Distinct cortical responses to Auditory Statistics: pre-attentive discriminations based on local and global representations

Author list: Martina Berto*[1], Emiliano Ricciardi[1], Pietro Pietrini[1], Davide Bottari[1]

Affiliations:

1 *Molecular Mind Lab, IMT School for Advanced Studies Lucca, Lucca, 55100, Italy*

Corresponding author
Martina Berto
martina.berto@imtlucca.it

ABSTRACT

The auditory system relies on local and global representations, depending on the amount of entering temporal details. When local information exceeds system constraints, acoustic details are summarized into a set of average statistics and global structure emerges. Whether computations subtending local and global representations of sounds can be distinguished at the cortical level is unknown. A computational auditory model was employed to extract auditory statistics from natural sound textures, such as fire, wind, and rain, and generate synthetic exemplars in which we had control of local and global properties. Participants were passively exposed to auditory streams with triplets of sounds including two standards and an oddball while the EEG was recorded. In Local Discrimination (exp 1) the oddball sound varied from standards only for its local features; in Global Discrimination (exp 2), it varied for its global structure. Different sound durations were employed to manipulate the amount of acoustic information and statistical variability between standards and oddballs. We expected short and long sounds to engage local (exp 1) or global (exp 2) representations, respectively. As predicted, neural responses associated with local discriminations exceeded global ones for short sounds, whereas the opposite pattern emerged for longer sounds. Electrical source estimates revealed that local discriminations mainly engaged auditory regions of the right hemisphere, while global discriminations of the left hemisphere. These findings revealed distinct neural correlates of acoustic discriminations guided by local or global structures. Ultimately results demonstrated that the auditory system developed a functional architecture to automatically detect changes occurring in the environment based on these two modes of sound representation.

SIGNIFICANT STATEMENT

Prior to this study, it was not known whether it was possible to directly measure auditory cortical representations of sound discriminations depending on local or global acoustic properties. Results unveiled the existence of a mechanisms which pre-attentively determines at which timescale (local or global) a change in a stream occurred based on statistical variability between sound excerpts. By performing source reconstruction, we additionally revealed that the response to local change detection pertains to different cortical structures as compared to global one, displaying also hemispheric asymmetries. These findings expand our knowledge on fundamental mechanisms characterizing auditory system functioning.

INTRODUCTION

The human auditory system is capable of discriminating sounds based on both local and global properties (McAdams, 1993; Griffiths, 2001). Local analysis relies on extracting and retaining fine-grained details to detect transient changes and amplitude modulations over time (Plomp, 1964; Dau, Kollmeier, and Kohlrausch, 1997). These temporal variations characterize different sound objects and help the system discerning across acoustic sources. However, environmental inputs typically comprise long lasting sounds in which the number of local features to be retained exceed sensory storage capacity. This prohibits discrimination based on local analysis to give way to global representations (McDermott, Schemitsch, and Simoncelli, 2013). In other words, as the duration of the entering sounds increases, global acoustic representations are built upon temporal features to condense information into a more compact, and retainable structure (Yabe et al., 1998; McDermott, Schemitsch, and Simoncelli, 2013). For stationary sounds (such as sound textures, e.g., rain, fire, waterfall, typewriting; Saint-Arnaud and Popat, 1995), characterized by a constant repetition of similar events over time, this form of compression consists of computing a set of auditory statistics which comprises time-averaged acoustic details. Consequently, local features are no longer retained after their average global representation emerges (McDermott, Schemitsch, and Simoncelli, 2013).

Computational synthesis approaches proved being fundamental for the characterization of local and global auditory analysis. The basic assumption derives from information theories (Barlow, 1961) and suggests that the brain represents a sensory input with a set of measurements. This would imply that signals containing values matching those measurements will be perceived as the same. Furthermore, computational approaches allow to extract auditory statistics from original recordings and impose them to synthetically produced sounds (McDermott and Simoncelli, 2011). By doing so, they provide the opportunity to create synthetic audio excerpts which sound like the original natural sounds, while controlling the embedded statistics. Previous behavioral studies have used this type of synthetic stimuli to characterize the temporal dynamics of the two computational modes, local and global representations, respectively (McDermott, Schemitsch, and Simoncelli, 2013; McWalter and McDermott, 2018). Two main questions remain unsolved. First, it is unclear if the two computational modes can occur automatically. Previous investigations (McDermott, Schemitsch, and Simoncelli, 2013; Berto et al., 2021) relied on psychophysics and required that participants paid attention to the stimuli throughout the experimental session to make deliberate judgments over the heard sounds. Based on this evidence alone, it is not possible to make assumptions whether the engagement of these modes can occur during passive exposure, i.e., automatically or require individual effort. Second, it is unknown whether the two basic auditory computations rely or not on the same neural correlates and subtend to similar cortical activations. That is, whether different neural fingerprints to the emergence of local and global representation can be detected. We hypothesized that the switch between the two modes of representation rely on different cortical networks in the auditory system.

We used a computational auditory model to extract auditory statistics from natural sounds and generate synthetic excerpts controlling local and global properties; we acquired the EEG of twenty-four participants being passively exposed to a stream composed by triplets of sounds presented at a fast stimulation rate (one sound every 500ms). Sounds were randomly drawn from a large set of different synthetic excerpts. Within the triplet, the first two sounds were identical (standards), while the third one (oddball) was a different one.

Two experiments were designed. In Local Discrimination, the oddball sound could vary for its local features with unaltered global structure; in Global Discrimination, the

oddball differed for global (and thus also local) properties as compared to standard sounds. As statistical variability is expected to vary as a function of sound duration (McDermott, Schemitsch, and Simoncelli, 2013), in separate blocks, stimuli of different lengths (40, 209 and 478ms) were presented. Participants were performing an orthogonal task so that any difference between experiments may be interpreted as automatic discriminative responses to the statistical variability in the oddball and the auditory mode elicited by it (local vs. global; see Materials and Methods).

MATERIALS AND METHODS

Participants

Twenty-four normal-hearing right-handed young adults (F= 12; mean age= 27.13 years, std= 2.83) took part in the experiment. This sample size was estimated through simulations. We used the procedure described in Wang and Zhang (2021; for details on sample size estimation see Supplementary Expanded Methods in appendix 1 and Figure S1). All participants were healthy, fully informed on the scopes of the experiment, signed a written informed consent prior to testing, and received a monetary compensation. The study was approved by the regional Ethical Committee (CEAVNO protocol n 24579) and protocol adhered to the guidelines of the Declaration of Helsinki (2013).

Experimental design and statistical analyses

Stimuli

Stimuli were synthesized with a previously validated computational auditory model (McDermott and Simoncelli, 2011; Figure 1A). A set of statistics which emulates basic auditory computations at the level of the cochlea (envelope, marginal moments and cross-correlation) and mid-brain (modulation power and two correlations) were extracted from original recordings of sound textures (Table 1) and imposed in an iterative way on four white noise samples. This allowed to generate different sounds varying for their local features but including the same long-term statistics (global structure). Synthetic sounds derived from the same original sound texture but built

on different white noise samples, were referred to as *exemplars* of the same sound texture and were perceptually very similar. As long-term statistics represent averages of local features over time (McDermott, Schemitsch, and Simoncelli, 2013), statistical similarity among exemplars of the same sound texture increases with duration. On the other hand, exemplars of different sound textures were derived from different original recordings and diverged for both local features and long-term statistics (Figure 1B and C).

This computational synthesis approach provides an unprecedented level of control over the statistics embedded in the sounds and allows to disentangle local features processing from their statistical averaging.

Synthesis toolbox is available here: http://mcdermottlab.mit.edu/downloads.html

To create experimental stimuli, exemplars were cut into excerpts of different lengths, equally spaced on a logarithmic scale, either 40, 209, and 478ms. Stimuli were equalized to the same root mean square (RMS= 0.1) and had a sampling rate of 20kHz.

Procedure

Participants were tested in an electrically shielded and sound isolated booth. After reading instructions on a monitor, they listened to the sound in the absence of retinal input (participants were blindfolded to prevent visual input).

A sound sequence lasting 108sec was presented. The sequence contained triplets of sounds (n= 216), presented one after the other to form an almost continuous sound stream in which each sound was comprised within a 500ms time-window (Figure 1D).

Two experiments were implemented. (1) In Local Discrimination, two sounds within a triplet were identical (standards), while the third one, oddball, comprised different temporal local features but same long-term statistics with respect to the other two; namely, it was an excerpt drawn from a different exemplar of the same sound texture. Thus, standards, and oddball were the same sound object but slightly differed in their local structure (e.g., both standards and oddball sounds are the sound of waterfall; see Table S1, column 1). (2) In Global Discrimination, oddball sound deviated from the other two for both its local features and global properties (auditory statistics) being it an excerpt extracted from an exemplar of a different sound texture compared to the other two identical sounds. This means that the

oddball was a different sound object compared to the standards (e.g., standard is the sound of a bulldozer, the oddball is waterfall).

All sounds were drawn from a large set of different synthetic excerpts of each original sound texture (Table S1). Sound Textures pairs were combined as in Table S1 (column 1 and 2).

