components at respective time points showed a topography that is similar to an M50-M100 complex (Figure 4A and 5A).

Figure 3 also shows that, among the positive changes to $r = 0.95$ (i.e., 0 → 0.95, 0.5 → 0.9, 0.8 → 0.95), the response peak latencies decreased and the peak magnitudes increased with increasing step size; repeated measures linear models (modeling step size as a linear within-factor) confirmed that the decreasing latency and increasing peak magnitude relationships are significant ($F[0.83,\ 9.97] = 1823.70,\ \eta_p^2 = 0.993,\ p < 10^{-13}$ for latencies; $F[0.83,\ 9.97] = 125.04,\ \eta_p^2 = 0.912,\ p < 10^{-6}$ for magnitudes; Figure 6).

Induced MEG responses

The CTPC differences (across-CTPC subtracted from within-CTPC; phase dissimilarity plots) averaged across subjects are shown in Figure 7. Increased CTPC was found mostly in delta and theta bands around the auditory cortices at various time points. Averaged into the five frequency bands over the whole trial (i.e., [0, 3] sec), the CTPC differences were tested

against zero using a one-sample T-test. For the number of subjects (n = 13), full permutations of flipping signs (2^{13} = 8,192) were used to compute p-values. After Bonferroni correction for the number of conditions, significant clusters were found for r = 0.8 (θ-band, right auditory channels, $T[12] = 5.98$, corrected $p = 0.004$) and r = 0.95 (θ-band, left auditory cortex, $T[12] = 6.38$, corrected $p = 0.011$; δ-band, left auditory cortex, $T[12] = 5.94$, corrected $p = 0.049$; Figure 8). However, no significant difference across correlation levels was found when tested using a one-way repeated-measure ANOVA (uncorrected $p > 0.01$). As a follow-up analysis on the significant CTPC in the theta band, we tested a linear dependency of the CTPC difference on correlation levels averaged across all channels using a repeated-measures regression model; this revealed a significant increase of the CTPC difference over the correlation coefficients ($F[1,12] = 39.35$, $p < 10^{-4}$; Figure 9).

The ITC differences (across-ITC subtracted from within-ITC) averaged across subjects are shown in Figure 10. Similar to CTPC, positive ITC differences were found in channels around the auditory cortices in the delta and theta bands. But when tested with a one-sample T-test, no significant cluster was found for any time intervals (corrected $p > 0.05$). A one-way repeated-measure ANOVA did not reveal any effect of correlation level either (uncorrected $p > 0.01$).

Discussion

The present study investigated neural signatures for the segregation and representation of amplitude correlation structure, or complex temporal modulation rate, in acoustic signals. With respect to segregation, the results revealed asymmetric and non-linear responses to changes in the amplitude correlation structure of acoustic signals. Specifically, random-to-correlated and correlated-to-random transitions displayed distinct neural signatures: random-to-correlated transitions produced a prominent evoked response at around 180 ms, while correlated-to-random transitions evoked an earlier response at around 70 ms, which resemble a canonical P50m-N100m complex. Further, the random-to-correlated transitions demonstrated a highly non-linear neural response, whereby even relatively small correlation transitions to the largest correlation coefficient used (r = 0.95) were much more salient than relatively large correlation transitions that did not involve a segment with r = 0.95. The analysis of induced responses, reflecting the representation of correlation structure, revealed phase tracking in the theta and delta frequency bands for the high correlations (r = 0.8 and 0.95), but no significant phase tracking in the gamma band for stimuli with no correlation (r = 0).

Evoked responses

The results are in broad agreement with a number of studies showing high sensitivity to slow modulation rates in human auditory cortex (Giraud *et al.*, 2000; Harms & Melcher, 2003; Boemio *et al.*, 2005; Harms *et al.*, 2005; Overath *et al.*, 2008; Overath *et al.*, 2012; Wang *et al.*, 2012; Teng *et al.*, 2018). In the present study, modulation rate within a channel or frequency was controlled via amplitude correlations between adjacent time frames within long sounds; large correlation coefficients produced slow temporal variations, while

temporal modulation rate increased as the correlation coefficient decreased. Here, the sounds with the largest correlation coefficient (or slowest temporal modulations) were those that participants were most sensitive to.

In the present data, dissociable responses were observed for random-to-correlated and correlated-to-random transitions with respect to latency (70 ms vs. 180 ms) and topography (P50m-like vs. N100m-like). The topography of evoked responses suggests that the dominant neural sources are located near the superior temporal cortices. Based on the fact that response latency is indicative of processing stage along the auditory neuraxis (Krumbholz *et al.*, 2007; Chait *et al.*, 2008), the major source of the earlier response at 70 ms is likely in primary auditory cortex (e.g. medial Heschl's gyrus), whereas the later response at 180 ms is most likely generated in non-primary auditory cortices such as lateral Heschl's gyrus and planum temporale (Lütkenhöner & Steinsträter, 1998; Godey *et al.*, 2001; Yvert *et al.*, 2005). Critically, the differential latencies of the evoked responses suggest that distinctive underlying mechanisms are involved in the detection of the appearance and disappearance of highly correlated modulation structure, which can be regarded as a form of auditory object or auditory stream.

The N1m-like response to random-to-correlated transitions in the current study has previously been found in transitions from randomly varying to constant tones (Chait *et al.*, 2007), regularly alternating or repeating tone-pips (Chait *et al.*, 2008; Barascud *et al.*, 2016), or the appearance of frequency components in a complex auditory scene (Sohoglu & Chait, 2016a; b; Teki *et al.*, 2016). It has been suggested that the N1m-like response occurs for disorder-to-order type "temporal edges" (Chait *et al.*, 2008). In the current study, the bias of modulation spectra was perceived non-linearly: transitions to $r = 0.95$ evoked a strong N100m-like response, whereas transitions to $r = 0.8$ evoked no significant response (see Figure 4C). This suggests a distinctive sensitivity to slow modulations in the non-primary auditory cortex that segregates $r = 0.95$ from other conditions. Moreover, it also suggests that there may exist a threshold for determining 'temporal regularity' within spectrotemporal modulation structure, and that this threshold lies between $r = 0.8$ and $r = 0.95$. Future studies will need to determine the precise location of such a threshold for detecting temporal regularity.

The latency of the N1m-like change response was previously found to be a function of the time required to infer regularity within the acoustic stimulus features (Chait *et al.*, 2007). In our data, the peak latency for a transition to $r = 0.95$ increased with an increasing correlation level of the preceding segment (i.e., decreasing contrast). This effect seems to be related to the window length of exponential decay within the stimuli (see Materials and Methods): the longer the window length in the previous segment, the more time is needed to determine a change in modulation structure.

