Dynamics underlying auditory-object-boundary detection in primary auditory cortex

Pradeep Dheerendra1,2,3 | Nicolas Barascud4,5 | Sukhbinder Kumar1,2 | Tobias Overath2,6 | Timothy D. Griffiths1,2

1Biosciences Institute, Newcastle University, Newcastle upon Tyne, UK
2Wellcome Centre for Human Neuroimaging, University College London, London, UK
3Institute of Neuroscience & Psychology, University of Glasgow, Glasgow, UK
4LSCP, Département d’Etudes Cognitives, ENS, EHESS, CNRS, PSL Research University, Paris, France
5Ear Institute, University College London, London, UK
6Department of Psychology and Neuroscience, Duke University, Durham, North Carolina, USA

Correspondence
Pradeep Dheerendra, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK.
Email: pradeep.dheerendra@gmail.com

Funding information
Wellcome Trust, Grant/Award Numbers: WT102561/Z/13/Z, WT106964MA

Edited by: Edmund Lalor

Abstract
Auditory object analysis requires the fundamental perceptual process of detecting boundaries between auditory objects. However, the dynamics underlying the identification of discontinuities at object boundaries are not well understood. Here, we employed a synthetic stimulus composed of frequency-modulated ramps known as ‘acoustic textures’, where boundaries were created by changing the underlying spectrotemporal statistics. We collected magnetoencephalographic (MEG) data from human volunteers and observed a slow (<1 Hz) post-boundary drift in the neuromagnetic signal. The response evoking this drift signal was source localised close to Heschl’s gyrus (HG) bilaterally, which is in agreement with a previous functional magnetic resonance imaging (fMRI) study that found HG to be involved in the detection of similar auditory object boundaries. Time–frequency analysis demonstrated suppression in alpha and beta bands that occurred after the drift signal.

KEYWORDS
auditory object, change detection, MEG, perceptual decision making, statistical learning

Abbreviations: BOLD, bold oxygenation level dependent; DC, direct current; DICS, dynamic imaging of coherent sources; EEG, electroencephalography; FM, frequency modulation; fMRI, functional magnetic resonance imaging; GS, greedy search; HG, Heschl’s gyrus; MEG, magnetoencephalographic; MNI, Montreal Neurological Institute; MSP, multiple sparse priors; RMS, root mean square; SFG, superior frontal Gyrus; SMG, supramarginal gyrus; SPM, statistical parametric mapping.

Pradeep Dheerendra and Nicolas Barascud have equal contribution.

[Correction added on 9 October 2021, after first online publication: The affiliation of fourth author has been corrected and the Present Address section has been removed.]
1 | INTRODUCTION

Humans and animals need to accurately detect the emergence of new sounds within the environment in order to determine their behavioural relevance: important for communications and the events such as sounds of predator or prey. The brain therefore requires a way of identifying any new sound as it appears in the acoustic environment, separating it from other auditory objects that are present and representing it as a stable perceptual unit in our mind: a distinct object. Thus, an auditory object is defined (Griffiths & Warren, 2004) as the computational result of the auditory system’s ability to detect, extract, segregate and group the spectrotemporal regularities in the acoustic environment into stable perceptual units. A critical aspect of the sound analysis (Bizley & Cohen, 2013; Griffiths & Warren, 2004) is the detection of boundaries between auditory objects: regions in frequency–time space at which there is a transition between different types of pattern distributed over frequency–time space with different statistical properties.

To understand the brain basis underlying auditory object detection, Overath et al. (2010) controlled the spectrotemporal statistics of synthetic sounds to form auditory objects (and auditory object boundaries). Auditory objects were defined based on the degree of spectrotemporal coherence between frequency ramps; whereas the onset times and starting frequency of each ramp were distributed randomly, their slope was controlled such that the slope was either identical between ramps or completely random (see Figure 1a). Recording functional magnetic resonance imaging (fMRI) while participants reported the detection of auditory object boundaries, Overath et al. (2010) found that bilateral activity in Heschl’s gyrus that reflected the detection of auditory object boundaries. Their data can be interpreted in terms of the recognition of a local rule change that signifies a boundary between objects before the analysis of the perceived timbral properties of the objects themselves in higher cortical areas.

A growing body of work using electroencephalography (EEG) and magnetoencephalography (MEG) has explored the neuronal dynamics underlying auditory object emergence using synthetic stimuli in both passive and active listeners. Prominent sustained neural responses have been observed in the magneto- or electroencephalographic activity as a low-frequency DC power offset. These sustained responses have been reported to be modulated by various manipulations of sound statistics ranging from very low-level features such as amplitude and duration in pure tones (Pantev et al., 1994; Picton et al., 1978a, 1978b) and temporal regularity using click trains (Gutschalk et al., 2002) or regular interval noise (Lütkenhöner et al., 2011) to higher order regularities such as changes in frequency-modulated (FM) coherence (Herrmann et al., 2021; Herrmann & Johnsrude, 2018), transitions between complex tone-pip patterns (Barascud et al., 2016; Herrmann & Johnsrude, 2018; Southwell et al., 2017) or changes in temporal coherence within tone clouds (Rezaeizadeh & Shamma, 2021; Teki et al., 2016).