In both experiments, excerpts were drawn from two exemplars randomly selected out of the four synthetic sounds originated from the four possible white noise samples. In this way, statistical variability within a triplet was kept constant but the sound streams were always different across blocks. In other words, stimuli *per se* were never the same between experiments, but the expected variability between standards and oddball was and discriminative responses emerging by the contrast of deviant and standard did not depend on specific properties (e.g., a change in frequency) but only on by their local or global statistical change. Moreover, to improve generalizability, the order of the triplets was shuffled for each participant. In sum, this allowed to test the processing (local features processing and/or global statistical averaging) independently from the specific stimuli that were presented, increasing generalizability of results.

Moreover, in the two experiments, the statistical variability between standard and oddball sounds changed as a function of sound duration in opposite directions: decreasing with duration in Local Discrimination, while increasing in Global Discrimination (Figure 1C).

Three different sound durations were used: short, intermediate, and long (40, 209, or 478ms, respectively). These sound lengths were chosen based on results in previous behavioral investigations (Berto et al., 2021; McDermott, Schemitsch, and Simoncelli, 2013).

A total of six conditions were employed, one for each of two experiments (Local and Global Discrimination) presented with stimuli of each of the three sound durations (40, 209, 478) separately. Two runs per condition (Experiment * Duration) were presented, for a total of twelve runs. The order of runs was randomized across participants and short breaks were foreseen in between runs. Disregarding sound duration, the number of standards was kept constant in each run (two standards in each triplet), to prevent potential differences in standard formation (i.e., Sussman and Gumeyuk, 2005) and expectancy effects from accounting for results.

Participants had to listen to the sound stream but were asked to perform an orthogonal task consisting of pressing a button when a beep sound appeared. The beep was a pure tone higher in pitch and intensity compared to the sound textures stream. The pure tone was 50ms in length, had a frequency of 2200Hz, an amplitude of 50dB, a sampling rate of 20kHz, and RMS of 5. The beep appeared randomly within the stimulation. The number of beeps varied randomly across runs, from zero to three. Detection was considered valid when participant pressed the key within an arbitrary window of 3 seconds from beep appearance (behavioral results are reported in the Supplementary Expanded Methods in appendix 1 and Figure S2).

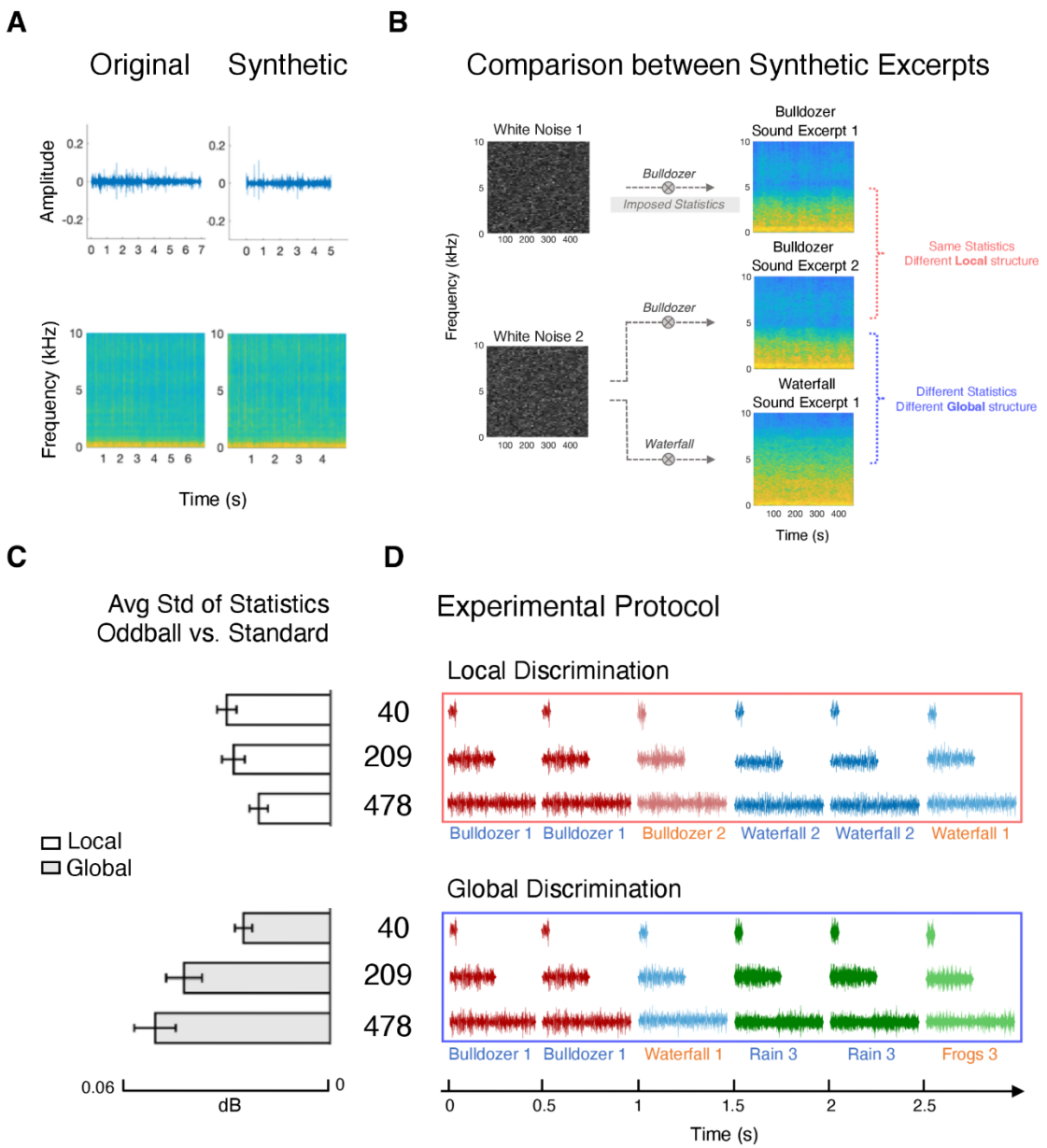


Figure 1. Synthetic stimuli and Experimental Protocol.
(A) Example of sound synthesis. Sound waveform of 7s original recording of a fire cracking sound (left panel) and of 5s synthetic excerpt originated from the same recording by extracting the set of auditory statistics with an Auditory Texture model (McDermott and Simoncelli, 2011). Spectrograms of each sound are displayed below for comparison. (B) Auditory statistics extracted from two original recordings (in this example, waterfall and bulldozer) are imposed to two white noise samples (their spectrograms are represented on the left). The outcomes of the imposition are displayed as spectrograms on the right. The two synthetic sound excerpts containing the *same set of statistics* have the same global structure but differ for their local properties as they are derived from different white noise samples. On the other hand, when *different statistics* are imposed on the same white noise both local and global structures vary. Stimuli with the same auditory statistics will be perceptually very similar and identified as the same sound object (i.e., both are instances of a bulldozer sound); different set of auditory statistics will generate a different sound object (i.e., bulldozer in one case, waterfall in the other). (C) Average standard deviation of auditory statistics between oddball and standard for each experiment at each sound duration. Variability increases with duration in Global Discrimination, as global structure emerges, and decreases in Local. (D) Experimental protocol. Triplets of sounds were presented at a fast rate (one sound every 500ms). Two sounds were identical (standard), while the third was different (oddball) and could vary for its local or its global properties, depending on the experiment (Local or Global Discrimination). Since statistical variability changes as a function of sound length, three sound durations equally spaced logarithmically (40, 209, and 478ms) were employed to tap into each auditory mode separately (local vs. global processing).

EEG recording

Electroencephalography (EEG) was recorded from an EGI's HydroCel Geodesic Sensor Net with 65 EEG channels and a Net Amps 400 amplifier (Electrical Geodesics, Inc., EGI, USA). Acquisition was obtained via EGI's Net Station 5 software (Electrical Geodesics, Inc., EGI, USA). Central electrode E65 (Cz) was used as reference. Four electrodes were located above the eyes and on the cheeks to capture eye movements. Electrode impedances were kept below 30 kΩ. Continuous EEG signal was recorded throughout the whole session with a sampling rate of 500Hz.

Experimental sounds were played from a stereo speaker (Bose Corporation, USA) positioned in front of the participant and at one meter distance from their eyes; sounds loudness was kept constant across participants and runs. The experiment ran on MATLAB (*R2018b*; Natick, Massachusetts: The MathWorks Inc.); written instructions were displayed only at the beginning of the experimental session, via Psychtoolbox version 3 (Brainard and Vision, 1997; PTB-3; http://psychtoolbox.org/).

EEG Data Analysis

Preprocessing

Data were preprocessed with a semi-automatic pipeline implemented in MATLAB (see Stropahl et al., 2018; Bottari et al., 2020). Preprocessing was performed using EEGLAB (Delorme and Makeig 2004; https://sccn.ucsd.edu/eeglab/index.php). Data were loaded excluding electrode E65 (Cz) which was the reference channel of our EEG setup (thus included only zero values).

To remove slow drifts and DC offset, a high-pass filter (windowed sinc FIR filter, cut-off frequency 0.1 Hz, filter order 500) was applied to the continuous signal. Signal was then re-sampled to 250Hz to reduce computational time of further analyses steps.

A first segmentation in time was performed, by epoching the signal according to event onset. To avoid boundary artifacts, for each run, the signal was cut two seconds before its onset event and until two seconds after the end of the presentation (thus, from -2 to +114 sec). For each participant, epochs were then merged in a single file containing only the parts of the signal referring to significant stimulation (thus excluding breaks in between trials).

