HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2018 October 01.

Published in final edited form as:

Neuroimage. 2017 October 01; 159: 195–206. doi:10.1016/j.neuroimage.2017.07.056.

Selective entrainment of brain oscillations drives auditory perceptual organization

Jordi Costa-Faidella[1,2], **Elyse S. Sussman**[3,4], and **Carles Escera**[1,2,5,*]

[1]Brainlab – Cognitive Neuroscience Research Group, Department of Clinical Psychology and Psychobiology, University of Barcelona, 08035-Barcelona, Catalonia-Spain

[2]Institute of Neurosciences, University of Barcelona, 08035-Barcelona, Catalonia-Spain

[3]Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY, 10461, USA

[4]Department of Otorhinolaryngology-HNS, Albert Einstein College of Medicine, Bronx, NY, 10461, USA

[5]Institut de Recerca Sant Joan de Déu, 08950-Esplugues de Llobregat, Catalonia-Spain

Abstract

Perceptual sound organization supports our ability to make sense of the complex acoustic environment, to understand speech and to enjoy music. However, the neuronal mechanisms underlying the subjective experience of perceiving univocal auditory patterns that can be *listened to*, despite *hearing* all sounds in a scene, are poorly understood. We hereby investigated the manner in which competing sound organizations are simultaneously represented by specific brain activity patterns and the way attention and task demands prime the internal model generating the current percept. Using a selective attention task on ambiguous auditory stimulation coupled with EEG recordings, we found that the phase of low-frequency oscillatory activity dynamically tracks multiple sound organizations concurrently. However, whereas the representation of ignored sound patterns is circumscribed to auditory regions, large-scale oscillatory entrainment in auditory, sensory-motor and executive-control network areas reflects the active perceptual organization, thereby giving rise to the subjective experience of a unitary percept.

Keywords

Active sensing; auditory; EEG; oscillations; perception

[*]**Correspondence:** Carles Escera, cescera@ub.edu, Brainlab – Cognitive Neuroscience Research Group, Department of Clinical Psychology and Psychobiology, University of Barcelona, Passeig Vall d'Hebron 171, 08035-Barcelona, Catalonia-Spain. Tel: +34 93 312 50 48; Fax: +34 934 021 584.

The authors declare no conflict of interests.

1. INTRODUCTION

Perception can be thought of as an act of inference (Gregory 1980; Helmholtz 1866). Modern neuroscience views the brain as a predictive machine, continuously generating internal models of the causal dynamics of the world in an attempt to interpret its observations (Bar 2009; Friston 2005). Although relevant to all sensory systems, this assumption especially applies to audition (Baldeweg 2006; Garrido et al. 2008; Winkler et al. 2012). Particularly, it applies to sequential organization, which refers to the sorting of interleaved sounds (Dowling 1973; Bregman 1990; Sussman et al. 1999; Shamma et al. 2011; Winkler et al. 2009). Meaningful auditory objects rely on binding distributed spectrotemporal patterns into coherent streams (Bregman 1990; Nelken and Bar-Yosef 2009; Sussman et al. 1999). Yet, auditory information can sometimes be feasibly explained by more than one internal model. For instance, in a musical piece, a single note from an instrument could belong simultaneously to a melodic line, to a harmonic progression and to a rhythmic pattern featuring several instruments. However, despite *hearing* all sounds, we consciously *perceive* univocal organizations that we can flexibly *listen to*. Our subjective experience therefore conforms to the Gestalt principle of exclusive allocation (Kohler 1947), which states that any sensory element should not be used in more than one description of the natural scene at a time. Whether this principle also applies at the neural level, specifying memory representations of the stimulus input (i.e., whether multiple internal models are held simultaneously or only the current attended one) is still a matter of intense debate (Sussman et al. 2014; Denham et al. 2014; Grossberg et al. 2004).

How the brain flexibly assigns individual events to any of the possible perceptual organizations they could fit into is optimally studied with ambiguous, multistable stimulation, because perception depends on the model currently explaining unchanging sensory input (Sterzer et al. 2009). Behavioral evidence on auditory spontaneous perceptual switches suggests that multiple alternative organizations are held simultaneously and compete to describe the acoustic scene (Denham et al. 2014; Pressnitzer and Hupe 2006; Sterzer et al. 2009; Sussman et al. 2014). Electrophysiological studies in humans have traditionally embedded violations of established regularities within the acoustic streams in order to use change detection auditory evoked potentials, such as the mismatch negativity (MMN) (Näätänen et al. 1978), as an index of sound organization (Sussman et al. 1998, 1999). However, besides yielding conflicting results, with some studies showing simultaneous encoding of alternative organizations (Pannese et al. 2015; Sussman et al. 2014) while others suggesting that only the currently perceived organization is represented (Sussman et al. 2002; Sussman 2013; Winkler et al. 2006), evidence of this nature is intrinsically indirect and does not inform about the neural mechanisms underlying the representation of sound organization.

Several studies have shown that any existing regularity in the auditory scene is reflected in oscillatory activity tuned to its temporal pattern (Henry et al. 2014; John et al. 2001, 2002; Luo et al. 2006, 2007; Pannese et al. 2015). This is an interesting observation because synchronized oscillatory activity has been proposed as an effective means for neuronal communication (Fries 2005). Moreover, since the high-excitability phase of ongoing low-frequency oscillations can be selectively entrained to events occurring in an attended stream

(Schroeder and Lakatos 2009), we speculate that neuronal entrainment could underlie our perceptual ability to flexibly reorganize sequential sounds.

We hereby designed a novel ambiguous sound sequence that allowed the study of active perceptual reorganization while controlling for sensory input. Given the quasi-rhythmic nature of most behaviorally relevant acoustic information (Patel 2008), rhythmic attention (Jones and Boltz 1989; Large and Jones 1999), and its neurophysiological counterpart oscillatory entrainment (Herrmann and Henry 2014; Schroeder and Lakatos 2009) would likely play a key role (Pannese et al. 2015). Nozaradan et al. (2011) demonstrated that oscillatory entrainment underlies meter imagery, the voluntary organization of musical beats. However, the imagined meter was imposed on a sound sequence with acoustic energy only at the main beat rate. This leaves open the question of whether oscillatory entrainment actually helps to disambiguate a rhythmic structure that has multiple potential meters. With energy at more than one possible meter, task demands may act to enhance the attended meter while suppressing the unattended one, rather than driving the overall meter of the sequence.

To target the dynamics of large-scale neuronal slow oscillatory activity, we combined spectral analyses with source localization of EEG data, seeking to explore the distinction between the neurophysiological nature of simultaneously encoded representations of the auditory scene, and the selected internal model underlying the perceived auditory object.

2. MATERIALS AND METHODS

2.1 Participants

Fourteen healthy volunteers (mean age: 28.9 years; age range: 24–38 years; 8 males; 2 left-handed) with no self-reported history of neurological, psychiatric, or hearing impairment and with normal or corrected-to-normal visual acuity participated in the experiment. All participants passed a hearing screening including pure tones of 500, 1000, 2000, and 4000 Hz at 20 dB HL prior to the recording session. One participant reported being an active amateur musician without formal training. Data from two participants were excluded due to poor task performance. All volunteers gave written informed consent in accordance with the guidelines if the Internal Review Board of the Albert Einstein College of Medicine (New York City, NY, USA) before their participation and after the procedures were explained to them. The study conformed to the Code of Ethics of the World Medical Association (Declaration of Helsinki). Data is fully available upon request.