The P50m-like response to correlated-to-random transitions showed similar latency and topography to the reported response at around 50 ms after the regular-to-random or constant-to-random transitions (i.e., disappearance of regularity) in previous studies (Chait *et al.*, 2007; Chait *et al.*, 2008; Barascud *et al.*, 2016; Sohoglu & Chait, 2016a; b). It is possible that the P50m-like component is due to the sharp, sudden change in energy within a channel

at the onset of an r=0 stimulus segment (Hari *et al.*, 1987; Pantev *et al.*, 1996; Pratt *et al.*, 2008). However, a P50m-like response to correlated-to-uncorrelated transitions was also observed in the absence of sudden increases in energy of specific channels where the interaural correlation of wideband noise was manipulated (Chait *et al.*, 2005). Thus, the P50m-like response to the correlated-to-random transition in the current data seems more attributable to the disappearance of regularity of correlated acoustic structure (Winkler *et al.*, 2009).

Behavioral vs. neural measures

The behavioral data show that performance for random-to-correlated transitions was better than for correlated-to-random transitions, and the evoked responses showed an even more pronounced asymmetry (stronger response to the random-to-correlated transitions). This asymmetry in neural responses suggests the involvement of distinctive mechanisms for detecting the appearance or disappearance of correlated modulation structure.

Previously, using a similar paradigm, Overath et al. (2010) found that acoustic changes across which spectrotemporal coherence increased were behaviorally more salient than those changes across which coherence decreased, and that this perceptual asymmetry was reflected in stronger responses in posterior temporal regions bilaterally. Similarly, a study that manipulated interaural correlation of noise (Chait *et al.*, 2005) also showed asymmetrical behavioral performance (difficulty in detecting decorrelation). More recent studies reveal both behavioral and neural asymmetries in the detection of appearing or disappearing regularity of tone-pips in statistically regular or irregular contexts (Cervantes Constantino *et al.*, 2012; Sohoglu & Chait, 2016b; a). In particular, Cervantes Constantino *et al.* (2012) demonstrated through various manipulations that the detection of an appearance of certain frequency components in a complex auditory scene is seemingly automatic (near-perfect performance level that was barely affected by the number of objects in the auditory scene), while this was not the case for detecting the disappearance of certain frequency components. The authors noted that detecting a disappearance of regularity is more computationally demanding because it requires the constant matching of expected inputs and responding to unexpected inputs (Cervantes Constantino *et al.*, 2012). Indeed, the maintenance of sound statistics is associated with increased phasic activity in the superior temporal gyrus (Barascud *et al.*, 2016; Southwell *et al.*, 2017), which might reflect enhanced inhibition, or an increased gain for specific target features, or other cognitive processes related to learning and working memory (Southwell *et al.*, 2017).

We should note that our initial prediction regarding the perceptual equidistance between “steps” (e.g., 0 → 0.5, 0.5 → 0.95, 0.8 → 0.95, belonging to the same step size and direction) was not supported by the behavioral and neural data, instead suggesting that perceptual distances were not equal. However, particularly for the evoked neural responses, the effect of absolute step size was not completely explained by the involvement of the highest correlation in transitions as there was a significant linear effect of absolute step size within transitions to the highest correlation level (i.e., 0 → 0.95 vs. 0.5 → 0.95 vs. 0.8 → 0.95; Figure 6). Future studies will need to determine r-values that are perceptually equidistant.

Phase coherence

We found significant theta and delta phase coherence for slow modulations, but no significant difference between modulation rates. One potential reason for this finding is the nature of the stimulus. As shown by Howard and Poeppel (2010), one major determinant of phase tracking is sharp envelope transitions across frequency bands, as they occur in speech. Such strong transients are largely absent in the present stimulus, for which the amplitude envelopes between frequency bands often run counter-correlated, thus rendering the average between-channel correlation close to zero. Relatedly, in the spectral domain, the stimuli in the current study have envelope modulation spectra without prominent peaks. This would have blurred prominent entrainment in specific frequency bands. Nonetheless, there was a linear dependency in theta phase coherence over the correlation levels, presumably reflecting heightened sensitivity to slow modulations at the level of auditory cortex (Giraud *et al.*, 2000; Harms & Melcher, 2003; Boemio *et al.*, 2005; Harms *et al.*, 2005; Overath *et al.*, 2008; Overath *et al.*, 2012; Wang *et al.*, 2012; Teng *et al.*, 2018).

Contrary to our hypothesis, we did not find significant phase coherence in the gamma band for sounds with fast temporal modulations. This is in line with Overath *et al.* (2008), who, using a similar stimulus, found increased blood-oxygenation-level dependent (BOLD) signal in the superior temporal sulcus for slow temporal modulations, but no BOLD signal increase for fast modulations anywhere in auditory cortex. In non-human primate brains, it has been shown that “synchronized” neural populations explicitly encode slow modulations via temporal coding, whereas “non-synchronized” populations of primary auditory neurons implicitly represented fast modulations via average discharge rates (Lu *et al.*, 2001a; Wang *et al.*, 2003). In human auditory cortex, concurrent M/EEG data revealed a non-phase-locked mode for rapid temporal modulation (Tang *et al.*, 2016). It is conceivable that the current results reflect these different neural coding algorithms.

In conclusion, the current study provides new evidence for the asymmetric and non-linear detection of temporal ‘edges’ based on spectrotemporal modulation structure in the human auditory cortex, with a preferential sensitivity for the emergence of slow modulations. The data demonstrate that the regularity of a complex auditory scene can be constructed based on temporal modulation rates, extending the notion of regularity that defines temporal acoustic edges. The results suggest that distinct neural populations in the primary and non-primary auditory cortices are involved in the detection of an emergence or disappearance of slow spectrotemporal modulations, reflecting a general principle of pattern extraction and pattern matching in the auditory cortex.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health [grant R01DC05660] to DP. We thank the anonymous reviewers for their comments that helped us in improving the earlier version of the manuscript.

Abbreviations

2I2AFC	two-interval two-alternative forced-choice
ANOVA	analysis of variance
CTPC	cross-trial phase coherence
DSS	denoising source separation
ICA	independent component analysis
IIR	infinite impulse response
ISI	interstimulus interval
ITC	inter-trial correlation
M/EEG	magneto-/electro-encephalography
MEG	magnetoencephalography
MMN	mismatch negativity
SD	standard deviation
SNS	sensor noise suppression
TSPCA	time-shift principal component analysis algorithm

References

Barascud N, Pearce MT, Griffiths TD, Friston KJ & Chait M. (2016) Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. Proceedings of the National Academy of Sciences, 113, E616–E625.