Independent Component Analysis (ICA) (Bell and Sejnowski, 1995; Jung et al., 2000a,b) was used to attenuate artifacts. To improve ICA decomposition, beforehand data were high-pass filtered (windowed sinc FIR filter, cut-off frequency 1Hz, filter order 500), low-pass filtered (windowed sinc FIR filter, cut-off frequency 40Hz, filter order 100) and segmented into consecutive dummy epochs of 1sec in order to spot non-stereotypical artifacts. Epochs with joint probability larger than 3 standard deviations were rejected (Bottari et al., 2020). PCA rank reduction was not applied prior to ICA, to avoid compromising its quality and effectiveness (Artoni, Delorme, and Makeig, 2018).

For each subject, ICA weights were computed using EEGLAB runica algorithm and assigned to the corresponding original raw (unfiltered) dataset. Topographies for each component were plotted for visual inspection. Artefacts associated to eye movements and blinks were expected. CORRMAP algorithm (Viola et al., 2009) was used to semi-automatically remove components associated to such artefacts. Component representing eye movements and blinks were manually identified from a single dataset and used as template. CORRMAP algorithm clusters ICA components

with similar topography across all datasets to highlight the similarity between the IC template and all the other ICs. A correlation of the ICA inverse weights was computed, and similarity allocated, with a threshold criterion of correlation coefficient being higher or equal to 0.8 (default value of CORRMAP; Viola et al., 2009). For all participants, on average 1.79 components were removed (std= 0.72; range= 0-3). Cleaned data were re-loaded and bad channels were interpolated after visual inspection. For most participants, a maximum of four channels were interpolated, except for one participant in which several noisy channels required interpolation (mean number of interpolated electrodes per participant= 1.67; std= 2.01; range= 0-8). Bad channels were interpolated via superfast spherical interpolation implemented in EEGLAB.

Finally, the channel E65 (Cz) was included back in the EEG data of each participant and datasets were re-referenced to average reference across channels.

Time-domain analysis

Preprocessed data were low pass filtered (windowed sinc FIR filter, cut-off frequency= 40Hz, filter order= 100). Consecutive epochs (from -0.1 to 0.5sec) were generated, including segments of either the oddball sounds or the second standard of the triplets for each participant and condition. Joint probability was used to prune non-stereotypical artifacts (i.e., sudden increment of muscular activation); rejection threshold was 4 standard deviations (Stropahl et al., 2018). For oddballs, on average 10.51 epochs per participant were removed (std=3.55; range 3-22) out of the 144 concatenated epochs that each Experiment * Duration comprised; for standards, on average 10.71 epochs were removed (std= 3.81; range 3-21), again out of 144 trials per each condition.

Datasets were converted from EEGLAB to FieldTrip (Oostenveld, Fries, Maris, and Schoffelen, 2011; http://fieldtriptoolbox.org). Grand averages across participants were computed for each experiment, duration, and stimulus type (either standard or oddball). Data across trials were averaged to compute event-related potentials (ERPs) at central electrodes (Cz and FCz). Before displaying results, the data were baseline corrected using the −0.1 to 0 sec pre-stimulus period.

Since all stimuli in the triplets were never the same across runs, we assessed whether differences exist between experiments (Local and Global) at sound duration by subtracting from the response to the oddball sound the response to the preceding

standard in each triplet. This was done to ensure that any potential difference in the perceived oddball sound between experiments, for each sound duration, was not driven by idiosyncratic differences in the stimuli that were presented in that specific run, but by the statistical difference to their preceding standards.

ERPs and grand-average topographies to oddball and standard sounds (prior to subtraction) are also available in the Supplementary information (Figure S3).

A non-parametrical permutation test was performed to check significances between experiments (Local vs. Global), for each duration (short, intermediate, and long) in the so-formed oddball – standard epochs. Each participant's data were summarized in an array of condition-specific averages (experiment*duration). Permutation test was carried out under the null hypothesis that probability distributions across condition-specific averages were identical across experiments. Noticeably, statistical analysis between experiments were performed only within duration, to control for refractoriness effects related to different interstimulus intervals (ISI) at long and short durations.

Thus, the contrasts of interest were: (1) Local short vs. Global short; (2) Local intermediate vs. Global intermediate; (3) Local long vs. Global long.

A series of cluster-based permutation tests (Maris and Oostenveld, 2007; cluster alpha threshold of 0.05 (two-tailed); 1000 permutation; minimum neighboring channels = 2) was performed. Cluster-based permutation analyses were performed without a priori assumptions regarding specific regions of interest, thus, including all channels. Based on extensive literature on auditory change detections, effects of Local and Global Discriminations were expected to occur around N1 and P200 components of auditory evoked potentials (see Näätänen et al., 2007 for review). Thus, cluster-based permutations were carried including all timepoints within the 0.1 to 0.3 s time-window. Significant positive clusters would indicate grater response at sensor level to Local as compared to Global Discrimination; significant negative clusters would imply the opposite trend, with grater response to Global as compared to Local.

Source Reconstruction

Source reconstruction was performed using the Brainstorm toolbox (Tadel et al. 2011; http://neuroimage.usc.edu/brainstorm) and following a validated pipeline (Stropahl et al., 2018; Bottari et al., 2020). This analysis was driven by results in the

time-domain at the sensor level, thus only performed for those conditions which yielded to significant differences after cluster-based permutation (see below and Figure 2). Preprocessed, epoched, and pre-stimulus baseline subtracted datasets were imported in Brainstorm. Electrode location was assigned, projected on the scalp, and MRI registered to a default anatomy (ICBM152; Fonov et al. 2009).
We computed noise covariance matrix (using a pre-stimulus period between -0.1 and -0.002 sec) and we averaged all epochs for each participant. The cortex surface was computed using OpenMEEG BEM (Gramfort et al. 2010); since default anatomy was used, this was done only for the first participant, and subsequently the head model was applied to all others. Sources were extracted using the minimum norm imaging inverse method and the dSPM inverse measure (Dale et al. 2000) with constrained (normal to cortex) dipole orientations. Subtraction (oddball – standard) was then performed in the source domain.
To assess significant voxels, we performed non-parametric tests (Pantazis et al., 2005a). Directionality of the tests and time window within which perform the statistics were driven by results in cluster-based permutation performed at the sensor level, that is to address the specific hypotheses yielded from time-domain analysis. Wilcoxon signed ranked test (MacFarland and Yates, 2016) was performed in Brainstorm to compare the two datasets at the whole brain level (across all voxels) to calculate t-statistics with a critical alpha of 0.05 (one-tailed). A number of 1000 permutations was implemented. Results were time-corrected for False Discovery Rate (FDR; Benjamini and Yekutieli, 2001); a cortical activation was considered only if above a minimum voxel-size of 20mm. Activity-based ROIs were defined after whole brain analysis. ROI selection was both data-driven according to the peak of activity and based on previous knowledge on auditory cortex organization and functioning.

RESULTS

Time-domain results

Cluster permutation revealed significant positive cluster for the short duration 40 ($p = 0.02$), lasting between 192ms and 224ms after stimulus onset. No significant cluster was found for the intermediate duration 209 (all $p > 0.17$), while a significant negative

cluster was found for the long duration 478 ($p < 0.001$) lasting between 196ms and 300ms after stimulus onset. Grand-average topographies at relevant latencies are displayed in Figure 2A. For visualization, we selected two central channels (Cz and FCz) and plotted the grand-average event-related potentials (ERPs) of oddball minus standard for all durations in each experiment (Figure 2B). These results support the hypotheses based on behavioral data (i.e., Berto et al., 2021; McDermott, Schemitsch, and Simoncelli, 2013) for which local processing is favored for brief sound excerpts while global processing emerges at longer durations and show that the same can hold even when participants are directly engaged to discriminate them. Discriminative responses can emerge automatically (pre-attentively) from stimulus properties as the amount of acoustic information changes.