In particular, Barascud et al. (2016) investigated how the brain detects temporal patterns in complex sound sequences in the absence of attention. They found a slow increased evoked signal at the transition of randomness to regularity using a sequence of tone pips that had evolved to a regular repeating pattern. The transitions in Barascud et al. (2016) could be considered an auditory object boundary due to the change in the spectrotemporal statistics. Teki et al. (2016) investigated brain mechanisms for figure–ground segregation in the absence of a relevant task. Using a random tone cloud containing temporally coherent tones at certain frequencies that repeated in time, they found a robust sustained evoked response that reflected the emergence of a figure (auditory object) from the random background. Using the same stimuli with a linear regression approach, O’Sullivan et al. (2015) had found increased neural response for coherent stimuli.

Sohoglu and Chait (2016) employed two kinds of auditory scenes: Random scenes consisted of seven to eight tones, each with a specific carrier frequency and duration but with random inter-tone interval; regular scenes were identical but for a constant inter-tone interval. They reported that the appearance of a new auditory object in an ongoing scene leads to increased evoked activity in a regular scene compared with a random scene in both active and passive conditions. These studies (Barascud et al., 2016; Herrmann & Johnsrude, 2018; O’Sullivan et al., 2015; Sohoglu & Chait, 2016; Teki et al., 2016) report a sustained MEG and EEG response associated with emergence of a new auditory object in an ongoing scene.

Although the increases of sustained activity are well documented in the context of regularity changes (Barascud et al., 2016; Herrmann & Johnsrude, 2018; Sohoglu & Chait, 2016; Teki et al., 2016), it remains unclear whether these responses are associated with an oscillatory signature (Auksztulewicz et al., 2017). Notably, alpha and beta oscillations have been linked with sound stimulation and processing in the auditory cortex (Billig et al., 2019) and the regulation of the formation of auditory objects (Griffiths et al., 2019; Strauß et al., 2014).

At the cortical level, the previous fMRI experiment (Overath et al., 2010) established the anatomical location of auditory boundary detection with similar stimuli and provides strong priors for the present analysis.
Other studies have suggested an involvement of sources in primary and secondary auditory cortices (Barascud et al., 2016; Gutschalk et al., 2002; Pantev et al., 1994, 1996), and possibly also frontal cortex, parietal cortex and hippocampus (Barascud et al., 2016; Teki et al., 2016; Tiitinen et al., 2012) in this process.

In this experiment, we sought to investigate the neural dynamics of the detection of object boundaries in auditory cortex, exploiting the temporal sensitivity of MEG. We recorded responses to synthetic stimuli containing changes in the spectrottemporal statistics while participants carried out an active boundary detection task. Based on prior studies (e.g. Barascud et al., 2016), we hypothesised that transitions in the spectrottemporal properties of the acoustic textures stimuli would be indexed by a slow DC drift in the neuromagnetic evoked responses. Our results also confirm the anatomical location of statistical boundary transitions suggested by the previous study (Overath et al., 2010) and characterise the dynamics of this process based on the evoked activity. Further, we demonstrate decrease in the induced activity in the alpha (10–15 Hz) and beta (>20 Hz) bands.

2 | MATERIALS AND METHODS

2.1 | Subjects

Sixteen right-handed paid participants (aged 18–41 years; mean age 25.3 years; nine females) with normal hearing provided written consent before taking part in the study. The study was approved by the Institute of Neurology Ethics Committee (London, UK). All subjects reported no history of audiological or neurological disorders.

2.2 | Stimuli

The synthetic ‘acoustic textures’ stimuli were constructed using randomly distributed linear FM ramps with varying trajectories. The percentage of spectrottemporal modulation that are coherent is determined by the proportion of ramps with identical direction (slope sign) and trajectory (slope value). This proportion can be systematically controlled, allowing the creation of auditory objects with different levels of coherence. Boundaries were created by juxtaposing acoustic textures of different coherence levels. In Figure 1a, a 3-s stimulus segment with 0% spectrottemporal coherence (ramps move independently of each other) is followed by a 3-s stimulus segment with 100% spectrottemporal coherence (all ramps move upwards and with the same trajectory).

Sound stimuli were created using scripts written in MATLAB (MathWorks, Natick, USA) Version 6.5 at a sample rate of 44.1 kHz and 16-bit resolution. Stimuli consisted of a dense texture of linear FM ramps; each ramp had a duration of 500 ms and started at a random time and frequency (passband, 250–2500 Hz), with a density of 30 ramps per second, roughly equalling one ramp per critical band. For ramps that extended beyond the
pass-around, that is, >2500 or <250 Hz, we implemented a wrap-around such that the ramps continued at the other extreme of the frequency band, that is, at 250 and 2500 Hz respectively. Further, the direction of ramps that extended beyond the object changed suddenly, indicating the occurrence of auditory object boundary.