Bartlett EL & Wang X. (2007) Neural representations of temporally-modulated signals in the auditory thalamus of awake primates. J. Neurophysiol, 97, 1005–1117. [PubMed: 17050830]

Baumann S, Griffiths TD, Sun L, Petkov CI, Thiele A. & Rees A. (2011) Orthogonal representation of sound dimensions in the primate midbrain. Nat. Neurosci, 14, 423–425. [PubMed: 21378972]

Baumann S, Joly O, Rees A, Petkov CI, Sun L, Thiele A. & Griffiths TD (2015) The topography of frequency and time representation in primate auditory cortices. Elife, 4, e03256.

Bendor D. & Wang X. (2007) Differential neural coding of acoustic flutter within primate auditory cortex. Nat. Neurosci, 10, 763–771. [PubMed: 17468752]

Boemio A, Fromm S, Braun A. & Poeppel D. (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat. Neurosci, 8, 389–395. [PubMed: 15723061]

Cervantes Constantino F, Pinggera L, Paranamana S, Kashino M. & Chait M. (2012) Detection of Appearing and Disappearing Objects in Complex Acoustic Scenes. PLoS One, 7, e46167.

Chait M, Poeppel D, de Cheveigné A. & Simon JZ (2005) Human Auditory Cortical Processing of Changes in Interaural Correlation. The Journal of Neuroscience, 25, 8518. [PubMed: 16162933]

Chait M, Poeppel D, de Cheveigné A. & Simon JZ (2007) Processing asymmetry of transitions between order and disorder in human auditory cortex. J. Neurosci, 27, 5207–5214. [PubMed: 17494707]

Chait M, Poeppel D. & Simon JZ (2008) Auditory temporal edge detection in human auditory cortex. Brain Res, 1213, 78–90. [PubMed: 18455707]

Cusack R, Deeks J, Aikman G. & Carlyon RP (2004) Effects of location, frequency region, and time course of selective attention on auditory scene analysis. J. Exp. Psychol. Hum. Percept. Perform, 30, 643–656. [PubMed: 15301615]

de Cheveigne A. & Simon JZ (2008) Denoising based on spatial filtering. J. Neurosci. Methods, 171, 331–339. [PubMed: 18471892]

de Cheveigné A. & Simon JZ (2007) Denoising based on time-shift PCA. J. Neurosci. Methods, 165, 297–305. [PubMed: 17624443]

de Cheveigné A. & Simon JZ (2008) Sensor noise suppression. J. Neurosci. Methods, 168, 195–202. [PubMed: 17963844]

Ding N, Melloni L, Zhang H, Tian X. & Poeppel D. (2015) Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci, 19, 158. [PubMed: 26642090]

Ding N. & Simon JZ (2013) Adaptive Temporal Encoding Leads to a Background-Insensitive Cortical Representation of Speech. The Journal of Neuroscience, 33, 5728–5735. [PubMed: 23536086]

Fisher RA (1915) Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. Biometrika, 10, 507–521.

Galambos R, Makeig S. & Talmachoff PJ (1981) A 40-Hz auditory potential recorded from the human scalp. Proceedings of the National Academy of Sciences, 78, 2643–2647.

Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R. & Kleinschmidt A. (2000) Representation of the temporal envelope of sounds in the human brain. J. Neurophysiol, 84, 1588–1598. [PubMed: 10980029]

Godey B, Schwartz D, De Graaf J, Chauvel P. & Liegeois-Chauvel C. (2001) Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. Clin. Neurophysiol, 112, 1850–1859. [PubMed: 11595143]

Gramfort A, Luessi M, Larson E, Engemann D, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L. & Hämäläinen M. (2013) MEG and EEG data analysis with MNE-Python. Front. Neurosci, 7.

Gross J. (2014) Analytical methods and experimental approaches for electrophysiological studies of brain oscillations. J. Neurosci. Methods, 228, 57–66. [PubMed: 24675051]

Hari R, Pelizzone M, Mäkelä JP, Hällström J, Leinonen L. & Lounasmaa OV (1987) Neuromagnetic responses of the human auditory cortex to on- and offsets of noise bursts. Audiology : official organ of the International Society of Audiology, 26, 31–43. [PubMed: 3593099]

Harms MP, Guinan JJ, Sigalovsky IS & Melcher JR (2005) Short-term sound temporal envelope characteristics determine multisecond time patterns of activity in human auditory cortex as shown by fMRI. J. Neurophysiol, 93, 210–222. [PubMed: 15306629]

Harms MP & Melcher JR (2003) Detection and quantification of a wide range of fMRI temporal responses using a physiologically-motivated basis set. Hum. Brain Mapp, 20, 168–183. [PubMed: 14601143]

Hautus MJ (1995) Corrections for extreme proportions and their biasing effects on estimated values ofd′. Behav. Res. Methods Instrum. Comput, 27, 46–51.

Howard MF & Poeppel D. (2010) Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. J. Neurophysiol, 104, 2500–2511. [PubMed: 20484530]

Krumbholz K, Hewson-Stoate N. & Schönwiesner M. (2007) Cortical Response to Auditory Motion Suggests an Asymmetry in the Reliance on Inter-Hemispheric Connections Between the Left and Right Auditory Cortices. J. Neurophysiol, 97, 1649–1655. [PubMed: 17108095]

Lu T, Liang L. & Wang X. (2001a) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat. Neurosci, 4, 1131–1138. [PubMed: 11593234]

Lu T, Liang L. & Wang X. (2001b) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat. Neurosci, 4, 1131–1138. [PubMed: 11593234]

Luo H. & Poeppel D. (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron, 54, 1001–1010. [PubMed: 17582338]

Lütkenhöner B. & Steinsträter O. (1998) High-Precision Neuromagnetic Study of the Functional Organization of the Human Auditory Cortex. Audiology and Neurotology, 3, 191–213. [PubMed: 9575385]

Maris E. & Oostenveld R. (2007) Nonparametric statistical testing of EEG-and MEG-data. J. Neurosci. Methods, 164, 177–190. [PubMed: 17517438]

Näätänen R, Paavilainen P, Rinne T. & Alho K. (2007) The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin. Neurophysiol, 118, 2544–2590. [PubMed: 17931964]

Oostenveld R, Fries P, Maris E. & Schoffelen J-M (2011) FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Comput. Intell. Neurosci, 2011, 9.