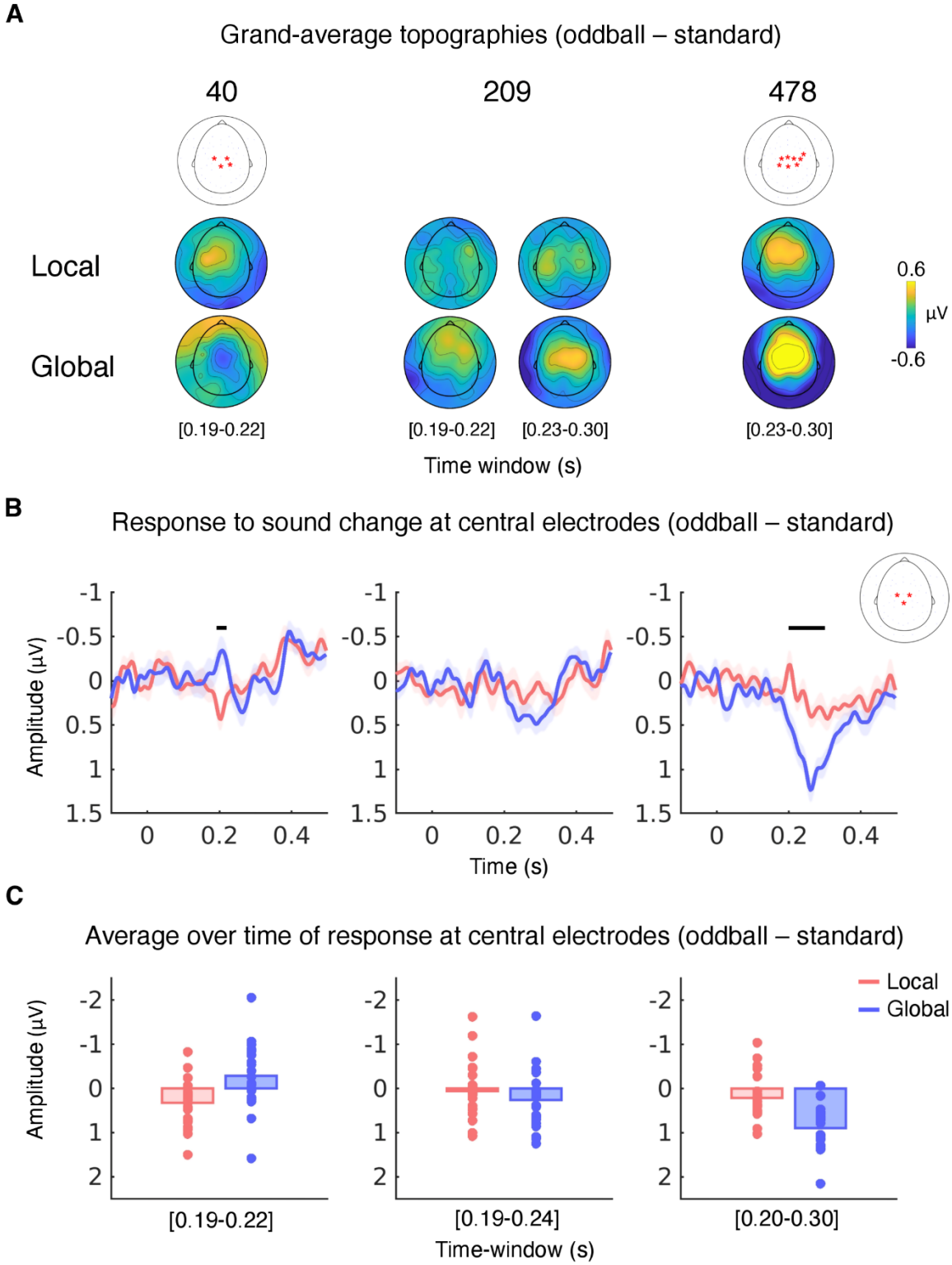


Figure 2. Results of time-domain analysis at sensor level.
(A) Grand-average topographies of the differential response associated with the sound change (oddball sound minus standard sound) at significant latencies, for each experiment and duration. For each latency, electrodes associated with significant clusters are displayed above as red stars on the scalp. * $p < 0.025$.
(B) Grand-average electrical activity of the differential response (oddball minus standard) at three central electrodes (E54, E7, E65, in red) which were significant for both short (40) and

long (478) duration. Shaded regions show interpolated standard error of the mean (SE) at each time point. Positive values indicate that oddball elicited greater response than standard. Results of cluster permutation are displayed as black bars extending through significant latencies. – *p < 0.025.*

(C) Grand-average amplitude of the differential response (oddball minus standard) across participants at the same three significant electrodes averaged within significant latencies for each experiment and duration. For visualization only, since no significant latencies existed for the duration 209, an intermediate time window was used, by computing an average between the time windows of significant latencies for duration 40 and 478. Bar plots represent average values across participants. Individuals' values are separately displayed by each point.

Source reconstruction results

Provided results at the sensor level, we investigated Local vs. Global for the duration 40 and 478. Specifically, for duration 40 we tested in which source estimates the activation in Local Discrimination was higher compared to Global Discrimination (H1= Local > Global) within the time window revealing the significant effect at the sensor level (0.192-0.224s; see Figure S2A); by contrast, for duration 478 we tested which sources had an activity which was higher for Global Discrimination as compared to Local (H1 = Global > Local) at latencies from 0.196 to 0.3 s (the time window comprising the significant effect at the sensor level).

At duration 40, whole-brain analysis showed significant voxels located around anterior superior temporal sulcus (STS) in the right hemisphere (Figure 3A, left panel). An activity-based ROI was drawn around the peak of activation at 208ms (Figure 3B, left panel).

At duration 478, the same procedure yielded to significant voxels around right medial superior temporal gyrus (STG), the anterior part of left STS, and left Heschl's gyrus corresponding to early auditory cortex (Figure 3A, right panel), with a peak of activation at 236ms (Figure 3B, right panel).

These results provide evidence that local and global processing represent alternative processes relying on partially different networks within the auditory system. These networks are inversely employed as a function of sound duration and stimulus properties.

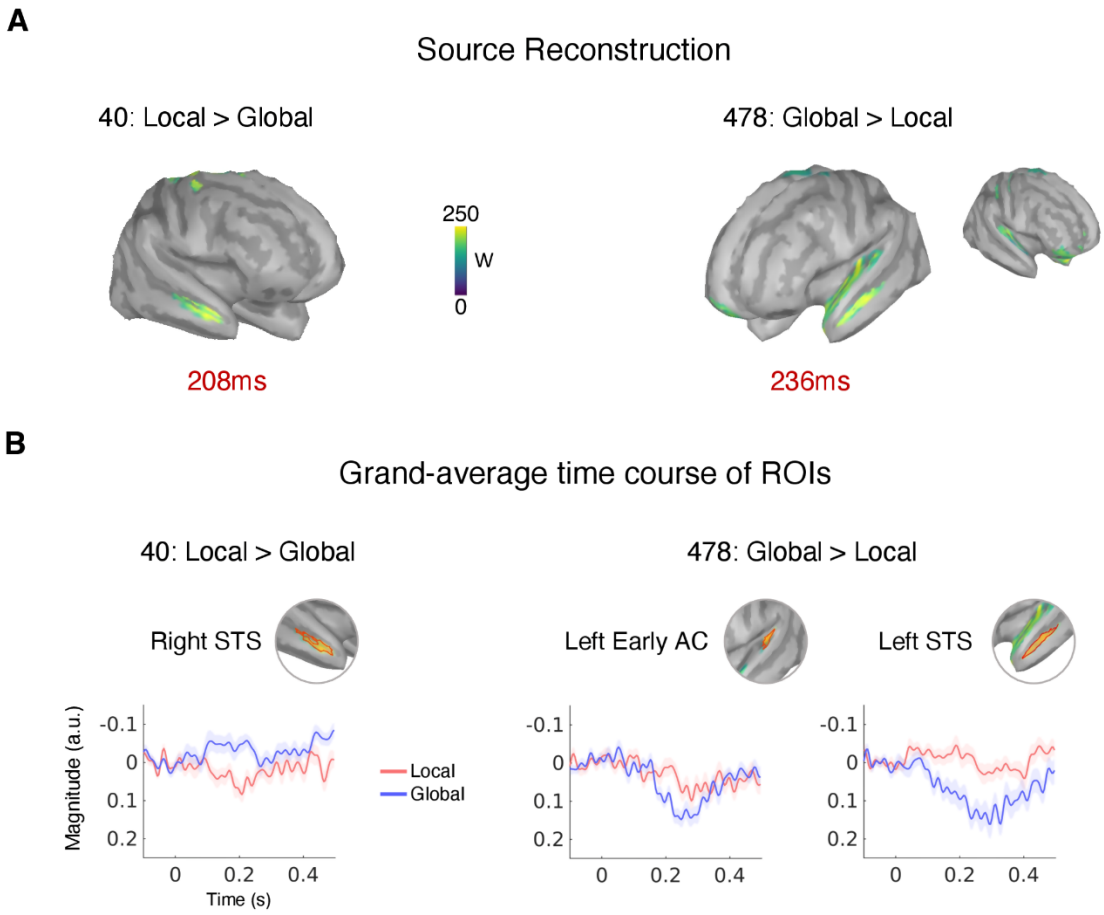


Figure 3. Results of Source Reconstruction.
(A) Group-level results of non-parametric permutation test performed within significant time windows at the whole brain level. Results were corrected for false discovery rate (FDR).
(B) Three ROIs were identified. Grand-average time course of source activity over the entire time window (0-0.5s) are plotted for each ROI. Shaded regions show interpolated standard error of the mean (SE) at each point.

DISCUSSION

The auditory system operates at different timescales extracting information at high (local) and low (global) temporal resolution. Global structure emerges at longer timescales, as the system extracts abstract information from local features. Here we aimed to (1) test the automaticity of sound discriminations based on local and global computations and (2) and understand if discriminative responses driven by local and global processes subtend to different neural correlates. We employed a computational model to control auditory statistics and synthetically create stimuli with the same global structure but emerging from different local features. We presented a

stream of sound triplets containing two consecutive standards followed by an oddball; oddballs could vary from standards for its local or global structure; while stimuli were equalized to the same RMS, local differences were more prominent at shorter durations and global differences at longer ones. Compatibly, results showed that when sound duration was short, changes in local properties elicited greater neural responses than different global ones. When duration of sounds was longer, the change in global structure elicited a higher response that was not detectable when only their local features changed (Figure 2A, B, C). The opposite trend in the neural response observed according to sound duration (see Figure 2C) provides evidence that the local or global processes are alternative and, pre-attentively selected merely based on the amount of information provided in the upcoming stimulation. Source modeling showed different putative underlying neural correlates distinguishing the two computational modes (local and global) and hemispheric asymmetries. The observed neural correlates indicated not only that a change had occurred, but also the nature of the change. Importantly, these responses could be measured without participants being asked to pay attention to any stimulus property.