2.2 Auditory stimuli

Sixty-four different pure sinusoidal tones (35000 Hz sampling rate) were generated with Matlab (R2008a; Mathworks) and delivered binaurally via insert earphones by the Stim interface system (NeuroScan Labs, Sterling, VA). The tones featured 16 different frequencies, ranging from 440 Hz (A4) to 1046.5 Hz (C6) in steps of one semitone, two different values of duration (40 and 120 ms) and two different intensities (70 and 85 dB SPL), with rise and fall times of 5 ms (Hanning window).

2.3 Sound sequence

Auditory stimuli were arranged in separate sequences (see Fig. 1A), each containing 12 repetitions of a four-tone melodic ascending-descending pitch pattern including three different tones (i.e., f1-f2-f3-f2). The frequency separation between adjacent tones was set to one semitone to facilitate sound integration into melodic patterns (Bregman 1990). Stimulus Onset Asynchrony (SOA) and inter-pattern interval were set to 200 ms. In order to minimize effects of neuronal adaptation across sequences, each sequence randomly presented a different set of three tone frequencies from the pool of 16, with the constraint that any frequency featured in a sequence could not appear in the subsequent one. Because all sequences would have 48 stimuli, in order to avoid the participants' expectancy of sequence length (which could influence task performance - see *Delayed response task*), we varied the number of tones in a sequence by randomly shortening or lengthening it by half a pattern (±2 tones). Tone duration alternated between short (40 ms) and long (120 ms) every four tones (800 ms), coinciding with the onset of the melodic pattern. A sequence started with a melodic pattern of short or long tones at random, with a 50% probability. Tone intensity varied in a three-tone pattern (600 ms) consisting of 1 loud tone (85 dB SPL) followed by 2 soft tones (70 dB SPL). The intensity of the first tone in the sequence was always a loud one. This arrangement of tone features resulted in a perceptually ambiguous sound sequence with a rhythm of 5 Hz (tone presentation), a rhythm of 1.25 Hz (corresponding to the duration/melodic pattern) and a rhythm of 1.67 Hz (corresponding to the intensity pattern), as illustrated by the sequence spectrum at Fig. 1D (*see also sound1.mp3*).

2.4 Delayed response task

In order to bias and stabilize the perceptual organization of the sequence, and to minimize muscle contamination during the EEG recording, participants were asked to perform a delayed response task associated with each of the two possible percepts. Therefore, hit rate measures, but not reaction times, could be analyzed. To ensure sound organization according to the duration/melodic pattern, participants were asked to attend to tone duration and detect whether a group of five consecutive tones of same length, appearing randomly between stimulus 36^{th} and 48^{th} (towards the end of the sequence) featured short or long duration. To ensure sound organization according to the intensity pattern, participants were asked to silently count the louder tones and report how many appeared in the sequence (either 14 or 15; from these to the end of the sequence all tones were soft in intensity). Importantly, the sound sequence was the same during the first 7.2 s regardless of the task. Two response buttons in a joystick were enabled at the end of each sequence and participants could respond until the next trial started. The order of the response buttons (left/right) was fixed during the experimental conditions and counterbalanced across participants. Participants used the left and the right thumbs to press the left and right buttons, respectively.

2.5 Procedure

Prior to recording, volunteers participated in a practice session in which they performed the duration pattern task in a sequence without loudness changes and the intensity pattern task in a sequence without duration changes. Once the tasks were clear, they practiced on an experimental (ambiguous) sequence until performing accurately (>75% hit rate). Practice

sequences were not used in the main experiment. At the recording session, participants sat in a comfortable chair in a sound-attenuated, electrically shielded room. A computer screen was placed in front of the chair at a distance of 1 meter at the height of the participant's eyes. The experiment started with 12 consecutive blocks of one condition (either duration or intensity pattern task), separated by resting pauses, followed by 12 blocks of the remaining condition. The order of the conditions was counterbalanced across subjects. Each block featured 12 trials, consisting in the presentation of a visual cue of 1s length, serving as a reminder of the required task and response button set, followed by a pause of 3, 3.5 or 4s length (rectangular distribution), followed by the sound sequence (9.2–10s, in 200ms steps; rectangular distribution), and finally followed by a pause of 1.5, 2 or 2.5s length (rectangular distribution). During the whole trial a white fixation cross was presented at the center of the screen with a black background. In the duration pattern task condition, the pre-sequence visual cue consisted in a green rectangle and a red square, both containing a white "5" number at the center, appearing at the left or right of the fixation cross, depending on response button set (e.g., the green rectangle appeared at the left of the cross if the left button was to be pressed when 5 long tones in a row were detected; the red square represented the response button for the 5 short tones). During the intensity pattern task condition, the pre-sequence visual cue consisted in a green square with a white "14" number at the center and a red square with a white "15" number at the center. Again, depending on the response button set the squares appeared either at the left or the right of the fixation cross (e.g. the green square appeared at the left of the cross if the left button was to be pressed when 14 loud tones were counted). Subjects were required to fixate the gaze at the cross during the experimental block and explicitly asked not to perform any movement (e.g., tapping with fingers or feet; moving the tongue while silently counting), being monitored with a closed-circuit television camera to ascertain so.

2.6 EEG acquisition and preprocessing

EEG was continuously recorded with frequency bandpass of 0.05–100 Hz and digitized at a sampling rate of 500 Hz by a SynAmps amplifier (NeuroScan Inc., Herndon, VA). Pure tin electrodes were used for the EEG acquisition, 29 of which were mounted in a nylon cap (Electro-Cap International, Eaton, OH) at standard locations according to the international 10–20 system. Additionally, two electrodes were positioned over the left and the right mastoids (M1 and M2, respectively). Vertical electro-oculogram was measured from an electrode placed below the right eye (VEOG) and FP2. The ground electrode was placed at P09 and the common reference electrode was placed at the tip of the nose. All impedances were kept below 5 kΩ during the whole recording session. Data preprocessing was performed offline using EEGlab v.7 software (Delorme and Makeig 2004) running under Matlab v7.6 (Mathworks). Continuous EEG data was downsampled to 250 Hz using EEGlab's function *pop_resample.m*, which applies an anti-aliasing filter, and high-pass filtered from 0.5 Hz (Kaiser window; w = 5.65; TBW = 0.5 Hz). Periods contaminated by non-stereotyped muscle artifacts were rejected by visual inspection. Independent component analysis decomposition was applied using the Infomax algorithm, removing blink-related independent components on the basis of their scalp topography and time course (Jung et al., 2000). After removing VEOG from the channel set, EOG artifact-corrected data were re-referenced to the average of all channels (n = 31). Continuous EEG data was cut in epochs

from −4 to 10 s time-locked to the onset of the first tone in a sequence and baseline corrected from −1 to 0 s. To ensure that brain activity corresponded to a successful perceptual reorganization of the sound sequence, only epochs containing a hit in the task were selected for further analyses. A participant was excluded from the sample if less than 8 blocks in a condition featured less than 75% hit rate (9 hits out of 12 trials per block). According to this criterion 2 participants were excluded. From the remaining participants (n = 12), epochs with an amplitude range exceeding ±75 μV were rejected. After rejection, a mean of 130.75 (*SD* = 14.73) and 135.25 trials (*SD* = 11.13) corresponding to the duration and the intensity pattern conditions respectively were included in subsequent analyses (no significant difference was found between the number of trials used in each condition; $t_{11} = -1.57$; $p = 0.14$).