Overath T, Kumar S, Stewart L, von Kriegstein K, Cusack R, Rees A. & Griffiths TD (2010) Cortical mechanisms for the segregation and representation of acoustic textures. J. Neurosci, 30, 2070–2076. [PubMed: 20147535]

Overath T, Kumar S, von Kriegstein K. & Griffiths TD (2008) Encoding of spectral correlation over time in auditory cortex. J. Neurosci, 28, 13268–13273. [PubMed: 19052218]

Overath T, Zhang Y, Sanes DH & Poeppel D. (2012) Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: fMRI evidence. J. Neurophysiol, 107, 2042–2056. [PubMed: 22298830]

Pantev C, Eulitz C, Hampson S, Ross B. & Roberts LE (1996) The auditory evoked “off” response: sources and comparison with the “on” and the “sustained” responses. Ear Hear, 17, 255–265. [PubMed: 8807267]

Pickett JM (1999) The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology. Allyn & Bacon, Boston, MA.

Poeppel D. (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. Speech Communication, 41, 245–255.

Pratt H, Starr A, Michalewski HJ, Bleich N. & Mittelman N. (2008) The auditory P50 component to onset and offset of sound. Clin. Neurophysiol, 119, 376–387. [PubMed: 18055255]

Rogers H. (2000) The Sounds of Language. An Introduction to Phonetics. Pearson, Harlow, Essex.

Ross B, Herdman AT & Pantev C. (2005) Stimulus Induced Desynchronization of Human Auditory 40-Hz Steady-State Responses. J. Neurophysiol, 94, 4082–4093. [PubMed: 16107530]

Sohoglu E. & Chait M. (2016a) Detecting and representing predictable structure during auditory scene analysis. ELife, 5, e19113.

Sohoglu E. & Chait M. (2016b) Neural dynamics of change detection in crowded acoustic scenes. Neuroimage, 126, 164–172. [PubMed: 26631816]

Southwell R, Baumann A, Gal C, Barascud N, Friston K. & Chait M. (2017) Is predictability salient? A study of attentional capture by auditory patterns. Philosophical Transactions of the Royal Society B: Biological Sciences, 372, 20160105.

Tang H, Crain S. & Johnson BW (2016) Dual temporal encoding mechanisms in human auditory cortex: Evidence from MEG and EEG. Neuroimage, 128, 32–43. [PubMed: 26763154]

Teki S, Barascud N, Picard S, Payne C, Griffiths TD & Chait M. (2016) Neural correlates of auditory figure-ground segregation based on temporal coherence. Cereb. Cortex, 26, 3669–3680. [PubMed: 27325682]

Teng X, Tian X, Doelling K. & Poeppel D. (2018) Theta band oscillations reflect more than entrainment: behavioral and neural evidence demonstrates an active chunking process. Eur. J. Neurosci, 48, 2770–2782. [PubMed: 29044763]

Teng X, Tian X, Rowland J. & Poeppel D. (2017) Concurrent temporal channels for auditory processing: Oscillatory neural entrainment reveals segregation of function at different scales. PLoS Biol, 15, e2000812.

van Zuijen TL, Sussman E, Winkler I, Näätänen R. & Tervaniemi M. (2004) Grouping of sequential sounds--an event-related potential study comparing musicians and nonmusicians. J. Cogn. Neurosci, 16, 331–338. [PubMed: 15068601]

Wang X, Lu T, Bendor D. & Bartlett E. (2008) Neural coding of temporal information in auditory thalamus and cortex. Neuroscience, 154, 294–303. [PubMed: 18555164]

Wang X, Lu T. & Liang L. (2003) Cortical processing of temporal modulations. Speech Communication, 41, 107–121.

Wang Y, Ding N, Ahmar N, Xiang J, Poeppel D. & Simon JZ (2012) Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: MEG evidence. J. Neurophysiol, 107, 2033–2041. [PubMed: 21975451]

Winkler I, Denham SL & Nelken I. (2009) Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn. Sci, 13, 532–540. [PubMed: 19828357]

Winkler I, van Zuijen TL, Sussman E, Horvath J. & Näätänen R. (2006) Object representation in the human auditory system. Eur. J. Neurosci, 24, 625–634. [PubMed: 16836636]

Yvert B, Fischer C, Bertrand O. & Pernier J. (2005) Localization of human supratemporal auditory areas from intracerebral auditory evoked potentials using distributed source models. Neuroimage, 28, 140–153. [PubMed: 16039144]

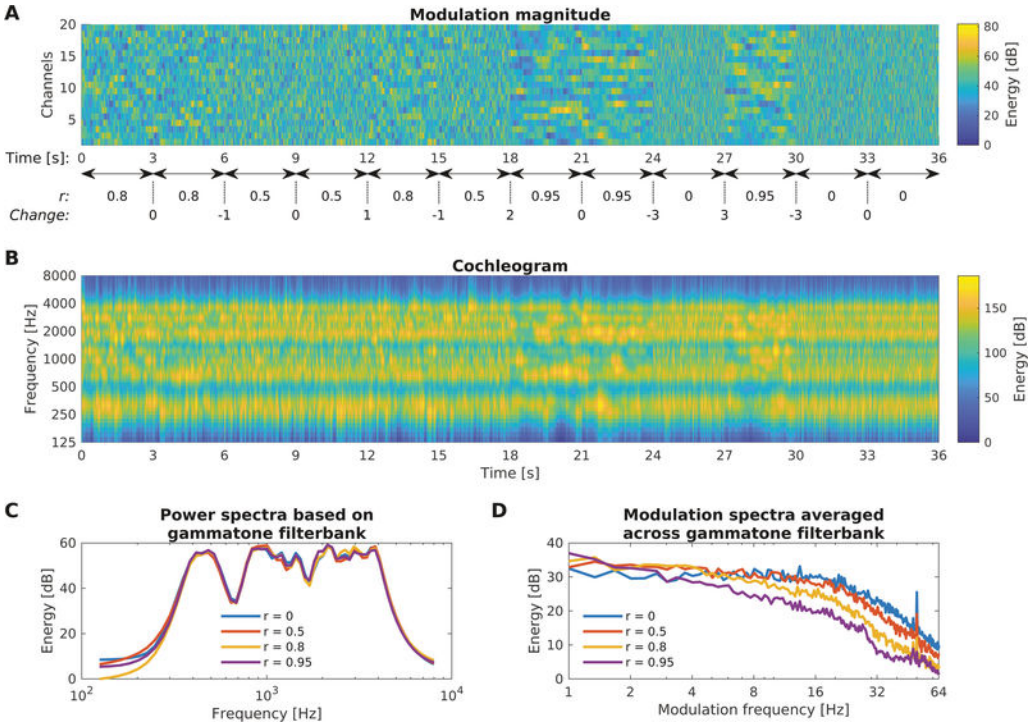


Figure 1. Schematic of the stimulus.