Automaticity of Local and Global Processing

Our findings show that different subprocesses (local vs. global) exist and are specifically tuned to detect different types of sound changes. The engagement of one processing strategy or the other occur automatically and is driven by properties of the change of sound. In fact, participants were engaged in an orthogonal task while exposed to the sound stream and were not instructed to perform any related task. Strikingly, the directionality and magnitude of the neural response was the same as in previous psychophysical investigations (Figure 2C; McDermott, Schemitsch, and Simoncelli, 2013; Berto et al., 2021) when participants were attentive to the sounds and performed deliberate judgments over two alternative forced choice (2AFC).

Attention plays a role in determining to which temporal scale the auditory system attend to (Lange, Roesler, and Roeder, 2003; Sanders and Astheimer, 2008). For instance, in a task involving rhythmic sound trains, listeners who were instructed to pay attention to long-term variations showed different patterns of responses as compared to those who paid attention to the global structure of the sound stream (e.g., Large and Jones, 1999). The auditory system switches between time scales

514 and what the listeners are paying attention to influences sound discrimination and
515 sensory processing. (Hink and Hillyard, 1976; Schwent et al., 1976; Hansen and
516 Hillyard, 1983; Lange, Roesler, and Roeder, 2003; Sanders and Astheimer, 2006).
517 Attention over selective features also engages different hemispheres, according to
518 the timescale at which participants are expecting sound changes to occur
519 (Brechmann and Scheich, 2005).
520 Nonetheless, in our experiment, we showed that the properties implicit in the sound
521 are sufficient to automatically elicit separate discriminative responses (Figure 1B, C,
522 D; Figure 2). Noticeably, the present results emerged despite triplets were
523 continuously changing (the only fixed parameter was either the local or the global
524 change) and thus can be generalized to multiple sound textures (Figure 1; see also
525 Supplementary Information, Table S1). These results suggest that the auditory
526 system is capable to automatically attune to specific type of sound changes, to
527 detect stimuli variances in the most efficient way and proved that this effect emerges
528 independently from attentional control. This expands seminal studies disclosing the
529 MisMatch Negativity response (MMN; Näätänen et al., 1978; Tiitinen et al., 1994), a
530 neural marker revealing automatic auditory change-detections. MMN is the outcome
531 of a process by which the system “scans” for regularities in the entering sounds and
532 uses them as reference to detect variations within the auditory scene (for reviews
533 see Näätänen et al., 2001, 2010).
534 MMN response to a deviant sound is not affected by prior expectations that the
535 oddball will occur (Rinne et al., 2001). In our study, expectations that a change would
536 occur in the third element of the triplet had a probability of 1 in each experiment
537 (Local and Global Discrimination; Figure 1D). Thus, our results cannot be simply
538 explained by spurious attentional effects. Differences between oddball responses,
539 compared to the previous standard, in our study can only be due to pre-attentive
540 discriminative responses to the properties of the deviant sound.
541 These results add up to the auditory MMN, which set the basis for the existence of a
542 primitive intelligence in the auditory system (for review see Näätänen et al., 2001).
543 Here, we showed that the auditory system not only orients the attention towards
544 relevant, deviant events in the environment but is also capable to automatically
545 categorize such acoustic difference according to the level of change and the
546 timescales at which deviation occurred.
547

Local and Global Discrimination subtend to alternative processes

Electrophysiological signatures of local and global sound representations automatically emerged from our protocol as a function of sound duration, advocating for the -at least partial- independence of the two computations (Figure 2).

As global representations are built upon local features (McDermott and Simoncelli, 2011; McDermott, Schemitsch, and Simoncelli, 2013), we could have expected that the two computations were cumulative processes, with global representation adding on top of local one and pertaining to similar neural activations. However, the processing of information at local or global timescale might differ not only quantitatively but also qualitatively (e.g., the response to small changes of pitch at global timescales -tend to be faster and more accurate as compared to the detection of a large pitch change occurring at a fast rate; Sanders and Poeppel, 2007). When the sensory information load is prohibitive, high temporal resolution information tends to get lost in the averaging process. Thus, since local features are not retained after averaging (McDermott, Schemitsch, and Simoncelli, 2013), it checks out that the two computations represent alternative processes, and that the system switches from one to the other depending on the intrinsic sound properties (e.g., statistical variability) and the amount of available information in the auditory stream (e.g., increased stimulation length). Recent evidence in the animal model showed that auditory statistics ensembles are represented by distinct responses at the level of the midbrain (Zhai et al., 2020). These neural representations to auditory statistics occur at dissociable time scales, permitting the discrimination and recognition of natural sounds, with local information supporting sounds differentiation while the emergence of global structure their recognition (Zhai et al., 2020). In our study, whole brain results in the source domain (Figure 3) support this suggestion of dissociable neural responses to local and global statistical properties. By comparing local and global discrimination at short duration we showed that discrimination based on local features entrained more the anterior part of the STS in the right hemisphere. Conversely, the same contrast at long durations linked the emergence of global discrimination predominantly to the left hemisphere; precisely response peaked both around Heschl's gyrus, corresponding to early auditory cortex, and in the STS and STG; Figure 3A). The hemispheric asymmetry emerging from source reconstruction falls into place with previous observations about the existence of an asymmetric

sampling of acoustic temporal information in right and left auditory cortices (Flinker et al., 2019).

The right STG has been identified as a potential generator for auditory MMN (Optiz et al., 2002; Doeller et al., 2003; Molhom et al., 2005). Here, the entrainment of right STS for the short stimuli is in line with a recent model suggesting a right hemisphere dominance for the processing of spectral content presented at a slow rate (Flinker et. al., 2019) and could account for higher sensitivity to local-acoustic-features-change detection in deviant sounds.

Previously, right hemisphere was found to be more sensible to global structure changes (Brechmann, Baumgart, & Scheich, 2002; Poeppel et al., 2004). This apparent discrepancy with the present results could be explained by considering that other studies addressed global processing by presenting variations of acoustic features at longer timescales (e.g., tracking slow rate modulation changes).

By contrast, for long durations, a change in global structure elicited greater neural response in the left hemisphere. This result may be surprising, as the left hemisphere has often been associated to the processing of fine-grained temporal details (Fiez et al., 1995; Johnsrude et al., 1997; Belin et al., 1998; Celsis et al., 1999) which perhaps underly its specialization in processing linguistic features (Tallal, Miller, & Fitch, 1993). However, a response in the left STS is in line with evidence appointing left STS as an area involved in sound object categorization (i.e., Liebenthal et al., 2010). Strikingly, we reported a peak of activity over the anterior and lateral parts of left Heschl gyrus and the anterior part of left STG (see Figure 3A). These areas have been previously referred to as the “what” auditory pathway, a network involved in sound identity (see Ahveninen et al., 2006; Bizley et al., 2013). Consequently, their involvement in global processing further supports that, as sound duration increases, auditory statistics emerge allowing for a different sound-object to be automatically detected (McDermott and Simoncelli, 2011). Interestingly, previous evidence showed that left STG can be compartmentalized into sub-regions partaking to different levels of processing, with the anterior part being associated with the integration of phonemes into more complex temporal patterns, such as words (DeWitt and Rauschecker, 2012). This specialization could extend to naturalistic sounds processing, beyond speech, and could represent a general abstraction mechanism to derive global abstract structure from local acoustic elements

CONCLUSION

Seminal studies on the auditory mismatch negativity (MMN) provided the basis for the notion of a primitive intelligence in the auditory system, dedicated to alerting and orienting towards changes of regularities in the environment. By combining electrophysiology with a computational approach, we revealed the cortical representations of the auditory system's ability to automatically differentiate changing sounds, depending on their statistical variability and amount of available information in the auditory stream (e.g., sound duration). Different cortical responses emerged as a function of specific stimulus-driven auditory representations, revealing different networks that represent discriminative responses to local vs. global changes. These responses appeared to occur pre-attentively and to be alternative to one another. These results open the possibility to use neural markers associated with local and global discriminations for further studying specific auditory computations in both basic and applied research.

ACKNOWLEDGMENTS

We thank all the students who helped with recruiting participants and/or data collection: Nicolò Castellani, Irene Sanchez, Chiara Battaglini, and Dila Suay.
Funding: Davide Bottari (PRIN 2017 research grant. Prot. 20177894ZH).

CONFLICT OF INTEREST

The authors declare no competing interest.