2.7 Spectral analysis

Following the procedures described in Elhilali et al. (2009), single trials obtained after preprocessing were cut from 0 to 7.2 s, thus avoiding the silent period and the inclusion of targets, and concatenated forming a single response separately per condition, channel and participant. A Fast Fourier Transform (FFT; Hanning windowed; using Fieldtrip software running under Matlab v7.6 [Mathworks], www.ru.nl/neuroimaging/fieldtrip, (Oostenveld et al. 2011) was applied on this single response, giving a single frequency spectrum from 0.5 to 100 Hz with a minimum frequency resolution of 0.0015 Hz (equivalent to a minimum response length (*T*) of 669.6 s, corresponding to one condition of one participant including 93 trials). Neural responses to the perceptually reorganized sequences were characterized by the magnitude of the frequency components at 1.25 Hz (duration pattern), 1.67 Hz (intensity pattern) and 5 Hz (tone presentation rate). The electric field strength was calculated as the product of the value of FFT times the sampling interval (1/250) and has units of μV/Hz. Power spectral density was calculated as the product of the inverse duration ($1/T$) times the modulus squared of the electric field strength, and has units of $\mu V^2/Hz$. The rest of the analysis was based on normalized neural responses, defined to be the squared magnitude of the frequency components at 1.25, 1.67 and 5 Hz, divided by the average squared magnitude of these frequency components ±1 Hz (but excluding the bins corresponding exactly to 1.25, 1.67 and 5Hz), averaged over the 5 channels with the strongest normalized neural responses for each participant, frequency component and condition. This method allows for inter-participant configuration variability across conditions and frequency components (Elhilali et al. 2009; Xiang et al. 2010). To characterize the effects of perceptual sound sequence reorganization due to selective attention, the relative change (dB) in normalized neural responses (nnr) between the two conditions (duration pattern task, DPT; intensity pattern task, IPT) was computed as 10*log(nnrDPT/nnrIPT) for each frequency component and participant. Furthermore, to evaluate the effect of selective attention at across frequencies, the same analysis was done at adjacent frequency bins around the rates of interest (±0.07 Hz). To assess the significance of effects, two-tailed Wilcoxon signed rank tests were used.

2.8 Source reconstruction and analysis

We used the standardized Low Resolution Electromagnetic Tomography software package (sLORETA; Fuchs et al. 2002; Jurcak et al. 2007; Pascual-Marqui 2002) to reconstruct the neural generators of entrained oscillatory activity. sLORETA solutions from simulated data

with noise are highly comparable between a 25 vs. 101 channel recording (Pascual-Marqui, 2002), and surface Laplacian estimates (CSDs) and PCA data decompositions of low- and high-density human EEG recordings (31 vs. 129 channels) are remarkably similar (Kayser & Tenke, 2006). Thus, based on these simulation and CSD studies, and given that group findings are the primary objective of our study, we feel sufficiently confident to report our source localization results with 31 channels.

sLORETA computes three-dimensional linear solutions for the EEG inverse problem with a three-shell spherical head model adapted to the probabilistic Talairach human brain atlas (Lancaster et al. 2000) digitized at the Brain Imaging Center of the Montreal Neurological Institute (MNI152 template; Mazziotta et al. 2001). The sLORETA solution space is restricted to cortical and hippocampal gray matter, comprising a total of 6239 voxels at 5 mm spatial resolution. The detailed description of the method can be found in Pascual-Marqui (2002). We computed the cross-spectral matrices of the EEG concatenated epochs per subject and condition (duration and intensity pattern tasks) for three frequency bands: 1.25 Hz (duration pattern), 1.67 Hz (intensity pattern) and 5 Hz (tone presentation rate). We included two additional baseline conditions, taking pre-sequence epochs from −1 to 0 s for each trial in each condition separately. Because the concatenated epochs featured different lengths that depended on the length of the single epoch (activity, 7.2 s; baseline, 1 s) and the number of trials per condition and subject, and because sLORETA poses computational constraints, we adopted a sliding window approach to compute the cross-spectra, akin to Welch's averaging method (Welch 1967). First, data were downsampled to 125 Hz using EEGlab's function *pop_resample.m*, which applies an anti-aliasing filter. Second, the cross-spectra were calculated in windows of 57.6 s, corresponding to 8 concatenated activity trials (and 57.6 baseline trials), using the sLORETA software. These windows were obtained by sliding along the concatenated data in jumps of a single trial length (i.e., activity window overlap: 87.5%; baseline window overlap: 98.2%). Third, we averaged the cross-spectra obtained from all windows. This approach allowed the inclusion of all trials per subject, matching signal length across conditions and thus frequency resolution (0.017 Hz). The averaged cross-spectral matrices for each subject in each condition were given as the input for sLORETA source analysis. The sLORETA yielded the spectral density of the current density at each voxel. sLORETA statistical contrast maps for each frequency band were assessed by voxel-by-voxel t-tests of the sLORETA solutions (subject-wise normalized; including all 6239 voxels), with the computed current density power log-transformed and a smoothing factor of 0.2, in a nonparametric permutation test with correction for multiple testing (SnPM; Nichols and Holmes 2001). The obtained t-values exceeding the critical probability threshold values, based on a test with 20000 permutations, were plotted onto a MRI template (MNI152 T2). For each frequency of interest, we compared both conditions against each other and against their baseline activity. According to the low spatial resolution of the sLORETA method, and only for reporting purposes, we merged the significantly activated voxels after the whole-brain analysis in a series of broad brain regions roughly based on their functional properties. These regions were defined according to Brodmann areas in the Talairach atlas for each hemisphere separately, as follows: Auditory Cortex (AC), including BAs 13/21/22/38/41/42/43; dorsal Premotor Cortex and Supplementary Motor Area (dPMC/SMA), including BAs 6/8 with voxels located above the inferior

junction of the superior frontal sulcus with the superior precentral sulcus, approximately at the $z = 50$ plane (Rizzolatti and Craighero 2004); ventral Premotor Cortex and BA44 (vPMC/BA44), including BAs 6/8 with voxels located below the $z = 50$ plane (Rizzolatti and Craighero 2004) and BAs 44/45/47; Dorsolateral Prefrontal Cortex (DLPFC), including BAs 9/46; Sensorimotor Cortex (SMC), including BAs 1/2/3/4/5; Anterior Cingulate Cortex (ACC), including BAs 24/25/32/33; Posterior Cingulate Cortex (PCC), including BAs 23/29/30/31; Inferior Parietal Lobule (IPL), including BA 40; Occipitotemporal Area (OT), including BAs 19/20/37; Parahippocampal Gyrus, including BAs 27/28/34/35/36; Precuneus/Cuneus, including BAs 7/17/18; Angular Gyrus, including BA 39 (excluding voxels located at Superior Temporal Gyrus); Orbitofrontal Area (OF), including BAs 11/12; and Anterior Prefrontal Cortex (APFC), including BA 10.