There were three texture conditions where the stimuli differed in terms of the coherent movement of the ramps: the percentage of ramps moving in the same direction for a given sound segment was either 0%, 100% or at detection threshold (X) that was determined individually for each participant. A boundary between the two different textures was created in the stimuli at half way along a 6-s-long stimuli. These three conditions will be referred to as ‘Transition 0–100’ condition, ‘No Transition 0–0’ condition and ‘Transition 0–X’ condition respectively.

2.3 | Task

Before the MEG experiment, participants were familiarised with the synthetic stimulus and then performed a one-interval two-alternative forced-choice (1I2AFC) adaptive tracking task to determine their perceptual object boundary threshold. Each 4-s stimulus started with 0% coherence and after 2 s transitioned to a different coherence level (100%, 80% ... 20%, 0%). Participants had to indicate whether they perceived two auditory objects (change detected) or one auditory object (no change detected). Each participant started with ‘Transition 0–100’ condition; participants had to get two consecutive correct responses before moving down a level, whereas one incorrect response resulted in moving up a level (two-down one-up adaptive tracking procedure). The last six reversal points were averaged to determine each participant’s perceptual threshold.

During the MEG experiment, the participants were asked to indicate whether or not they detected a change in the sound structure in this synthetic stimulus and convey their decision through a button press. They pressed two different buttons using their right hand for either a ‘change detected’ or ‘no change detected’ case. Participants were instructed to only respond after the sound ended to avoid motor response overlap with the auditory perceptual responses.

The data were collected in six sessions of 60 trials each, for a total of 360 trials per subject. The average inter-trial interval was 2 s with a random jitter within ±100 ms. For the ‘Transition 0–X’ condition, the behavioural result of the previous session was used to titrate and adjust the difficulty level for detecting transition in the next session such that the individual subject performance for the ‘Transition 0–X’ condition was expected to be around threshold on average, that is, 50%.

2.4 | MEG recording and data analysis

Magnetic signals were recording using a CTF-275 whole head MEG system (axial gradiometers, 274 channels; 30 reference channels; VSM MedTech) and analysed using SPM12 (Litvak et al., 2011) (http://www.fil.ion.ucl.ac.uk/spm/) in MATLAB 2019 (MathWorks, Inc.). Acoustic stimuli were presented diotically via MEG-compatible pneumatic tubes (EARTONE 3A 10 Ohm, Etymotic Research) using Cogent toolbox (http://www.vislab.ucl.ac.uk/cogent.php) running in MATLAB. Acquisition was continuous, with a sampling rate of 600 Hz and a 100-Hz hardware low-pass filter. Offline high-pass filtering was applied at 0.1 Hz, and low-pass filtering was applied at 30 Hz for all time-domain analysis (two-pass, fifth-order Butterworth). The data were then down-sampled to 200 Hz. Next, SPM’s built-in eyeblink artefact rejection was applied using a channel close to the eye (MRT21) because no electro-oculogram (EOG) or eye-tracking data were collected. Manual artefact trial rejection was applied using z-score metric, and any trials beyond seven were rejected (2.2% or 109 trials were rejected in total). Two participants were rejected as they had too many artefacts due to the presence of metal in their clothing. The trials were epoched from 0.5 before to 6.2 s after the onset of the acoustic stimuli. Epochs were baseline-corrected to the prestimulus interval −0.5–0 s.

2.5 | MEG evoked response

The root mean square (RMS) of the field strength across the temporal sensors was computed for each time sample as auditory cortex was shown (Overath et al., 2010) to be involved in detection of auditory object boundary. The time course of the RMS, reflecting the instantaneous amplitude of the neural responses, was used as a measure of neuronal responses evoked in the auditory system. For illustration purposes, the group RMS (RMS of individual subject RMSs) is shown, but the statistical analysis was always performed across subjects.

To statistically evaluate the latency of the transition response, the time points at which the ‘Transition 0–100’ condition (or ‘Transition 0–X’ condition) first showed a statistical difference from ‘No Transition 0–0’ control condition, the (squared) difference between RMSs of the two conditions evaluated across temporal sensors was
calculated for each participant and subjected to bootstrap resampling (Efron & Tibshirani, 1994), 1000 iterations—balanced implementation that ensures that all trials are utilised at the end of resampling. This difference was considered significant if the proportion of bootstraps that fell above or below zero was >99.5% for 15 or more consecutive samples (75 ms). The bootstrap analysis was run over the entire epoch duration, and all significant intervals identified in this way are indicated as horizontal bars at the bottom of the panel.