FIGURES TITLES AND LEGENDS

Figure 1. Synthetic stimuli and Experimental Protocol.
(A) Example of sound synthesis. Sound waveform of 7s original recording of a fire cracking sound (left panel) and of 5s synthetic excerpt originated from the same recording by extracting the set of auditory statistics with an Auditory Texture model (McDermott and Simoncelli, 2011). Spectrograms of each sound are displayed below

for comparison. (B) Auditory statistics extracted from two original recordings (in this example, waterfall and bulldozer) are imposed to two white noise samples (their spectrograms are represented on the left). The outcomes of the imposition are displayed as spectrograms on the right. The two synthetic sound excerpts containing the *same set of statistics* have the same global structure but differ for their local properties as they are derived from different white noise samples. On the other hand, when *different statistics* are imposed on the same white noise both local and global structures vary. Stimuli with the same auditory statistics will be perceptually very similar and identified as the same sound object (i.e., both are instances of a bulldozer sound); different set of auditory statistics will generate a different sound object (i.e., bulldozer in one case, waterfall in the other). (C) Average standard deviation of auditory statistics between oddball and standard for each experiment at each sound duration. Variability increases with duration in Global Discrimination, as global structure emerges, and decreases in Local. (D) Experimental protocol. Triplets of sounds were presented at a fast rate (one sound every 500ms). Two sounds were identical (standard), while the third was different (oddball) and could vary for its local or its global properties, depending on the experiment (Local or Global Discrimination). Since statistical variability changes as a function of sound length, three sound durations equally spaced logarithmically (40, 209, and 478ms) were employed to tap into each auditory mode separately (local vs. global processing).

Figure 2. Results of time-domain analysis at sensor level.

(A) Grand-average topographies of the differential response associated with the sound change (oddball sound minus standard sound) at significant latencies, for each experiment and duration. For each latency, electrodes associated with significant clusters are displayed above as red stars on the scalp. * $p < 0.025$.

(B) Grand-average electrical activity of the differential response (oddball minus standard) at three central electrodes (E54, E7, E65, in red) which were significant for both short (40) and long (478) duration. Shaded regions show interpolated standard error of the mean (SE) at each time point. Results of cluster permutation are displayed as black bars extending through significant latencies. – $p < 0.025$.

(C) Grand-average amplitude of the differential response (oddball minus standard) across participants at the same three significant electrodes averaged within significant latencies for each experiment and duration. For visualization only, since

no significant latencies existed for the duration 209, an intermediate time window was used, by computing an average between the time windows of significant latencies for duration 40 and 478. Bar plots represent average values across participants. Individuals' values are separately displayed by each point.

Figure 3. Results of Source Reconstruction.

(A) Group-level results of non-parametric permutation test performed within significant time windows at the whole brain level. Results were corrected for false discovery rate (FDR).

(B) Three ROIs were identified. Grand-average time course of source activity over the entire time window (0-0.5s) are plotted for each ROI. Shaded regions show interpolated standard error of the mean (SE) at each point.

SUPPLEMENTARY EXPANDED METHODS (Appendix 1)

Sample size

The number of participants was estimated through simulations. We used the procedure described in Wang and Zhang (2021) and simulated a dataset with two conditions (Local and Global) of ERP data. First, we selected a group of electrodes of interest at central locations (E7, E65, E54; see Figure 2B and S2B in red) which typically capture auditory responses. For the simulation, we selected a time window between 0.1 and 0.3s, based on previous MMN studies (see Näätänen et al., 2007 for review). The amplitude values at the electrodes of interest for the two conditions were sampled from a bivariate normal distribution (within-subject design), which mean and standard deviation were chosen based on results of four pilot datasets (mean Local= 0.16; mean Global= 0.56; sd Local= 0.52; sd Global= 0.54).

We then ran cluster-based permutation on simulated datasets, to test whether any statistical cluster (t-values) exhibited significant difference between the two conditions with an alpha level of 0.05. The procedure starts with a sample size of 10 and increases in steps of 1 until it reaches a power of 0.80. We run 1000 simulations for each sample size and calculated the power as the proportion of the number of times finding significant clusters in these 1000 simulations. The simulation results

showed that in order to obtain a power above 0.8, a sample size of N= 24 was required (see Figure S1).

The algorithm to perform such analyses can be downloaded from this link: https://osf.io/rmqhc/.

Behavioral Results

For each condition, percentage of correct beep detections was above 90% (Local 40: mean= 0.99, std= 0.03; Local 209: mean= 0.99, std= 0.05; Local 478: mean= 1, std= 0; Global 40: mean=0.99, std= 0.05; Global 209: mean= 0.97, std=0.08; Global 478: mean= 0.97, std= 0.11; Figure S2). We run a two-way ANOVA for repeated measures with factors Experiment (2 levels, Local vs. Global) and Duration (3 levels, 40, 209, and 478) to address whether experiment type and stimulus length impact on beep detection and participants' attention to the task. No significant main effects were observed (Experiment, $F_{(1,23)}= 3.62$, $p =0.07$, $n^2= 0.14$; Duration, $F_{(2,46)}= 0.58$, $p= 0.56$, $n^2= 0.3$), nor their interaction (Experiment*Duration, $F_{(2,46)}= 0.45$, $p= 0.64$, $n^2= 0.2$).

These behavioral results provide evidence that participants were attentive and responsive during sound presentation throughout the experiment and that attention to this orthogonal task was not influenced by experiment and sound duration.

REFERENCES

McAdams, S. (1993). Recognition of sound sources and events. Thinking in sound: The cognitive psychology of human audition, 146-198.

Griffiths, T. D. (2001). The neural processing of complex sounds. Annals of the New York Academy of Sciences, 930(1), 133-14

Plomp, R. (1964). Rate of decay of auditory sensation. The Journal of the Acoustical Society of America, 36(2), 277-282.

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. The Journal of the Acoustical Society of America, 102(5), 2892-2905.

McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. Nature neuroscience, 16(4), 493-498.

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. Psychophysiology, 35(5), 615-619.

Saint-Arnaud, N., Popat, K. (2021). Analysis and synthesis of sound textures D.F. Rosenthal, H.G. Okuno (Eds.), Computational Auditory Scene Analysis, CRC Press (2021), pp. 293-308

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages Sensory Communication ed WA Rosenblith. Cambridge, MA: MIT Press), 7, 1-34.

McDermott, J. H., & Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. Neuron, 71(5), 926-940.

McWalter, R., & McDermott, J. H. (2018). Adaptive and selective time averaging of auditory scenes. Current Biology, 28(9), 1405-1418.

Berto, M., Ricciardi, E., Pietrini, P., & Bottari, D. (2021). Interactions between auditory statistics processing and visual experience emerge only in late development. Iscience, 24(11), 103383.

World Medical Association. (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. Jama, 310(20), 2191-2194.

Wang, C., & Zhang, Q. (2021). Word frequency effect in written production: Evidence from ERPs and neural oscillations. Psychophysiology, 58(5), e13775.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clinical neurophysiology, 118(12), 2544-2590.

Sussman, E. S., & Gumenyuk, V. (2005). Organization of sequential sounds in auditory memory. Neuroreport, 16(13), 1519-1523.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. Spatial vision, 10(4), 433-436.

Stropahl, M., Bauer, A. K. R., Debener, S., & Bleichner, M. G. (2018). Source-Modeling auditory processes of EEG data using EEGLAB and brainstorm. Frontiers in neuroscience, 12, 309.

Bottari, D., Bednaya, E., Dormal, G., Villwock, A., Dzhelyova, M., Grin, K., … & Röder, B. (2020). EEG frequency-tagging demonstrates increased left hemispheric involvement and crossmodal plasticity for face processing in congenitally deaf signers. NeuroImage, 223, 117315.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of neuroscience methods, 134(1), 9-21.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. Neural computation, 7(6), 1129-1159.

Jung et al., 200 a,b
Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V., & Sejnowski, T. J. (2000a). Removing electroencephalographic artifacts by blind source separation. Psychophysiology, 37(2), 163-178.

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000b). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clinical Neurophysiology, 111(10), 1745-1758.

Artoni, F., Delorme, A., & Makeig, S. (2018). Applying dimension reduction to EEG data by Principal Component Analysis reduces the quality of its subsequent Independent Component decomposition. NeuroImage, 175, 176-187.

Viola, F. C., Thorne, J., Edmonds, B., Schneider, T., Eichele, T., & Debener, S. (2009). Semi-automatic identification of independent components representing EEG artifact. Clinical Neurophysiology, 120(5), 868-877.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational intelligence and neuroscience, 2011.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. Journal of neuroscience methods, 164(1), 177-190.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis computational intelligence and neuroscience. *Hindawi Publishing Corporation*, 1-13.

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage, (47), S102.

Gramfort, A., Papadopoulo, T., Olivi, E., & Clerc, M. (2010). OpenMEEG: opensource software for quasistatic bioelectromagnetics. Biomedical engineering online, 9(1), 1-20.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron, 26(1), 55-67.

Pantazis, D., Nichols, T. E., Baillet, S., & Leahy, R. M. (2005). A comparison of random field theory and permutation methods for the statistical analysis of MEG data. Neuroimage, 25(2), 383-394.

MacFarland, T. W., & Yates, J. M. (2016). Wilcoxon matched-pairs signed-ranks test. In Introduction to Nonparametric statistics for the biological sciences using R (pp. 133-175). Springer, Cham.

Benjamini, Y. & Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics. 29(4), 1165-1188.

Lange, K., Rösler, F., & Röder, B. (2003). Early processing stages are modulated when auditory stimuli are presented at an attended moment in time: An event-related potential study. Psychophysiology, 40(5), 806-817.