(A) The modulation magnitude values of the 20-ms frames are depicted. Note that these are only relative magnitudes, as the actual loudness of the sounds in the MEG scanner was scaled to approximately 75 dB. The 20 rows indicate frequency channels (whose actual values in Hz varied between sound blocks, see Materials and Methods). (B) Cochleogram of a stimulus generated from the modulation magnitude shown in A (64 gammatone filters over 20–8000 Hz; low-frequency range (<125 Hz) is truncated for visualization). (C) Power spectrum of the sound with the frequencies shown in B, calculated separately for the four levels of correlation. D) Modulation spectra averaged across gammatone filters for four levels of correlations. Modulation spectra were computed for each segment and averaged across exemplars. Note that the decay of the modulation power over frequency accelerates as *r* increases. Also, note that the peaks at 50 Hz were originated from the length of 20-ms frames but they were not reflected in the MEG data after preprocessing (see Supplementary Figure S1).

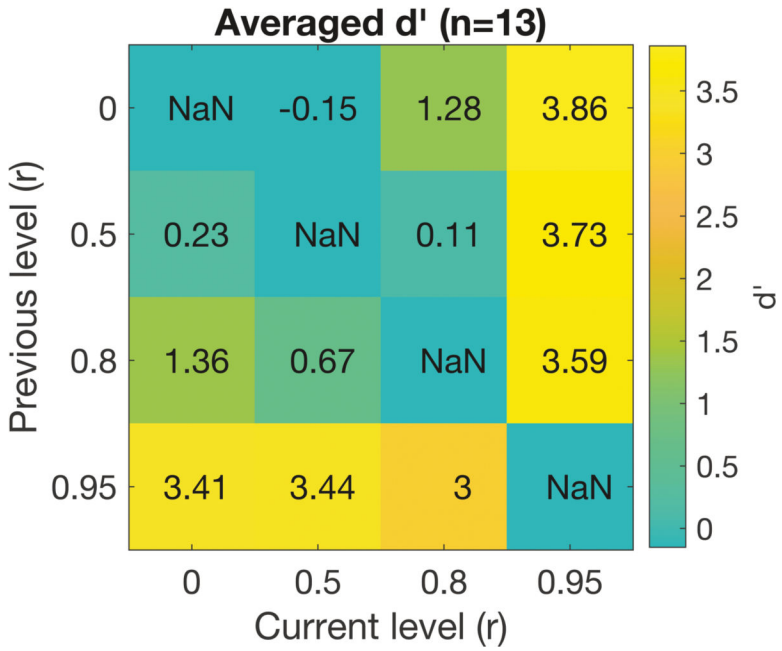


Figure 2. Behavioral performance.

Mean d-prime scores for all types of transitions are shown in 4-by-4 matrix.

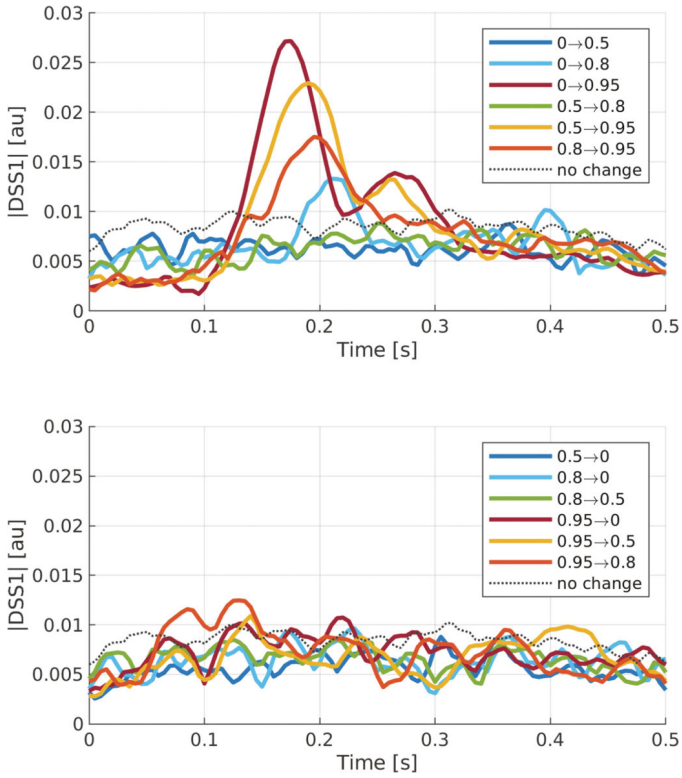


Figure 3. Average absolute DSS1.

Conditions with positive changes (upper) and negative changes (lower) are plotted separately. For the “no change” condition, DSS1s were extracted from conditions without changes (e.g., 0 → 0, 0.5 → 0.5, …) separately, then averaged.