2.6 | MEG evoked response source localisation

Source localisation was performed using the greedy search (GS) option of the multiple sparse priors (MSP) algorithm (Friston, Harrison, et al., 2008). The MSP algorithm (López et al., 2012) requires hundreds of cortical source patches to be defined a priori, each corresponding to a single potentially activated region of cortex. SPM12 uses 512 patches covering entire cortical surface. An optimisation process follows whose objective is to obtain a set of hyper-parameters that maximise the evidence for the data using free energy as cost function (Friston et al., 2007). The negative variational free energy is a trade-off between the accuracy of the model and complexity of achieving that accurate model. For this, the GS scheme was adopted, which uses informative hyper-priors that ensure that most of the hyper-parameters shrink to zero, thereby providing a minimal sparse solution in source space.

In practice, first, the sensor positions were co-registered with 8196 vertex cortical mesh template (provided by SPM12 and defined in Talairach and Tournoux stereotaxic space) using the free fiducial marker locations (Mattout, Henson, & Friston, 2007). Next, a spherical head model was used while computing the gain matrix of the lead field model. Then, the source estimate was obtained on the cortical mesh via inversion of the forward model (Mattout et al., 2006; Mattout, Phillips, et al., 2007) using MSP method (Friston, Harrison, et al., 2008) with GS option (Friston, Chu, et al., 2008) without specifying any priors but under group constraints (Litvak & Friston, 2008) as implemented in SPM12. During source inversion, the frequency window of interest was 0–1 Hz, whereas the time window of interest was guided by the duration over which the two conditions under consideration diverged in the sensor level time-domain results. In the second level random effects analysis, the T-contrast used was ‘Transition 0–100’ condition ≥ ‘No Transition 0-0’ condition. Because individual MRI structural scans were not collected from the participants, the co-registration of the MEG data with the MNI template was based on approximate fiducial locations.

2.7 | Induced responses

Time–frequency representations (TFRs) were computed for each sensor separately for all trials using the multitaper convolution algorithm (‘mtmconvol’ method) implemented in Fieldtrip (Oostenveld et al., 2011) using Hanning tapers ranging from 1 to 30 Hz with a sliding 500-ms time window, in steps of 1 Hz and 100 ms. A non-parametric cluster-based permutation resampling approach using Fieldtrip (Popov et al., 2018) was adopted, which corrected for multiple comparisons across time–frequency-channel clusters. Monte Carlo estimates of the significance probabilities from the permutation distribution were derived from 1000 random draws, and a two-sided statistical test with alpha = .025 was adopted.

2.8 | MEG-induced response source localisation

Source localisation of induced responses was performed using the DAiSS toolbox (https://github.com/SPM/DAiSS) implemented in SPM12. The dynamic imaging of coherent sources (DICS) beamformer algorithm (Gross et al., 2001) was used to generate maps of source activity at specific time interval and frequency range for each condition. DICS beamformer operates in the frequency domain but has the same filtering principle as linearly constrained minimum variance (LCMV) beamformer (van Veen et al., 1997) operating in time domain, which estimates weights that linearly map the MEG sensors to source space. So, power is estimated at each source location while simultaneously minimising interference through suppression of contributions from other sources, resulting in an enhanced detection of the target source activity.

Sources were estimated for both conditions together using common filter weights, thus ensuring that differences in source activity were not related to spatial filter differences. The results from the sensor level time–frequency–channel cluster-based permutation test informed the source localisation. Source activity was estimated for 10–15 Hz (close to alpha) and 23–29 Hz (upper beta), which survived the time–frequency analysis at sensor level. The time interval chosen for localisation was between 5.5 and 6.0 s after sound onset where the difference in conditions emerged in the sensor-level analysis. In the second level random effects analysis, the
T-contrast used was ‘Transition 0–100’ ≤ ‘No Transition 0-0’ because only deactivations were seen at the sensor-level analysis.

3 | RESULTS

3.1 | Behaviour

Behavioural results obtained during scanning for the detection of change in spectrotemporal coherence was robust (Figure 1a). Hit rates for the ‘Transition 0–100’ condition (mean = 98.49%, s.e.m. = 0.58%, paired-sample t-test: t(13) = 37.22, p < 1e–14) and the ‘Transition 0–X’ condition (mean = 54.59%, s.e.m. = 3.76%, paired-sample t-test: t(13) = 13.42, p < 5e-9) were significantly different from the false alarm rate for the ‘No Transition 0–0’ condition (mean = 15%, s.e.m. = 2.11%). The d-prime scores for the ‘Transition 0–100’ condition (mean = 3.87, s.e.m. = 0.29, one-sample t-test: t(13) = 9.68, p < 2e-7) and the ‘Transition 0–X’ condition (mean = 1.23, s.e.m. = 0.1, one-sample t-test: t(13) = 2.38, p < 0.0332) were statistically significant from d-prime of 1.