In the results section, we provide the activated regions, the maximum t value within a region, and the cluster size (number of significantly activated voxels within a defined region; reported minimum cluster size of 5 voxels).

2.9 Time-Frequency Analysis

In order to ascertain whether entrained oscillations reflected phase reorganization only or concomitant increments in total power (phase-locked and non-phase-locked), and to observe their time course along the sequence, we performed time-frequency (TF) analyses on the single trials obtained after EEG preprocessing (−4 to 10 s long), locked to the onset of each sequence for each condition separately. The complex Fourier spectrum was obtained by convolving single trials with complex Morlet wavelets with a linearly increasing number of wavelet cycles from 5 to 20 as frequencies ranged from 1 to 100 Hz, in 150 exponentially spaced frequency bins. Estimates of phase-locking factor (PLF), which is a measure of phase alignment independent of power with values ranging between 0 (no phase alignment) and 1 (complete phase alignment), and total power, calculated by averaging the squared absolute values of the convolutions over trials, were computed in a time window ranging from −1s to 7s relative to sequence onset, thus avoiding wavelet edge artifacts and excluding any target related activity, according to the procedures described by Tallon-Baudry et al., (1996). Statistical comparisons were performed using non-parametric cluster-based permutation tests, following the procedure described by Maris and Oostenveld (2007). This method effectively controls for multiple comparisons and allows the identification of clusters with significant group differences in 3D (time, frequency, and electrode). Neighboring electrodes were defined using a Delaunay triangulation over a 2D projection of the electrode montage, which connects nearby electrodes independently of their physical distance. We set a minimum of 2 nearby electrodes per cluster. In order to increase the sensitivity of the tests, we performed 2D cluster-based analyses (time, electrode) separately for the frequency bands of interest (1.25, 1.67 and 5 Hz) in a time-range between 0 and 7s. As no a priori hypotheses were drawn for higher frequency bands, we performed a 3D cluster-based analysis including three frequency bands: alpha (8–13 Hz), beta (13–30 Hz) and gamma (30–100 Hz). We compared statistically the activity drawn from one condition against the other (duration pattern task vs. intensity pattern task) and, whenever statistical differences were found, we compared each condition against its respective baseline (time-averaged; −1 to −0.5s) to ascertain that effects were not due to noise. For each comparison, the resulting power/PLF

values at each TF point and electrode underwent a two-tailed dependent t-test assessed with the non-parametric Monte Carlo method. The Monte Carlo significance probability (p-value) was determined by calculating the proportion of 2D/3D samples from 5000 random partitions of the data that resulted in a larger test statistic than those on the observed test statistic. Then, clusters were created by grouping adjacent 2D/3D points exceeding a significance level set to 0.05. A cluster-level statistic was calculated by taking the sum of the t-statistics within every cluster. The significance probability of the clusters was assessed with the described non-parametric Monte Carlo method. Values of $p < 0.025$ were considered significant. All analyses (time-frequency transformation and statistics) were performed using Fieldtrip software running under Matlab v7.6 (Mathworks), www.ru.nl/neuroimaging/fieldtrip, (Oostenveld et al. 2011). As no significant differences were found in the alpha, beta or gamma range between conditions, we only report results on the frequency bands of interest for entrained oscillatory activity (1.25, 1.67 and 5 Hz).

3. RESULTS

To investigate the neuronal mechanisms underlying the representation and active selection of competing parallel models of sound input, we asked participants to actively listen to an ambiguous sound sequence that could be perceived in two mutually exclusive ways (Fig. 1A; *sound1.mp3*). The sequence consisted of a melodic ascending-descending pitch pattern with three different tone frequencies separated by one semitone each, at a tone presentation rate of 5 Hz (200 ms SOA) (Fig. 1A). Tone duration changed every four tones (120 ms and 40 ms), reinforcing the melodic pattern, and one tone every three featured a louder intensity (82dB vs. 70dB SPL), breaking the melodic pattern. The rates of the duration and the intensity patterns were non-multiples (1.25 and 1.67 Hz respectively), and thus the perceptual interpretation of the sound sequence was ambiguous. In other words, perception of the sound organization depended on the listener's selective attention of the individual sounds within the sequence as one of two mutually exclusive rhythmic organizations. Importantly, the three main rhythms (1.25, 1.67 and 5 Hz) were *physically* present in the acoustic signal, as revealed by a frequency decomposition of the sound sequence using a Fast Fourier Transform (FFT; Fig. 1D). To ensure perceptual stability, we asked participants to perform a delayed response task on the duration and the intensity patterns (Fig. 2A; see *Materials and Methods* for task description). Targets appeared towards the end of the sound sequence to separate button-press muscle activity from perception-related neuronal activity. Performance (hit rate) on the intensity pattern task (IPT) was higher (94.73%, SD = 5.32%) than for the duration pattern task (DPT) (89.99%, SD = 7.82%). However, this difference did not quite reach significance ($t_{11} = -2.19$; $p = 0.051$).

3.1 Brain oscillations entrain selectively to the rate of the task-specific sound organization

We recorded scalp EEG from our participants to reveal changes in neuronal activity due to selective attention and perceptual reorganization of the sound sequence. These changes were visible in individual subjects and are illustrated for a representative participant (Fig. 2A–B). The auditory evoked response (i.e., averaged single trials) band-pass filtered from 4–6 Hz revealed a sustained oscillation to tone presentation rate regardless of the performed task. However, the auditory evoked response band-pass filtered from 1–2 Hz mirrored the

perceived sequence organization by entraining brain oscillations to either the duration or the intensity pattern rates. Such differences are clearly illustrated in the frequency spectrum of the EEG signal, calculated with a FFT on a long data epoch of concatenated single trials, as a fine-tuned selective enhancement of power spectral density in the frequency component corresponding to the attended rate. Pooled data across subjects also shows the same pattern of power spectral density (Fig. 2C) and the sharpness of the oscillatory activity enhancement at the rates of interest, as frequency components only 0.07 Hz away show close to 0 neural responses (Fig. 3A).

Group analyses on relative neural response enhancement between the DPT vs. IPT conditions (Fig. 2B) revealed a significant increase at 1.25 Hz (median = 29.97 dB; interquartile range = 19.45 dB; Wilcoxon signed rank test = 78; p = 0.00048); a significant decrease at 1.67 Hz (median = −24.5 dB; interquartile range = 23.71 dB; Wilcoxon signed rank test = 7; p = 0.0093) but no differences at 5 Hz (median = 5.01 dB; interquartile range = 9.64 dB; Wilcoxon signed rank test = 63; p = 0.064), nor at any of the computed adjacent (Δf = 0.07Hz) frequency bins. To further demonstrate the dissociation of neural response at sound pattern rates depending on attention and task demands, we compared the relative neural response enhancement (DPT vs. IPT) at 1.25Hz with that at 1.67 Hz (Wilcoxon signed rank test = 77; p = 0.0009). This strong interaction between task and frequency can be reliably observed in all the participants in the sample (Fig. 3C).