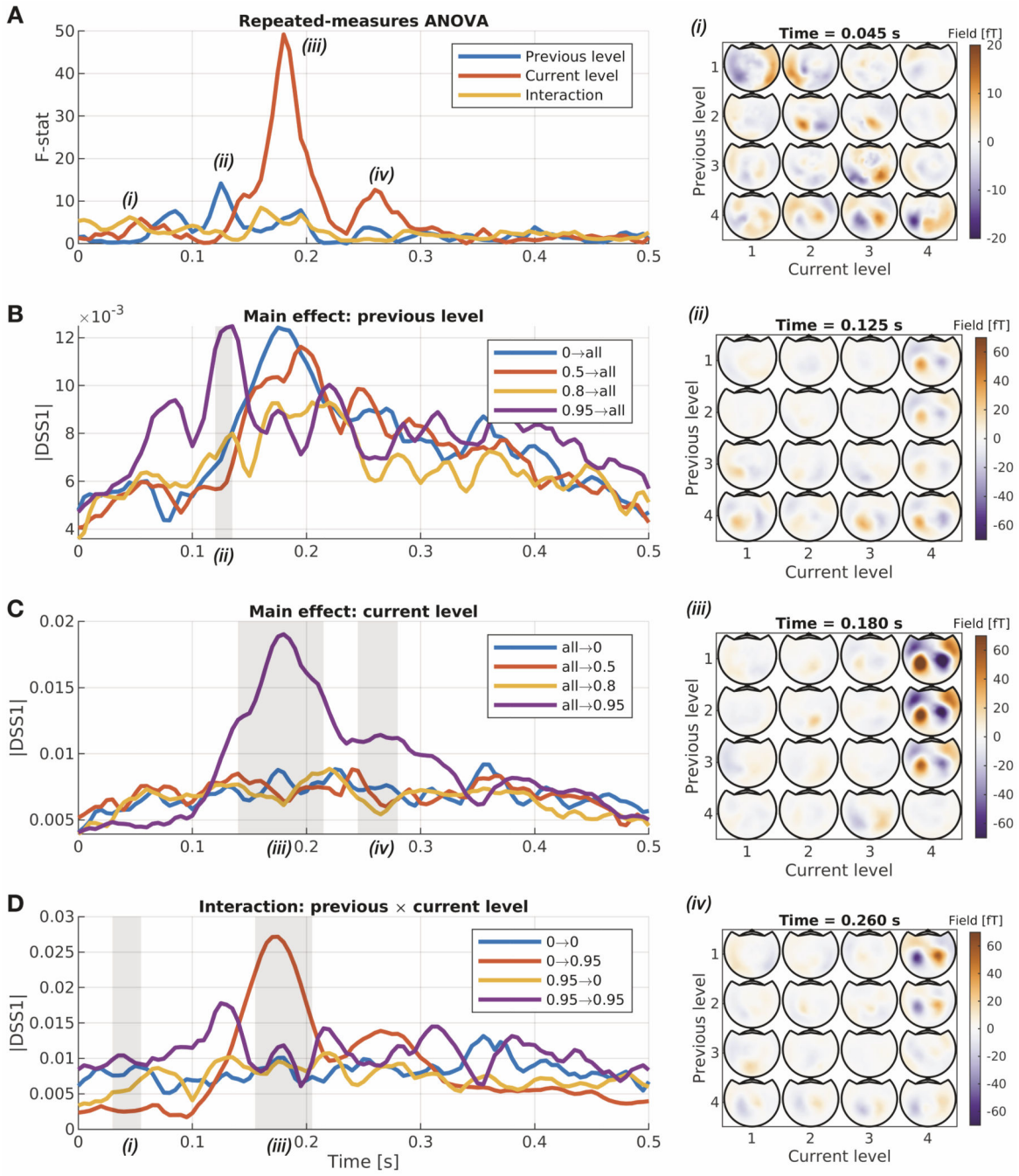


Figure 4. Repeated-measures ANOVA for Previous and Current Levels.

(A) F-statistics for main effects and their interaction, (B–D) averaged DSS1 time course to demonstrate each effect are plotted with significant clusters marked in gray shades. Topography of the projected DSS1 components for the time points (i, ii, iii, iv) with significant effects are shown on the right.

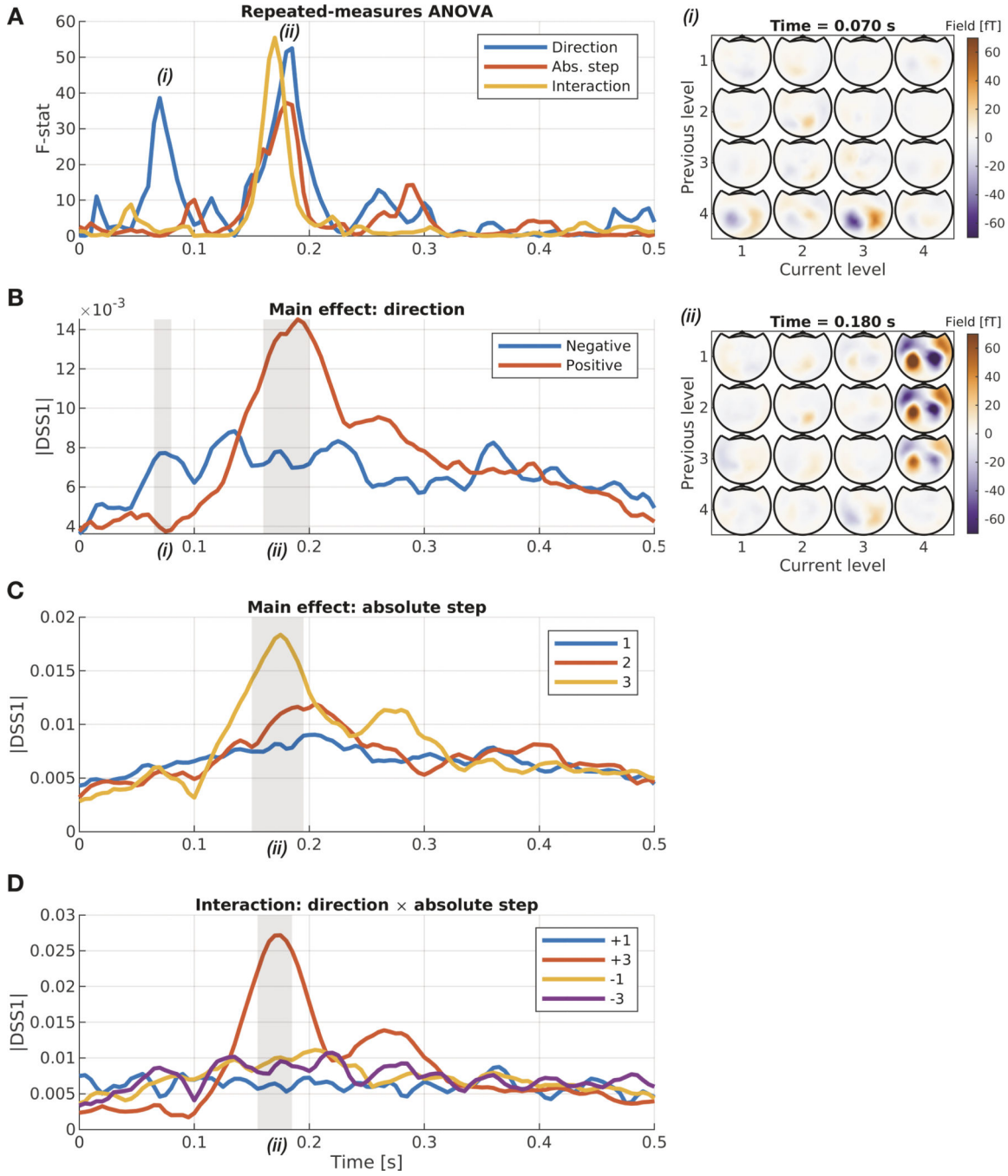


Figure 5. Repeated measures ANOVA for Direction and Absolute Step.