3.2 | Evoked response

RMS field strength across all temporal sensors in both hemispheres was calculated for all three conditions in each subject. The time course of the RMS, reflecting the instantaneous amplitude of neural response, is a measure of neuronal responses evoked by the stimulus (Figure 2a). The difference between conditions was subjected to bootstrap resampling—the ‘Transition 0–100’ condition significantly differed from ‘No Transition 0–0’ control condition, 255 ms after the onset of the texture boundary and stayed on for 1.31 s before decaying back to the ‘No Transition 0–0’ control condition highlighted by yellow horizontal bar at the bottom of Panel A in Figure 2.

The ‘Transition 0–X’ condition did not differ from ‘No Transition 0–0’ condition. Further, when the trials of the ‘Transition 0–X’ condition were split based on participant’s percept as either trials with correct response (Hit) or trials with incorrect response (Miss), their evoked response did not differ significantly. Additionally, when the trials of the ‘Transition 0–X’ condition were split across median into two groups based on their average single-trial RMS amplitude during 3.5- to 4.5-s time interval (Figure 2a), the detection performance between the two groups (mean = 56% high-RMS group vs. mean = 53% in the low-RMS group) did not differ significantly (paired t-test, t(13) = 0.76, p = .46).

3.3 | Evoked response source localisation

Source localisation applied to the interval over which conditions ‘Transition 0–100’ and ‘No Transition 0–0’ diverged showed increased cortical activity in an area very close (9 mm) to Heschl’s gyri bilaterally (Figure 2b and Table 1).

3.4 | Induced response

The induced responses were computed at sensor level using multi-tapers and then subjected to a time–channel–frequency cluster-corrected permutation test using Fieldtrip (Figure 3a). This exploratory analysis revealed that the ‘Transition 0–100’ condition was suppressed relative to the ‘No Transition 0–0’ condition in the frequency ranges 10–15 Hz, overlapping with the alpha band, and 23–29 Hz, corresponding to the upper beta band over the 4.0- to 6.0-s peri-stimulus time interval.

3.5 | Induced response source localisation

Using DICS beamformer, the cortical sources underlying the suppression seen in induced responses were localised as summarised in Figure 3b and Table 2. The 10- to 15-Hz (~alpha) suppression seen at the sensor level was source localised to right superior frontal gyrus (SFG) and right supramarginal gyrus (SMG), whereas the 23- to 29-Hz upper beta suppression was source localised to right angular gyrus.

4 | DISCUSSION

This work examines the cortical activity accompanying the fundamental process of detecting auditory object boundaries in humans using MEG. We used synthetic stimuli with statistical properties that were manipulated to create auditory object boundaries by abruptly varying the spectrotemporal coherence. We observed a distinct response to this sudden change in coherence, visible both in the evoked response and the time–frequency domain. The transition response was manifest as a slow (<1 Hz, ~1 s long) drift in the evoked response directly following the object boundary, which was source localised to auditory cortex bilaterally. This drift was followed by a decrease in alpha (10–15 Hz) and high beta (23–29 Hz) frequency bands.

The previous MRI study (Overath et al., 2010) performed a systematic manipulation of spectrotemporal
Characterisation of the evoked response. (a) Neuromagnetic evoked group RMS response averaged across temporal sensors for ‘Transition 0–100’ condition in yellow trace, ‘Transition 0–X’ condition in orange trace and for ‘No Transition’ control condition (at 0% coherence) in blue trace. Responses to ‘Transition’ condition significantly differed (highlighted through yellow bar at the bottom) from ‘No Transition’ condition from 255 ms after object boundary for 1.31 s before decaying to control condition. (b) Source localisation results showing bilateral activation around Heschl’s gyri for four different axial slices. The un-thresholded in orange transparent overlays, whereas the voxels surviving FWE correction are highlighted in red.

<table>
<thead>
<tr>
<th>Location</th>
<th>Side</th>
<th>MNI coordinates</th>
<th>Peak-level stats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heschl’s gyrus</td>
<td>Right</td>
<td>X: 52 Y: -14 Z: 16</td>
<td>T: 4.26 p: .000</td>
</tr>
<tr>
<td></td>
<td>Left</td>
<td>X: -54 Y: -12 Z: 16</td>
<td>T: 4.03 p: .001</td>
</tr>
</tbody>
</table>

Notes: Because individual structural MRI data were not used, the ROIs should be regarded as approximations of particular cortical areas only. The main aim of this study was to reveal the underlying temporal dynamics.
FIGURE 3 Legend on next page.
coherence in textures stimuli to delineate the brain circuits involved in the detection of change in spectrotemporal characteristics of the acoustic signal, from those involved in the encoding of absolute spectrotemporal coherence in itself. Brain activation in response to changes in spectrotemporal coherence was found in Heschl’s gyri bilaterally, whereas the spectrotemporal coherence itself was encoded in non-primary areas. The present results for perceiving a change in spectrotemporal coherence are consistent with this previous fMRI study that used the same synthetic stimuli (Overath et al., 2010) subject to the correspondence between evoked MEG response and BOLD activity.