Sanders, L. D., & Astheimer, L. B. (2008). Temporally selective attention modulates early perceptual processing: Event-related potential evidence. Perception & psychophysics, 70(4), 732-742.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. Psychological review, 106(1), 119.

Hink, R. F., & Hillyard, S. A. (1976). Auditory evoked potentials during selective listening to dichotic speech messages. Perception & Psychophysics, 20(4), 236-242.

Schwent, V. L., Hillyard, S. A., & Galambos, R. (1976). Selective attention and the auditory vertex potential. I. Effects of stimulus delivery rate. Electroencephalography and Clinical Neurophysiology, 40(6), 604-614.

Hansen, J. C., & Hillyard, S. A. (1983). Selective attention to multidimensional auditory stimuli. Journal of Experimental Psychology: Human Perception and Performance, 9(1), 1.

Brechmann, A., & Scheich, H. (2005). Hemispheric shifts of sound representation in auditory cortex with conceptual listening. Cerebral Cortex, 15(5), 578-587.

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. Acta psychologica, 42(4), 313-329.;

Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. Nature, 372(6501), 90-92.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). ‘Primitive intelligence’in the auditory cortex. Trends in neurosciences, 24(5), 283-288.

Näätänen, R., Astikainen, P., Ruusuvirta, T., & Huotilainen, M. (2010). Automatic auditory intelligence: An expression of the sensory–cognitive core of cognitive processes. Brain research reviews, 64(1), 123-136.

Rinne, T., Antila, S., & Winkler, I. (2001). Mismatch negativity is unaffected by top-down predictive information. NeuroReport, 12(10), 2209-2213.

Sanders, L. D., & Poeppel, D. (2007). Local and global auditory processing: Behavioral and ERP evidence. *Neuropsychologia*, *45*(6), 1172-1186.

Zhai, X., Khatami, F., Sadeghi, M., He, F., Read, H. L., Stevenson, I. H., & Escabí, M. A. (2020). Distinct neural ensemble response statistics are associated with recognition and discrimination of natural sound textures. Proceedings of the National Academy of Sciences, 117(49), 31482-31493.

Flinker, A., Doyle, W. K., Mehta, A. D., Devinsky, O., & Poeppel, D. (2019). Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. Nature human behaviour, 3(4), 393-405.

Opitz, B., Rinne, T., Mecklinger, A., Von Cramon, D. Y., & Schröger, E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. Neuroimage, 15(1), 167-174.

Doeller, C. F., Opitz, B., Mecklinger, A., Krick, C., Reith, W., & Schröger, E. (2003). Prefrontal cortex involvement in preattentive auditory deviance detection:: neuroimaging and electrophysiological evidence. Neuroimage, 20(2), 1270-1282.

Molholm, S., Martinez, A., Ritter, W., Javitt, D. C., & Foxe, J. J. (2005). The neural circuitry of pre-attentive auditory change-detection: an fMRI study of pitch and duration mismatch negativity generators. Cerebral Cortex, 15(5), 545-551.

Brechmann, A., Baumgart, F., & Scheich, H. (2002). Sound-level-dependent representation of frequency modulations in human auditory cortex: a low-noise fMRI study. Journal of Neurophysiology, 87(1), 423-433.

Poeppel, D., Guillemin, A., Thompson, J., Fritz, J., Bavelier, D., & Braun, A. R. (2004). Auditory lexical decision, categorical perception, and FM direction discrimination differentially engage left and right auditory cortex. Neuropsychologia, 42(2), 183-200.

Paavilainen, P., Mikkonen, M., Kilpeläinen, M., Lehtinen, R., Saarela, M., & Tapola, L. (2003). Evidence for the different additivity of the temporal and frontal generators of mismatch negativity: a human auditory event-related potential study. Neuroscience letters, 349(2), 79-82.

Fiez, J. A., Raichle, M. E., Miezin, F. M., Petersen, S. E., Tallal, P., & Katz, W. F. (1995). PET studies of auditory and phonological processing: Effects of stimulus characteristics and task demands. Journal of cognitive neuroscience, 7(3), 357-375.

Johnsrude, I. S., Zatorre, R. J., Milner, B. A., & Evans, A. C. (1997). Left-hemisphere specialization for the processing of acoustic transients. NeuroReport, 8(7), 1761-1765.

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A. A., Masure, M. C., & Samson, Y. (1998). Lateralization of speech and auditory temporal processing. Journal of cognitive neuroscience, 10(4), 536-540.

Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J. P., Berry, I., Nespoulous, J. L., & Chollet, F. (1999). Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. Neuroimage, 9(1), 135-144.

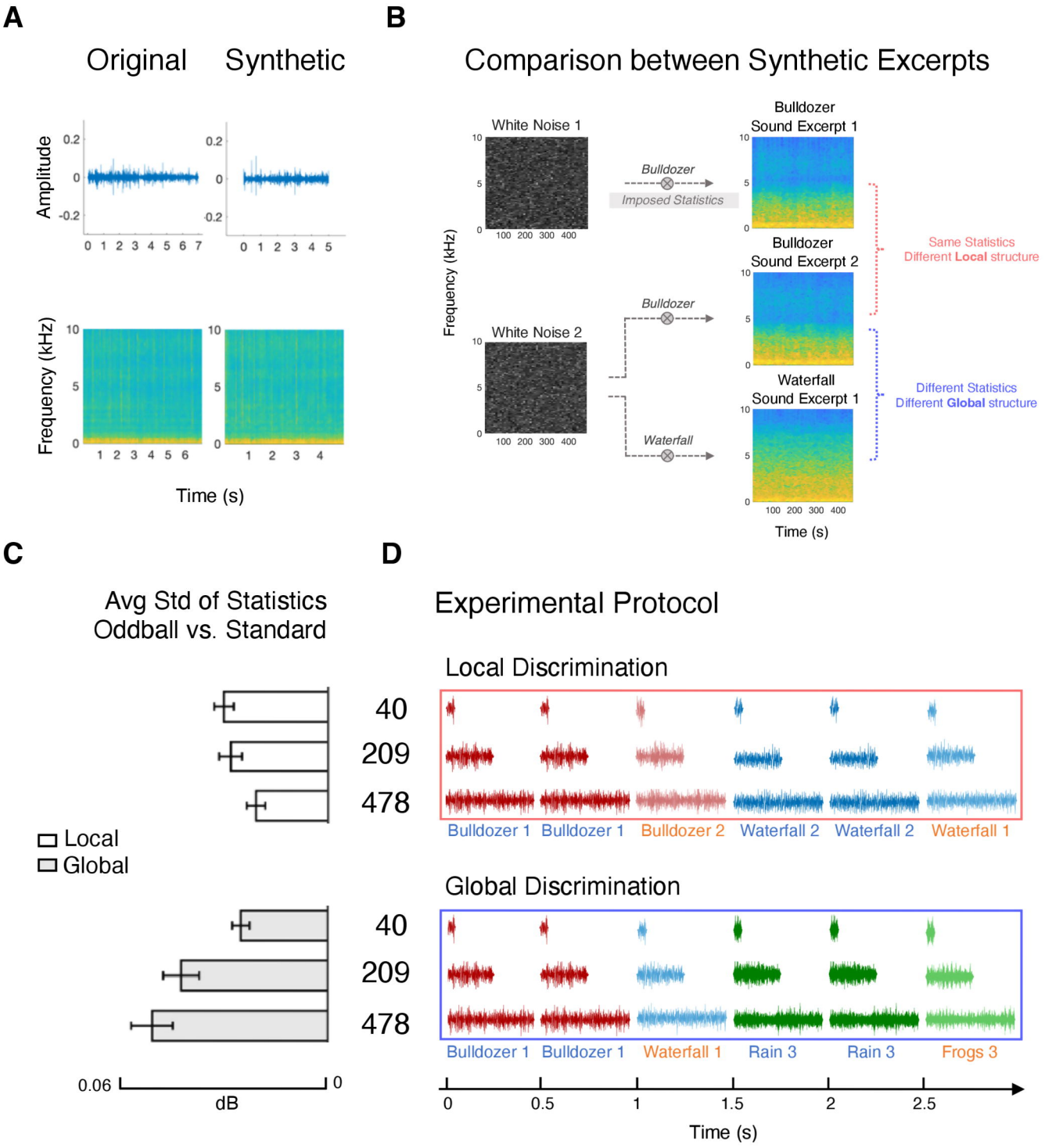
Tallal, P., Miller, S., & Fitch, R. H. (1993). Neurobiological basis of speech: a case for the preeminence of temporal processing. Annals-New York Academy of Sciences, 682, 27-27.