(A) F-statistics for main effects and their interaction, (B–D) averaged DSS1 time course to demonstrate each effect are plotted with significant clusters marked in gray shades. Topography of the projected DSS1 components for the timepoints (i, ii) with significant effects are shown on the right.

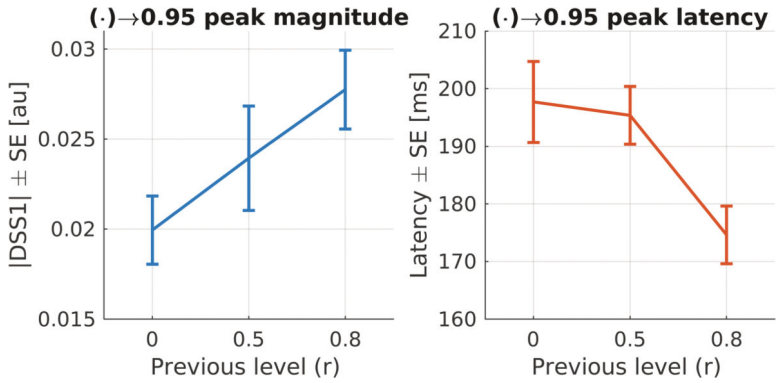


Figure 6. Dependency of peak magnitudes and latencies of evoked responses to previous levels of transitions to *r* = 0.95.

Maximal peaks were identified from |DSS1| components for a time-window of [130, 250] ms.

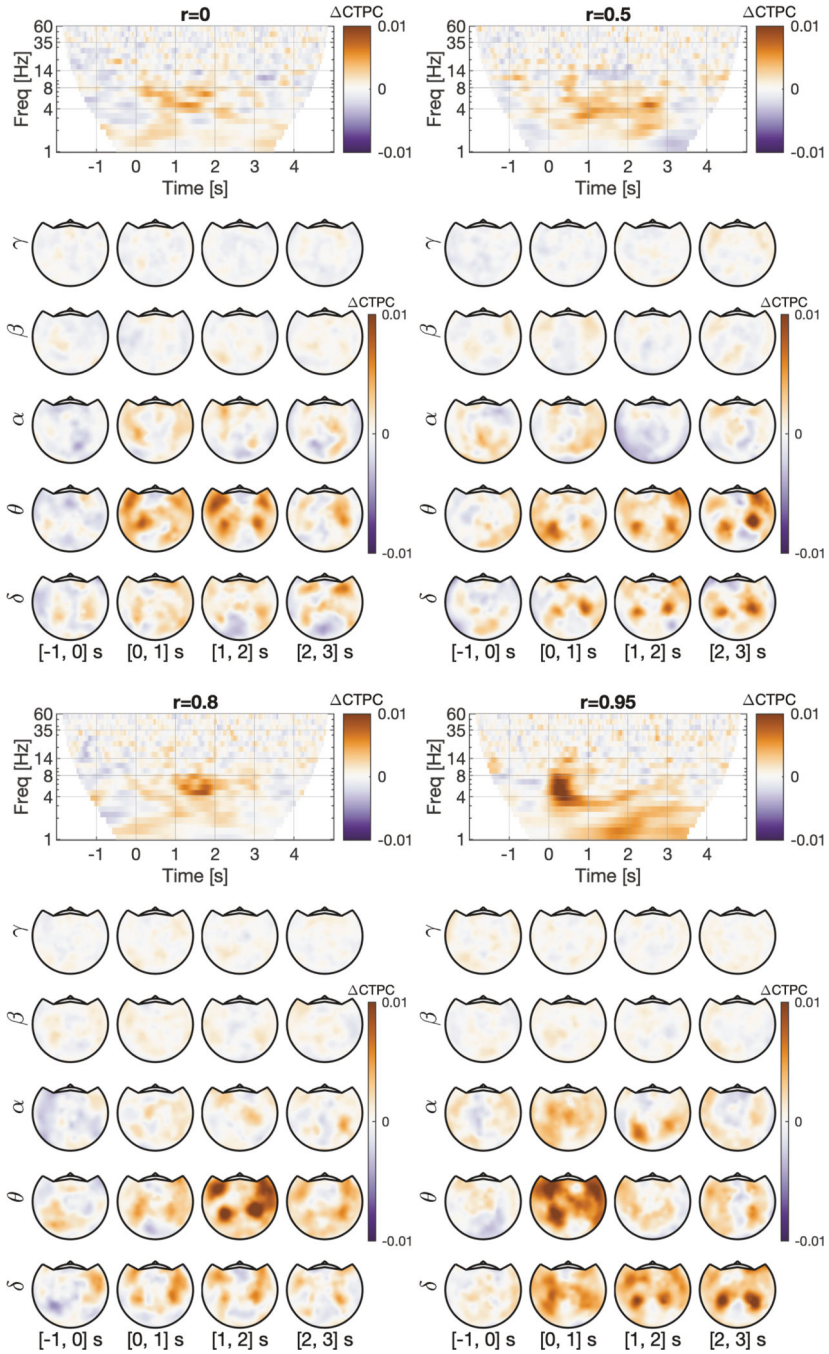


Figure 7. Average cross-trial phase coherence (CTPC) difference (across-CTPC subtracted from within-CTPC).

CTPC differences averaged across all channels are shown on time-frequency planes. Averaged in 1-sec time-window (for each column) and frequency band (for each row), the topography of CTPC differences is also displayed below each time-frequency plot for each correlation level.

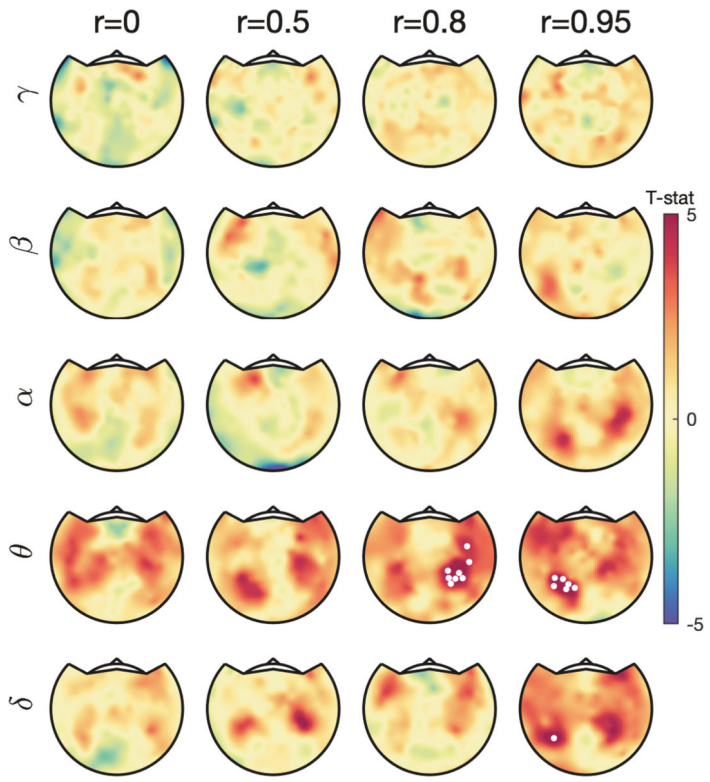


Figure 8. One-sample T-test for the significance of CTPC difference.