4.1 Drift response: Similarities to previous work

The present MEG study was not designed to address the question of encoding of absolute spectrotemporal coherence but instead aimed at characterising the temporal dynamics of boundary detection. Our results show increased sustained MEG activity for spectrotemporally coherent textures compared with random textures. This observation is consistent with numerous EEG and MEG studies reporting enhanced sustained responses for sounds containing a temporally regular pattern (Barascud et al., 2016; Gutschalk et al., 2002; Herrmann et al., 2021; Herrmann & Johnsrude, 2018; Pantev et al., 1994; Pantev et al., 1996; Sohoglu & Chait, 2016; Southwell et al., 2017; Southwell & Chait, 2018; Teki et al., 2016).

To explain the slow drift signal seen in the current MEG evoked response, previous research has invoked a predictive coding account of perceptual inference (Friston, 2005; Friston & Kiebel, 2009). In this account, our brains are not just passively reacting to the incoming sensations but are constantly predicting it. This ‘predictive-coding’ account suggests that, in the continuous ongoing predict–compare–update process (Friston, 2010), temporally regular acoustic inputs have higher relevance for predicting how the auditory world would change compared with temporally irregular sounds. Thus, temporally regular sounds are up-weighted, whereas temporally random sounds are down-weighted. In this framework, ‘precision’ signals corresponding to the inverse variance of prediction errors vary with the level of regularity of sensory inputs, such that temporally regular (irregular) sounds are afforded higher (lower) precision. Thus, precise prediction errors become more relevant during belief updating. In this framework, sensory attention is formalised as a process that infers this level of predictability of sensory inputs (Heilbron & Chait, 2018) and results in up- or down-weighting the prediction errors according to the stimulus precision.

There is prior literature to support enhancement or up-weighting of temporally regular sounds. Sohoglu and Chait (2016) showed the effect of scene regularity by employing regular and random auditory scenes and reported increased evoked response in regular scenes compared with random scenes in both active and
passive listeners. Consistent with this result, Barascud et al. (2016) found a slow increased evoked signal marking the transition of randomness to regularity using a sequence of tone pips that had evolved to a regular repeating pattern. This sustained activity was localised to a network that included auditory cortex. The regular sound sequence post-transition can be considered as an emergence of an auditory object. Similar to O’Sullivan et al. (2015), Teki et al. (2016) observed a robust sustained evoked response that reflected the emergence of a figure (auditory object) from random background. This sustained activity was source localised to a network that included planum temporale, consistent with their previous fMRI results (Teki et al., 2011). The onset of an auditory figure or coherent repeating tone pips can be interpreted as emergence of an auditory object. These and other studies (Barascud et al., 2016; Herrmann et al., 2021; Herrmann & Johnsrude, 2018; Sohoglu & Chait, 2016; Southwell et al., 2017; Southwell & Chait, 2018; Teki et al., 2016) relate to the detection of changes in statistical rules in frequency–time space so as to identify the emergence of auditory objects. These studies all report a steady-state signal that occurs at or shortly after the boundary, localised to a cortical region at or near primary auditory cortex. To illustrate the relationship between sustained activity changes and perceptual learning, Herrmann et al. (2021) used transitions between random and regularly repeating tone-pip patterns and found increased neural responses when the regular patterns were completely novel to the subject as opposed to when the pattern had reoccurred at various points through the experiment (see also Andrillon et al. (2017)). Our results (increased sustained response to novel statistical regularities in the stimulus) are in agreement with this line of research and extend previous findings to another type of statistical regularity, namely, spectrotemporal coherence.

4.2 | Drift response: Differences from previous work

It is worth noting that unlike the above studies (Barascud et al., 2016; Herrmann et al., 2021; Sohoglu & Chait, 2016; Teki et al., 2016), the transition response in the ‘Transition 0–100’ condition was not fully steady state: It did not persist until the end of the stimulus. One reason for this could be the stimulus. Here, although the auditory object beyond the transition in ‘Transition 0–100’ trials is arguably more regular than before, it is still not fully regular because, for example, the onset time and frequency were all determined randomly. This structural difference in our stimulus might explain why the response we observe resembles a drift but does not reach full steady state unlike previous studies that used temporally regular patterns (Barascud et al., 2016; Teki et al., 2016).

Sohoglu and Chait (2016) also reported an ‘appearance effect’ where the appearance of a new auditory object in an ongoing scene leads to an increased evoked activity in a regular scene compared with a random scene in both active and passive listeners. They interpreted this ‘overshoot’ as evidence for mechanism that infers precision of sensory input that is used to upregulate the processing of more reliable sensory signals. They also showed that the appearance of a new auditory object in an ongoing scene is characterised by a canonical succession of M50/M100/M200 event-related fields seen at object onset in both active and passive listeners. However, due to the stochastic nature of the stimuli employed in our study, where the FM ramps start independently, these appearance-evoked responses were not seen in our ‘texture’ stimuli.