These results provide strong evidence in favor of brain oscillatory entrainment as a mechanism to resolve internal model competition, as shown by the highly selective neuronal synchronization to the rhythm of the perceived sequence organization used to perform the task.

3.2 Selective entrainment of brain oscillations to the task-relevant sequence organization recruits an audio-motor network

Several neuroimaging studies have shown the activation of a sensory-motor network recruiting the auditory cortex (AC) as well as the supplementary-motor area (SMA), the dorsal and ventral premotor cortex (dPMC, vPMC), the cerebellum and the basal ganglia, that is engaged when listening to rhythms, even when they involve sounds that are not meaningful for the motor system (Chen et al. 2008a, 2008b). However, due to the coarse temporal resolution of fMRI, these studies could not access the temporal fine-tuning of synchronized oscillatory activity to attended rhythms. To reveal the network of brain regions exhibiting oscillatory entrainment, we computed the cross-spectral matrices of the EEG concatenated single trials per participant and condition (DPT and IPT) for the three frequency bands of interest (1.25, 1.67 and 5 Hz) and reconstructed their sources with sLORETA (see *Materials and Methods*). The following results correspond to significantly activated regions after a whole-brain analysis using a non-parametric statistical test with correction for multiple comparisons (limited to cortical and hippocampal gray-matter).

Fig. 4A–B illustrate the cortical regions differentially activated when performing the DPT vs. IPT for both the duration and the intensity pattern rhythms respectively (1.25 and 1.67 Hz; *see also* Fig.S1A–B). At 1.25 Hz (Fig. 4A), oscillatory entrainment was enhanced when performing the DPT vs. IPT ($t > 3.25$; $p < 0.05$), engaging a network of brain regions

comprising strong AC, dPMC/SMA, vPMC and Anterior Cingulate Cortex (ACC) activations. Other weaker but significant activations included the Dorsolateral Prefrontal Cortex (DLPFC) and the sensory-motor cortex (SMC). Conversely, at 1.67 Hz (Fig. 4B), oscillatory entrainment was enhanced when performing the IPT vs. DPT ($t > 3.4$; $p < 0.05$) in a network including AC and vPMC activations. Other weaker significant activations comprised DLPFC and SMC. The complete list of activations including cluster size and peak voxel value is summarized in Table 1. To confirm that differences between conditions were not due to noise in the data, we compared each condition against silence (*see* Fig.S2 A and C). We found widespread strong activations, as it would be expected considering the high signal-to-noise ratio of our data (*see* Fig.3A).

At the 5 Hz tone presentation rate, there were no significant differences between conditions. However, when comparing both the DPT and the IPT against a silent baseline, previous to the onset of the sound sequence, we found the strongest activated region located in the AC ($t > 9.5$; $p < 0.01$)(Fig. 4C; Table 1). These results, comparing each condition separately against its baseline, can be found in Fig.S2B–D.

Another interesting finding is illustrated in Fig. 4D. When comparing the DPT vs. silence at the intensity pattern rhythm (1.67 Hz), we found a significant activation circumscribed to the AC ($t > 4.1$; $p < 0.05$)(Table 1). Conversely, comparing the IPT vs. silence at the duration pattern rhythm (1.25 Hz) did not reveal any significant activation.

Taken together, these results suggest that whereas non-attended sound organizations are simultaneously represented by the oscillatory dynamics of the auditory system, the currently selected internal model is represented in a pulsating audio-motor network that additionally recruits other areas related to executive control. The presence of significantly activated auditory regions for the high-order unattended intensity pattern, in contrast to the absence of significant activations for the unattended duration pattern, is consistent with the notion that melody and duration are comparatively less robust cues than loudness for subjective sound grouping (Repp 2007).

3.3 Entrainment as phase reorganization of brain oscillations to the task-relevant sequence organization

To observe the time course of activity, we computed estimates of phase-locking factor (PLF; ranging from 0 to 1) and total power for the frequency bands of interest (1.25, 1.67, and 5Hz) on single trials of the whole sound sequences (−1 to 7s); *see Materials and Methods*). Thus, we assessed whether entrained oscillatory activity reflected phase reorganization, concomitant increments in total power (phase-locked and non-phase-locked), or both. Non-parametric cluster-based permutation tests were used to correct for multiple comparisons in time and electrode dimensions.

Fig. 5A illustrates the evolution of these clusters in time and space. The comparison between PLF obtained when performing the DPT vs. IPT revealed a positive cluster at 1.25Hz ($T = 219080$; $p < 0.0001$; from 0.618 to 7s) and a negative cluster at 1.67Hz ($T = -119030$; $p < 0.0005$; from 1.558 to 7s), but no significant differences at 5Hz. To confirm that differences between conditions were not due to noise, and to affirm the direction of the effects, we

separately compared the PLF elicited when performing the DPT and the IPT against the silent baseline. The analysis of the DPT vs. silence yielded a positive cluster at 1.25Hz (T = 248060; $p < 0.0001$; from 0 to 7s), a positive cluster at 1.67Hz (T = 129660; $p < 0.0005$; from 0 to 5.786s) and a positive cluster at 5Hz (T = 459880; $p < 0.0005$; from 0 to 7s). The evolution of these clusters in time and space is illustrated at Fig. 5B. The analysis of the IPT vs. silence yielded a positive cluster at 1.25Hz (T = 64954; $p < 0.001$; from 0 to 1.576s), a positive cluster at 1.67Hz (T = 253380; $p = 0.001$; from 0 to 7s) and a positive cluster at 5Hz (T = 423750; $p < 0.0005$; from 0 to 7s). The evolution of these clusters in time and space is illustrated at Fig. 5C.

The comparison between total power obtained when performing the DPT vs. IPT only revealed a negative cluster at 1.67Hz (T = 65067; $p < 0.01$; from 1.082 to 4.678s), which was rather due to a decrease in power at that frequency band during the DPT (1 negative cluster against the silent baseline: T = 58491; $p < 0.05$; from 1.934 to 7s) than to any increase of power during the IPT (no significant differences against the silent baseline).

Together, these results indicate that active organization of a sound sequence entails phase reorganization to the rhythm of the perceived pattern without concomitant power increments. Actually, the decrease of power at the frequency of unattended patterns may be indexing a tuning of delta band resources to the attended organization rhythm. Furthermore, developing such selective phase reorganization requires more than one second of time. This is illustrated by the fact that while the initial sounds in the sequence are processed equally regardless of the task performed, PLF is increased to the task-relevant and decreased to the task-irrelevant organization rates within the sound sequence. This time course may be related to the buildup of stream segregation (Bregman 1978). Besides, while ignoring the loud tone pattern does not fully prevent its representation until after several seconds, the representation of the unattended duration pattern is much briefer (~1.6s), in line with our source localization results (Fig. 4D). Moreover, cluster topographies at 5Hz (and 1.67 Hz when ignoring that pattern rate) are consistent with auditory cortex generators (Näätänen and Picton 1987) reflecting local activity, whereas topographies at the rhythms of the perceived patterns are widely distributed, reflecting the involvement of a large-scale pulsating network.