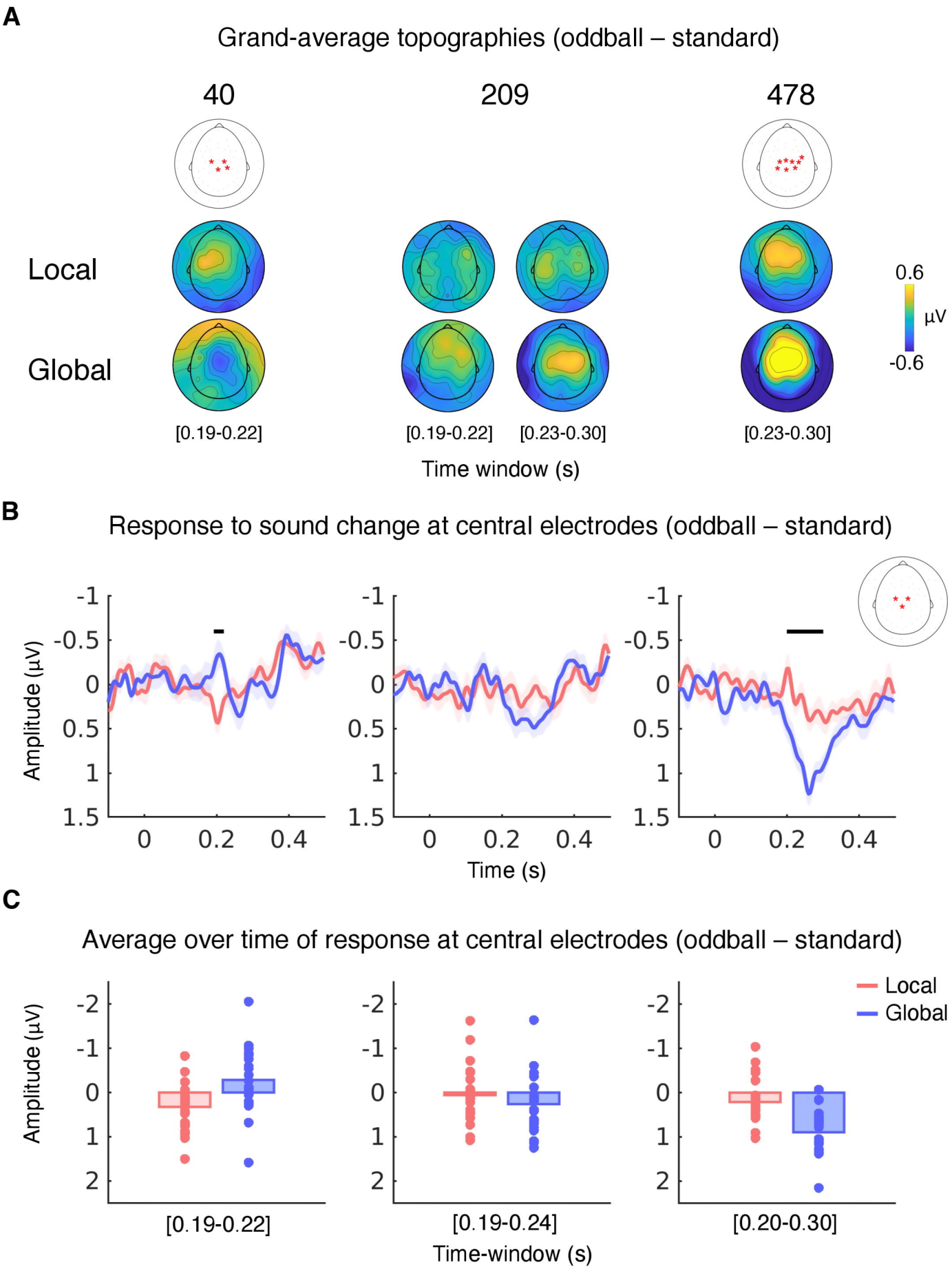
Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., & Binder, J. R. (2010). Specialization along the left superior temporal sulcus for auditory categorization. Cerebral Cortex, 20(12), 2958-2970.

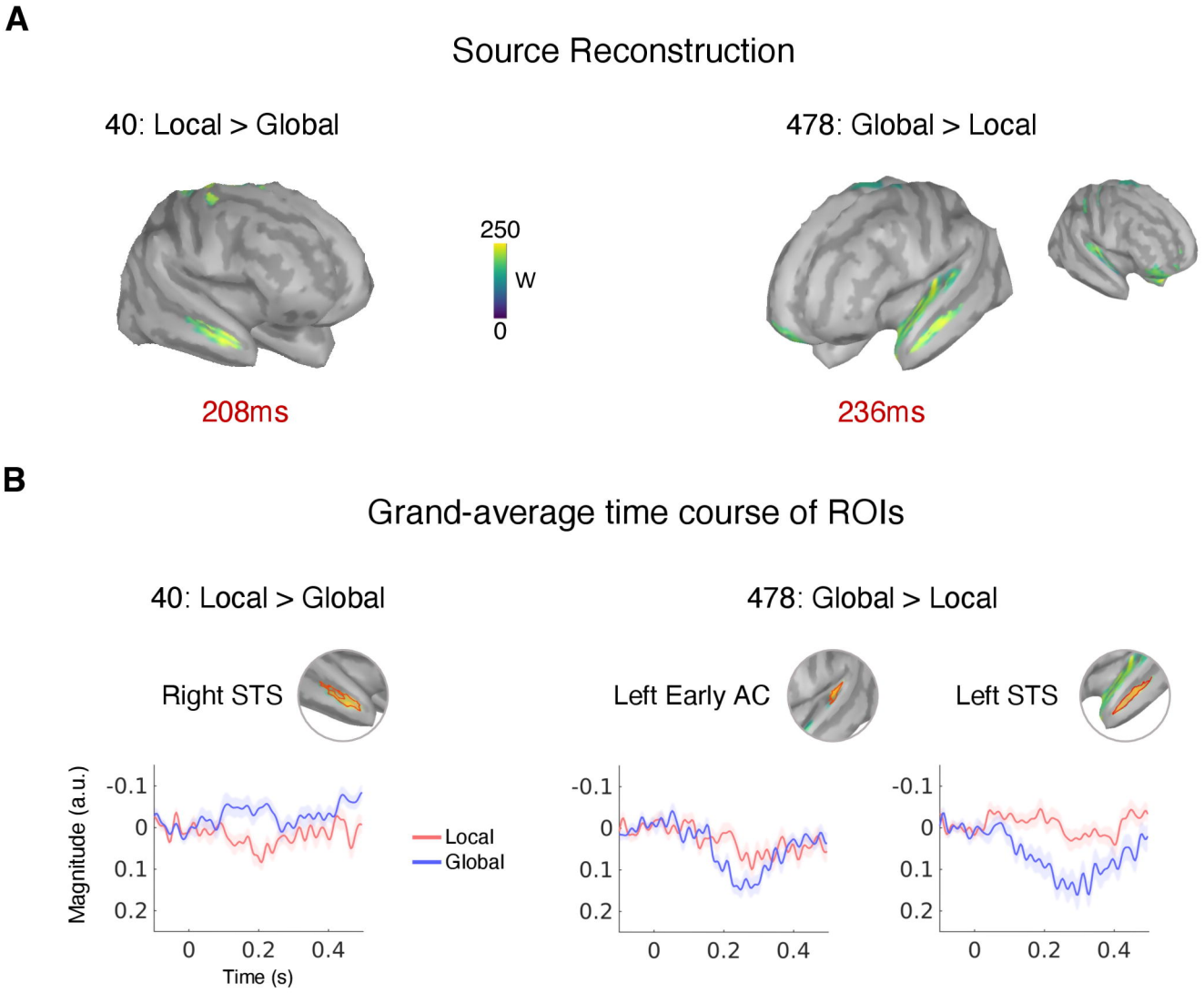
Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., ... & Belliveau, J. W. (2006). Task-modulated “what” and “where” pathways in human auditory cortex. Proceedings of the National Academy of Sciences, 103(39), 14608-14613.

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. Nature Reviews Neuroscience, 14(10), 693-707.

DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. Proceedings of the National Academy of Sciences, 109(8), E505-E514.


A
Original
Synthetic
Amplitude
Frequency (kHz)
Time (s)
B
Comparison between Synthetic Excerpts
White Noise 1
White Noise 2
Bulldozer
Imposed Statistics
Waterfall
Bulldozer Sound Excerpt 1
Bulldozer Sound Excerpt 2
Waterfall Sound Excerpt 1
Same Statistics
Different Local structure
Different Statistics
Different Global structure
Frequency (kHz)
Time (s)
C
Avg Std of Statistics
Oddball vs. Standard
40
209
478
Local
Global
0.06
0
dB
D
Experimental Protocol
Local Discrimination
Bulldozer 1
Bulldozer 1
Bulldozer 2
Waterfall 2
Waterfall 2
Waterfall 1
Global Discrimination
Bulldozer 1
Bulldozer 1
Waterfall 1
Rain 3
Rain 3
Frogs 3
0
0.5
1
1.5
2
2.5
Time (s)


A
Grand-average topographies (oddball – standard)
40
209
478
Local
Global
0.6
µV
-0.6
[0.19-0.22]
[0.19-0.22]
[0.23-0.30]
[0.23-0.30]
Time window (s)
B
Response to sound change at central electrodes (oddball – standard)
Amplitude (µV)
-1
-0.5
0
0.5
1
1.5
0
0.2
0.4
Time (s)
C
Average over time of response at central electrodes (oddball – standard)
Local
Global
Amplitude (µV)
-2
-1
0
1
2
[0.19-0.22]
[0.19-0.24]
[0.20-0.30]
Time-window (s)


A
Source Reconstruction
40: Local > Global
478: Global > Local
250
W
0
208ms
236ms
B
Grand-average time course of ROIs
40: Local > Global
478: Global > Local
Right STS
Left Early AC
Left STS
Magnitude (a.u.)
-0.1
0
0.1
0.2
0
0.2
0.4
Time (s)
Local
Global