The very slow ramping up of evoked response visible in both conditions that occurs from the transition point could be due to evidence accumulation for making a decision and motor preparation for button press (Jahanshahi & Hallett, 2003). On the other hand, the transitory drift is very slow (<1 Hz) compared with a slow (200–500 ms) rise in the evoked response reported in the above studies. This could be due to differences in the stimulus employed—current study employed 500-ms FM ramps compared with 25- to 50-ms tone pips employed in the previous studies (Barascud et al., 2016; Sohoglu & Chait, 2016; Teki et al., 2016).

4.3 | Oscillatory signatures of auditory statistical boundary detection

Previous studies of the temporal dynamics of object detection have mostly focused on evoked activity. In this work, we also examined oscillatory dynamics following boundary detection. Our results show that the drift in the evoked response is followed by decreases in alpha and beta activity.

Beta oscillations are often linked with the motor system and typically decrease in power in anticipation of upcoming sensory–motor processing (for an overview, see Jensen et al. (2019)). In particular, orienting to an upcoming motor event involves a contralateral suppression of alpha- and beta-band oscillations within sensorimotor cortex (van Ede et al., 2011). Although our participants had to indicate their decision with a button press, they did not know which button to press prior to the transition and could thus not prepare any specific
motor response. Furthermore, our results primarily show an ipsilateral suppression, and the fact that we see a difference between ‘Transition 0–0’ and ‘Transition 0–100’ conditions is not entirely consistent with a purely motor preparation interpretation. Perhaps more relevant to the present discussion, higher level cognitive studies point to a role for beta oscillations in decision making, and beta activity has been proposed to be involved in the accumulation of evidence when perceptual decisions, and motor responses on those decisions, have to be made (Donner et al., 2009). The findings on decision making and beta oscillations give a strong processing connotation to the beta-band activity, which somehow is in contrast to observed functions of the motor cortical beta activity. Along those lines, beta activity has been strongly associated with expectation and attention in the brain (Tavano et al., 2019). For instance, Engel and Fries (2010) have suggested that beta activity reflects the maintenance of a status quo, whereby increases in beta would be observed during the persistence of a sensory or cognitive state, whereas decreases in beta would signal that an unexpected change has been detected. Our results show beta suppression in both transition and no-transition conditions, consistent with the idea that participants have arrived at a decision on whether a change in the stimulus statistics occurred. However, our results indicate that this beta suppression is significantly larger in the Transition 0–100 condition compared with control. We speculate that the salient novel and highly structured information in the ‘Transition 0–100’ condition might trigger the bottom-up capture of attention, which could explain the larger beta suppression observed in the ‘Transition 0–100’ condition relative to the ‘Transition 0–0’ control.

Concomitant with this beta suppression effect, we also found a relative decrease in alpha activity following the evoked drift in the Transition 0–100 condition relative to control condition. Alpha has long known to be a strong marker of cognitive effort and attention in MEG and EEG. Classically, higher alpha was thought to reflect an ‘idling’ state, and it has been acknowledged that during auditory perception (Billig et al., 2019) or while engaging in a task such as perceptual judgement, an increased attentiveness leads to a decrease in alpha power (Pfurtscheller et al., 1996). Attentional demand is known to modulate the extent of alpha power enhancement (for review, see Foxe and Snyder (2011)), and alpha power has been implicated in functional inhibition of task-irrelevant distraction in both vision (Snyder & Foxe, 2010) and audition (Wöstmann et al., 2017). Consistent with our results, alpha (and beta) decrease concomitant with enhanced DC evoked responses has already been reported by Aukstulewicz et al. (2017) in the context of auditory pattern detection for temporally repeating versus random tone-pip sequences. Importantly, decrease in alpha power was also linked to increasing stimulus predictability (Bauer et al., 2014; Wöstmann et al., 2019) as well as with increasing decision confidence (Wöstmann et al., 2015), which is consistent with our results because the statistical structure of post-transition texture in the T0–100 condition is more predictable.

Although all these accounts are more anticipatory—insofar as they pertain to upcoming sensory events—alpha has also been hypothesised to accompany the regulation of the formation of auditory objects (Strauß et al., 2014). As such, alpha would not only reflect preparatory attention but also ongoing attentive stimulus processing. For instance, alpha decreases have been documented for tasks that recruit the sensory representation (Hauswald et al., 2020). That we observe a decrease in alpha activity is consistent with these theses and might reflect an engagement of top-down reactive attention following the conscious detection of the newly formed auditory object. This might also explain why the observed alpha decrease occurred relatively later in the response.

Taken together, our alpha and beta effects integrate well with the recent frameworks proposed in the context of working memory (van Ede, 2018) and auditory perception (Griffiths et al., 2019) and could point to a role in stimulus processing. For instance, van Ede (2018) suggested that alpha power increases for tasks with sensory disengagement, although it decreases for tasks that recruit the sensory representation, when the stimulus is being prioritised based on current attentional demands. In a similar vein, Griffiths et al. (2019) also recently proposed that reduced alpha and beta oscillations may enhance information processing, effectively serving as an index of the fidelity of the stimulus representation. This interpretation also fits well with recent work linking changes in sustained response amplitude to perceptual learning (Andrillon et al., 2017; Herrmann et al., 2021).