4. DISCUSSION

The goal of the current study was to investigate neuronal mechanisms underlying the active selection of competing concurrent internal models of the auditory scene, and the neural representations reflecting the sequential organization of the sounds. Our results revealed that multiple competing sound organizations are concurrently represented as entrained oscillatory activity to their intrinsic rhythms. However, we found a clear dissociation between the ignored (non-task-specific) and active (task-specific) organizations. Responses to task-irrelevant organizations were restricted to auditory cortices, whereas responses to task-relevant organizations recruited additional regions from the motor system and the executive control network. These results are consistent with our previous studies showing differential processing of sounds in passive and active listening conditions (Sussman and Steinschneider 2009, 2011). We extend these findings and provide novel evidence showing

temporal dynamics of large-scale networks in complex sound scenes induced by task-specific sound grouping.

4.1 "Winner-takes-all" or concurrent encoding of multiple internal models of the auditory scene

In our initial spectral analysis, we found exquisitely tuned oscillatory activity to the rates of the attended organizations (1.25 Hz and 1.67 Hz), along with increased synchronization at the rate of stimulus presentation (5 Hz). While this analysis disclosed that higher-order non-task relevant sound organizations were not represented in the neural activity, further source location and PLF analyses showed concurrently synchronized activity originating from auditory cortical areas. However, the task-irrelevant responses were weaker and fading over the course of the sequence. The overall pattern of results fits with previous research revealing that the auditory system is able to hold parallel acoustic regularities in multiple time-scales (Costa-Faidella et al. 2011; Horvath et al. 2001; Pannese et al. 2015; Sussman et al. 2014; Ulanovsky et al. 2004; Winkler et al. 2009). Our results support the idea that auditory streaming builds-up over time (Bregman 1978; Sussman et al. 2007; Sussman-Fort and Sussman 2014; cf., Deike et al. 2012). Further, as revealed by a decrease in PLF and oscillatory power to the rate of task-irrelevant organizations, our results suggest that attention inhibits the representation of the unattended organization rather than enhancing that of the attended one, consistent with behavioral results on spontaneous switching (Pressnitzer and Hupe 2006).

How does evidence of simultaneous representation of multiple sound organizations correspond with our subjective experience of a unitary percept in awareness? The goal of the auditory system is to provide veridical representations of the sound sources in the environment, sorting out the individual auditory objects from the intricate sound mixtures, a feat known as *auditory scene analysis* (Bregman 1990). Holding multiple representations of the environment simultaneously available would therefore enable the flexible identification of sound sources (Sussman et al. 2014). According to our data, proto-objects that may compete to describe the acoustic scene are encoded in the activity of the auditory system (Bregman 1990; Winkler et al. 2012). Success in the competition would be attained by engaging a global network that brings the winning representation to the foreground of awareness (Baars 2005; Dehaene et al. 1998; Dehaene & Naccache 2001). This idea could provide a plausible explanation to previous studies showing that while deviant sounds embedded in task-relevant and non-task-relevant organizations elicit a change detection signal, only those marked as targets in the attended organization elicit a P3b auditory evoked potential (Sussman et al. 2014). Although change detection is considered a pervasive property of the auditory system (Escera and Malmierca 2014), P3 component neural sources relate to those identified as the executive control system (Eichele et al. 2005) and to predictive processing of temporal structure such as the cerebellum (Kotz et al. 2014).

4.2 Large-scale network entrainment to task-relevant sequential organization: an 'active sensing' perspective

The network entrained to the rates of the task-relevant perceptual organizations exhibited major activations in auditory and premotor cortical regions. Despite coarse spatial resolution

of EEG and sLORETA inverse solution, this pattern of activations fits strikingly well with existing literature in rhythm perception and selective attention (Chen et al. 2008a, 2008b; Besle et al. 2011). In addition, there were smaller but consistent activations at the SMA, the SMC, the DLPFC, the IPL and the CC. Slight functional differences among task conditions were observed, in which the dPMC, the SMA and the CC were only involved when performing the duration task. This may be due to differences in task requirements, possibly indexing an enhanced activation of the executive control network to deal with greater conflicting stimulus information (Fan et al. 2005). The duration task required grouping of all adjacent tones into short patterns while inhibiting the intrusion of salient loudness differences, whereas the intensity task required inhibiting the intrusion of tone duration differences in the detection of intensity patterns. This difference is consistent with fMRI evidence showing increased sensitivity of the dPMC to the hierarchical structure of rhythmic sequences (Chen et al. 2008a, 2008b). Although there was no significant performance difference between tasks, there was a trend toward significance (p=0.051). Thus, it leaves open the question as to how much the difficulty level contributed to the difference in activation observed. Results from previous studies using stimulus duration tasks have shown increased SMA activations with task difficulty (Ferrandez et al. 2003; Livesey et al. 2007; Tregellas et al. 2006). However, even if the DPT was more difficult than the IPT, the finding of increased SMA activation occurred at exactly 1.25 Hz was the duration pattern rate. Thus, a very finely tuned effect was associated with the DPT, suggesting that a network pulsating at the attended pattern rates includes all regions necessary to fulfill the rhythmic task. Regardless of task differences, the described network greatly overlaps with that related to “attention in time” (Besle et al. 2011; Nobre et al. 2007).

Our findings provide compelling support for *active sensing* (Schroeder et al. 2010), a theoretical proposition that harmonically combines the roles of predictive processing and selective attention in perception. Active sensing refers to the use of motor sampling routines, generally rhythmic, to acquire sensory information. It capitalizes on the idea that the motor system generates predictions guided by top-down attentional modulation to control sensory input by directing sensory organs. Crucially for audition, it also influences sensory processing via corollary discharge signals marking the timing of behaviorally relevant events (Crapse and Sommer 2008; Morillon et al. 2015). Indeed, behavioral evidence shows that producing a rhythmic movement sharpens the temporal selection of auditory information (Morillon et al. 2014) and resets the perceptual organization of an auditory scene (Kondo et al. 2012). Whereas our study precludes ascertaining the directionality of the influences between the auditory and motor systems, we suggest that salient features of the stimulation, such as presentation rate and loudness differences, would drive rhythmic entrainment of lower sensory levels and the larger-scale network would couple and reinforce them at a later processing stage (Morillon et al. 2015). Conversely, sequence reorganization according to weaker melodic cues (e.g., tone duration) would necessitate a stronger imposition of a motor sampling routine. Further research using functional connectivity analysis with EEG/MEG data could help in elucidating this conjecture.

4.3 Integrating theoretical views

As discussed above, our key finding is the selective, goal-directed entrainment of an auditory-sensorimotor network underlying the task-based organization of a sound sequence. But, what do we specifically mean by entrainment? And, why do we consider it instrumental for the selection of internal models of the auditory scene?