T-statistic maps on sensor space are shown for each correlation level (for each column) and frequency band (for each column) for a time-window of [0, 3] s. Channels belonging to a significant cluster are marked by white dots (corrected $p < 0.05$).

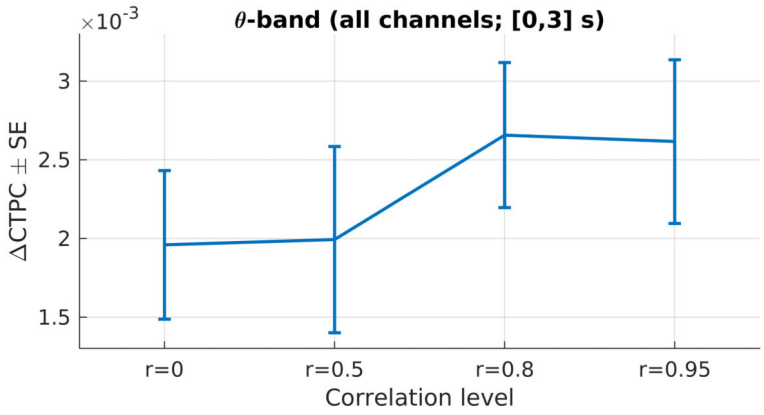


Figure 9. Linear dependency of theta-band CTPC difference on correlation levels.
Averaged across all channels over time-window of [0, 3] s after transition showed a significant increase as revealed by a repeated-measures regression model ($p < 10^{-4}$).

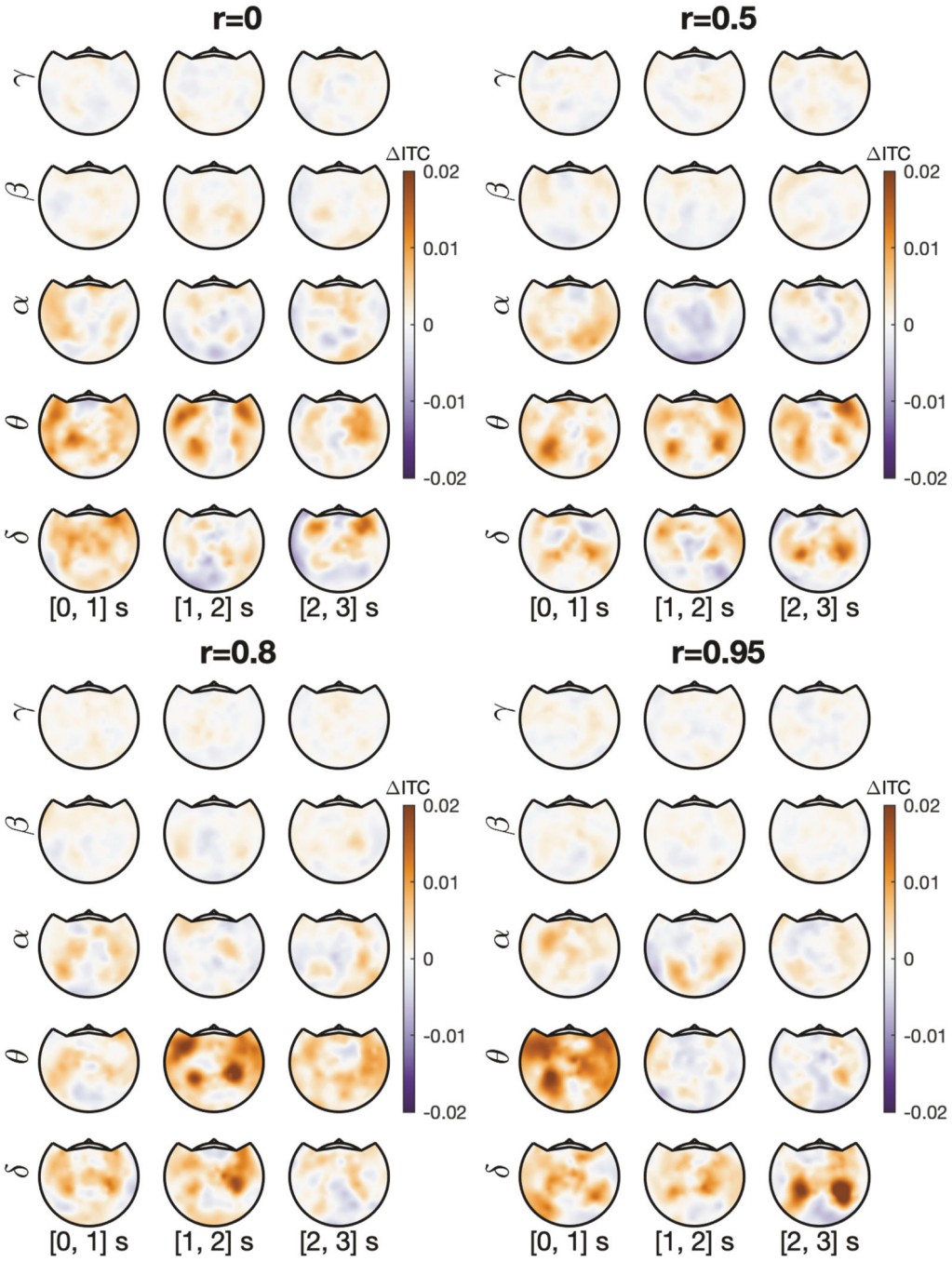


Figure 10. Average inter-trial correlation (ITC) difference (across-ITC subtracted from within-ITC).

ITC differences computed for 1-sec time-window (for each column) and frequency band (for each row) are shown for each correlation level.