4.4 Limitations

There are some limitations in the interpretation of the present results. First, because the auditory object boundary was created by manipulating regularity in the synthetic stimulus, the slow drift could be attributed to detection of regularity as much as detection of texture boundary. Because our study did not create object boundaries by reduction or elimination of regularity (i.e. ‘Transition 100–0’ condition or 100% coherence to 0% coherence transition), the respective contributions of object boundary detection and regularity detection are
confounded and cannot be disambiguated by the present study alone.

In the source localisation of the evoked activity, the peak cluster activity was found slightly outside of Heschl’s gyri, which was our prior hypothesis based on the previous fMRI study (Overath et al., 2010). This could be due to the fact that template MRIs were used for the forward model instead of individual structural scans, which reduced our spatial resolution. Further, the fact that the clusters did not survive family-wise error correction can also be explained by the relatively limited number of subjects in this study (14).

The lack of a difference between the ‘Transition 0-X’ and ‘No Transition 0-0’ conditions was unexpected. It is possible that this is due to the cancellation of neural responses between the detected (Hit) and undetected trials (Miss): The evoked response to the detection of a texture transition might be nullified during averaging with trials where the texture transition was not perceived by the participants. In the ‘Transition 0-X’ condition, the amount of change was held constant so that participants responded ~50% of the time that they heard a change in coherence (and 50% of the time that they did not), we had hypothesised that the neural response to those trials in which participants detected a change in coherence would be more similar to the Transition 0–100 condition, and those trials in which participants failed to detect a change in coherence would be more similar to the No Transition 0-0 condition. However, dividing the ‘Transition 0-X’ condition based on percept also did not yield a significant difference between Hit and Miss trials (results not shown). One possible explanation for this is the considerable variability in the participants’ individual performance accuracy for the ‘Transition 0-X’ condition (visible in Figure 2), such that the number of Hit and Miss trials is very different, which would undoubtedly affect detection power. This variability occurred despite careful individual threshold calibration and despite our efforts to readjust the threshold between each session. A continuous feedback (for each trial instead of each session) for the titration of the difficulty level for detection of texture transition might have proved more powerful for the ‘Transition 0-X’ condition. Another possibility for this lack of effect in the 0–X condition might be that the sustained response is directly related to the detection of stimulus coherence. Because the only condition that yielded an increased sustained response is Transition 0–100, it is possible that the observed overshoot in the RMS might indicate early decision making or motor preparation, whereas the 0–X and 0–0 conditions might require more time to be correctly identified (because they require detecting an absence of transition in the 0–0 case, and a much harder stimulus in the 0–X case). However, although we cannot entirely rule out this possibility on the basis of data collected in this study alone, many studies have found sustained response increases under passive listening conditions (Barascud et al., 2016; Herrmann & Johnsrude, 2021; Teki et al., 2016), and our own source localisation results (Figure 2b) are not entirely consistent with a sensorimotor generator site for this evoked activity.

4.5 | Summary

We created an auditory object boundary via the emergence of a regular auditory object in an ongoing random acoustic scene. Detecting such complex spectrottemporal changes relies on mechanisms that are fundamental for the analysis of ecologically valid sounds in a dynamic auditory environment. In predictive coding account of perceptual inference, precision is a long-term second-order statistic that represents the level of regularity of sensory inputs, which serves to determine the relevance of incoming inputs when updating the internal model. We speculate that the observed drift signal could be related to the increase of precision following changes in the coherence of the acoustic spectrum enabling the detection of auditory object boundary. Our results along with the spatial priors from the previous fMRI study are consistent with the notion that the auditory cortex plays a central role in the detection of complex changes in the higher order statistical properties of an auditory object. This detection appears to be mediated by infra-hertz activity originating from sensory cortices at an initial evidence accumulation phase, followed by alpha and beta oscillatory signatures in the subsequent perceptual decision stage.

ACKNOWLEDGEMENT

This research was funded by Wellcome Trust PhD studentship to P.D. (WT102561/Z/13/Z) and Wellcome Trust Senior Clinical Fellowship awarded to T.D.G. (WT106964MA).

CONFLICT OF INTEREST

The authors disclose that there is no conflict of interest.

AUTHOR CONTRIBUTIONS

Pradeep Dheerendra and Nicolas Barascud contributed equally. T.O. and T.D.G. designed the study. T.O. and N.B. collected the data. P.D. and N.B. analysed the data. P.D. wrote the original draft. All authors reviewed and edited the draft. T.D.G. acquired funding.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ejn.15471.
DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID
Pradeep Dheerendra https://orcid.org/0000-0002-1575-154X
Nicolas Barascud https://orcid.org/0000-0003-1495-561X

REFERENCES


---