When stimuli appear in a rhythmic fashion, which is often the case in audition, neuronal oscillations can be entrained (i.e., synchronized) to the external rhythms in a way such that their phase aligns predictively to the incoming inputs (Lakatos et al. 2008). Entrainment results in a spectrotemporal filter that amplifies stimuli falling in high excitation phases of the underlying oscillation and inhibits those falling in low excitation phases (Lakatos et al. 2013). This provides a neurophysiological mechanism for attention by which top-down and bottom-up signals find their intended targets, regulating network interactions and enabling a flexible change in functional connectivity (Fries 2005). Selective entrainment of oscillatory activity has been shown to exert a mechanistic role in neuronal response regulation, behavior, and even speech intelligibility (Besle et al. 2011; Ding and Simon 2012; Giraud and Poeppel 2012; Henry et al. 2014; Horton et al. 2013; Mesgarani and Chang 2012; Schroeder and Lakatos 2009; Stefanics et al. 2010; Xiang et al. 2010; Zion Golumbic et al. 2013). Several studies have shown the involvement of oscillatory entrainment in selecting and enhancing the processing of an attended speaker in competing speech streams settings over the unattended speech, akin to the “cocktail-party” phenomenon. For instance, it has been shown that the attended and the unattended speaker streams are processed independently (Ding and Simon 2012). Representation of the attended stream is distributed over a wider neural network involving high-level regions, while representation of the unattended stream is attenuated (Horton et al. 2013) and remains circumscribed to auditory cortices (Zion Golumbic et al. 2013). Our results using pure tone patterned sequences replicate these findings. A difference between using pure tones and speech is that the spectrotemporal dynamics of a speaker present dissociable features from those of another speaker, where pure tones do not have this level of stream coherence. Thus, even if input from speech streams are greatly overlapping, there is only one feasible model per speaker to explain the sensory input, in contrast, pure tones can be ascribed to more than one equally possible internal model. This renders our oscillatory entrainment results even more compelling, because they appear to underlie the ability to perceptually reorganize an ambiguous acoustic sequence.

It is important to discern *genuine* oscillatory entrainment from a possible *artefactual* steady-state oscillation that could arise from the differential processing of a particular tone in the sequence that is attended at regular times (e.g. the pattern downbeat). This is because the neural activity that a sound evokes is a non-sinusoidal waveform with rich spectral content. Thus, to provide additional evidence to support our interpretation that sound pattern perception is based on genuinely entrained oscillatory activity, and not the enhanced evoked activity, we have appended a full set of analyses on evoked responses as Supplementary Material (text and Figs.S3–S4). For the sake of brevity, these will not be further discussed here. Oscillatory entrainment has been proposed as the neuronal correlate of dynamic attending (Herrmann and Henry 2014), a theory that conceptualizes rhythmic waxing and

waning of attention as a dynamic system that can be described in terms of an oscillation coupled to temporally regular stimulation (Jones and Boltz 1989; Large and Jones 1999). This theory, and its neurophysiological counterpart, can provide meaningful insights into our results because there is a crucial difference between the widely used ambiguous streaming paradigms (the "van Noorden ABA_ paradigm", van Noorden 1975), and our sequence. Whereas the "ABA" entails spontaneous perceptual switches (Denham et al. 2014; Pressnitzer and Hupe 2006; Sterzer et al. 2009; Winkler et al. 2012) attributed to low-level physiological factors, such as noise levels and adaptation (Kang and Blake 2010), our sequence was designed to be voluntarily reorganized by selectively attending to mutually exclusive rhythmic organizations (Sussman et al. 2002; Sussman et al. 2014). In this regard, our paradigm sits between the phenomena of streaming and that of metrical interpretation (Repp 2007), the subjective allocation of a beat or a temporal anchor around which other events are organized. Indeed, previous research has demonstrated a link between metrical interpretation and oscillatory entrainment by showing that neuronal oscillations not only entrain to an existing imposed repetitive beat but also to its imagined metrical organization (Nozaradan et al. 2011).

Linking the concepts of dynamic attending, active sensing, and metrical interpretation with oscillatory entrainment provides an integrative theoretical view to our results. While feature and temporal relations between stimuli in our ambiguous sequence remain constant, and are encoded simultaneously in the auditory system, dynamic attention via motor corollary discharges allocates a subjective anchor to which subsequent events unfolding predictively are bound by a large-scale entrained brain oscillation. The resulting global representation describing the auditory scene would underlie what we subjectively experience as an auditory object. Simply put, "my experience is what I agree to attend to" (James 1890).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health (R01 DC004263), the SGR2014-177 grant from the Generalitat de Catalunya, the grant PSI2015-63664-P from MINECO, a FPU grant AP2007-01084 awarded to J.C.F, and the ICREA Acadèmia Distinguished Professorship awarded to C.E.

References

Baars BJ. Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. Prog Brain Res. 2005; 150:45–53. [PubMed: 16186014]

Bar M. The proactive brain: memory for predictions. Philos Trans R Soc Lond B Biol Sci. 2009; 364:1235–1243. [PubMed: 19528004]

Baldeweg T. Repetition effects to sounds: evidence for predictive coding in the auditory system. Trends Cogn Sci. 2006; 10:93–4. [PubMed: 16460994]

Besle J, Schevon CA, Mehta AD, Lakatos P, Goodman RR, McKhann GM, Emerson RG, Schroeder CE. Tuning of the human neocortex to the temporal dynamics of attended events. J Neurosci. 2011; 31:3176–85. [PubMed: 21368029]

Bregman, AS. Auditory scene analysis : the perceptual organization of sound. Cambridge (MA): A Bradford book, The MIT Press; 1990.

Bregman AS. Auditory streaming is cumulative. J Exp Psychol Hum Percept Perform. 1978; 4:380–387. [PubMed: 681887]

Chen JL, Penhune VB, Zatorre RJ. Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. J Cogn Neurosci. 2008a; 20:226–239. [PubMed: 18275331]

Chen JL, Penhune VB, Zatorre RJ. Listening to musical rhythms recruits motor regions of the brain. Cereb Cortex. 2008b; 18:2844–2854. [PubMed: 18388350]

Costa-Faidella J, Grimm S, Slabu L, Díaz-Santaella F, Escera C. Multiple time scales of adaptation in the auditory system as revealed by human evoked potentials. Psychophysiology. 2011; 48:774–783. [PubMed: 20946129]

Crapse TB, Sommer MA. Corollary discharge across the animal kingdom. Nat Rev Neurosci. 2008; 9:587–600. [PubMed: 18641666]

Dehaene S, Kerszberg M, Changeux JP. A neuronal model of a global workspace in effortful cognitive tasks. Proc Natl Acad Sci USA. 1998; 95:14529–34. [PubMed: 9826734]

Dehaene S, Naccache L. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. Cognition. 2001; 79:1–37. [PubMed: 11164022]

Deike S, Heil P, Böckmann-Barthel M, Brechmann A. The Build-up of Auditory Stream Segregation: A Different Perspective. Front Psychol. 2012; 3:461. [PubMed: 23118731]

Delorme A, Makeig S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. 2004; 134:9–21. [PubMed: 15102499]

Denham S, Bõhm TM, Bendixen A, Szalárdy O, Kocsis Z, Mill R, Winkler I. Stable individual characteristics in the perception of multiple embedded patterns in multistable auditory stimuli. Front Neurosci. 2014; doi: 10.3389/fnins.2014.00025

Ding N, Simon JZ. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci U S A. 2012; 109(29):11854–9. [PubMed: 22753470]

Dowling WJ. The perception of interleaved melodies. Cogn Psychol. 1973; 5:322–337.

Eichele T, Specht K, Moosmann M, Jongsma MLA, Quiroga RQ, Nordby H, Hugdahl K. Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI. Proc Natl Acad Sci U S A. 2005; 102:17798–803. [PubMed: 16314575]

Elhilali M, Xiang J, Shamma SA, Simon JZ. Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. PLoS Biol. 2009; 7:e1000129.doi: 10.1371/journal.pbio.1000129 [PubMed: 19529760]

Escera C, Malmierca MS. The auditory novelty system: An attempt to integrate human and animal research. Psychophysiology. 2014; 51:111–123. [PubMed: 24423134]

Fan J, McCandliss BD, Fossella J, Flombaum JI, Posner MI. The activation of attentional networks. Neuroimage. 2005; 26:471–479. [PubMed: 15907304]

Ferrandez AM, Hugueville L, Lehéricy S, Poline JB, Marsault C, Pouthas V. Basal ganglia and supplementary motor area subtend duration perception: an fMRI study. Neuroimage. 2003; 19(4): 1532–44. [PubMed: 12948709]

Fries P. A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. Trends Cogn Sci. 2005; 9:474–480. [PubMed: 16150631]

Friston K. A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci. 2005; 360:815–36. [PubMed: 15937014]

Fuchs M, Kastner J, Wagner M, Hawes S, Ebersole JS. A standardized boundary element method volume conductor model. Clin Neurophysiol. 2002; 113:702–712. [PubMed: 11976050]

Garrido M, Kilner J, Stephan K, Friston K. The mismatch negativity: a review of underlying mechanisms. Clin Neurophysiol. 2009; 120:453–63. [PubMed: 19181570]

Giraud AL, Poeppel D. Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci. 2012; 15:511–7. [PubMed: 22426255]

Gregory RL. Perceptions as hypotheses. Philos Trans R Soc Lond B Biol Sci. 1980; 290:181–197. [PubMed: 6106237]

Grossberg S, Govindarajanb L, Wysec K, Cohen M. ARTSTREAM: a neural network model of auditory scene analysis and source segregation. Neural Netw. 2004; 17:511–536. [PubMed: 15109681]

von Helmholtz, H. Treatise on Physiological Optics. Vol. III. New York (NY): Dover Publications; 1866.

Henry MJ, Herrmann B, Obleser J. Entrained neural oscillations in multiple frequency bands comodulate behavior. Proc Natl Acad Sci U S A. 2014; 111:14935–14940. [PubMed: 25267634]

Herrmann B, Henry MJ. Low-Frequency Neural Oscillations Support Dynamic Attending in Temporal Context. Timing Time Percept. 2014; 2:62–86.

Horton C, D'Zmura M, Srinivasan R. Suppression of competing speech through entrainment of cortical oscillations. J Neurophysiol. 2013; 109(12):3082–93. [PubMed: 23515789]

Horvath J, Czigler I, Sussman E, Winkler I. Simultaneously active pre-attentive representations of local and global rules for sound sequences in the human brain. Cogn Brain Res. 2001; 12:131–144.

James, W. Principles of Psychology. Vol. I. New York (NY): Dover Publications; 1890.

John MS, Dimitrijevic A, van Roon P, Picton TW. Multiple auditory steady-state responses to AM and FM stimuli. Audiol Neurootol. 2001; 6:12–27. [PubMed: 11173772]

John MS, Dimitrijevic A, Picton TW. Auditory steady-state responses to exponential modulation envelopes. Ear & Hearing. 2002; 23:106–117. [PubMed: 11951847]

Jones MR, Boltz M. Dynamic Attending and Responses to Time. Psychol Rev. 1989; 96:459–491. [PubMed: 2756068]

Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clin Neurophysiol. 2000; 111:1745–1758. [PubMed: 11018488]

Jurcak V, Tsuzuki D, Dan I. 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. Neuroimage. 2007; 34:1600–1611. [PubMed: 17207640]

Kayser J, Tencke CE. Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: II. Adequacy of low-density estimates. Clin Neurophysiol. 2006; 117:369–80. [PubMed: 16356768]

Kang MS, Blake R. What causes alternations in dominance during binocular rivalry? Atten Percept Psychophys. 2010; 72:179–186. [PubMed: 20045887]

Kohler, W. Gestalt psychology: An introduction to new concepts in modern psychology. New York (NY): Liveright; 1947.

Kondo HM, Pressnitzer D, Toshima I, Kashino M. Effects of self-motion on auditory scene analysis. Proc Natl Acad Sci U S A. 2012; 109:6775–80. [PubMed: 22493250]

Kotz SA, Stockert A, Schwartze M. Cerebellum, temporal predictability and the updating of a mental model. Philos Trans R Soc Lond B Biol Sci. 2014; 369:20130403. [PubMed: 25385781]

Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE. Entrainment of neuronal oscillations as a mechanism of attentional selection. Science. 2008; 320:110–3. [PubMed: 18388295]

Lakatos P, Musacchia G, O'Connell MN, Falchier AY, Javitt DC, Schroeder CE. The Spectrotemporal Filter Mechanism of Auditory Selective Attention. Neuron. 2013; 77:750–761. [PubMed: 23439126]

Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT. Automated Tailairach Atlas Labels for Functional Brain Mapping. Hum Brain Mapp. 2000; 10:120–131. [PubMed: 10912591]

Large EW, Jones MR. the Dynamics of Attending: How People Track Time-Varying Events. Psychol Rev. 1999; 106:119–159.

Livesey AC, Wall MB, Smith AT. Time perception: manipulation of task difficulty dissociates clock functions from other cognitive demands. Neuropsychologia. 2007; 45(2):321–31. 2007. [PubMed: 16934301]

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

Luo H, Wang Y, Poeppel D, Simon JZ. Concurrent encoding of frequency and amplitude modulation in human auditory cortex: MEG evidence. J Neurophysiol. 2006; 96:2712–2723. [PubMed: 16510774]

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

Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, et al. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos Trans R Soc Lond B Biol Sci. 2001; 356:1293–322. [PubMed: 11545704]

Mesgarani N, Chang EF. Selective cortical representation of attended speaker in multi-talker speech perception. Nature. 2012; 485:233–6. [PubMed: 22522927]

Morillon B, Hackett TA, Kajikawa Y, Schroeder CE. Predictive motor control of sensory dynamics in auditory active sensing. Curr Opin Neurobiol. 2015; 31:230–238. [PubMed: 25594376]

Morillon B, Schroeder CE, Wyart V. Motor contributions to the temporal precision of auditory attention. Nat Commun. 2014; 5:5255. [PubMed: 25314898]

Näätänen R, Gaillard AWK, Mantysalo S. Early Selective Attention Effect on Evoked Potential Reinterpreted. Acta Psychol (Amst). 1978; 42:313–329. [PubMed: 685709]

Näätänen R, Picton T. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology. 1987; 24:375–425. [PubMed: 3615753]

Nelken I, Bar-Yosef O. Neurons and objects: The case of auditory cortex. Front Neurosci. 2009; 2:107–113.

Nichols TE, Holmes AP. Nonparametric Permutation Tests for {PET} functional Neuroimaging Experiments: A Primer with examples. Hum Brain Mapp. 2001; 15:1–25.

Nobre A, Correa A, Coull J. The hazards of time. Curr Opin Neurobiol. 2007; 17:465–470. [PubMed: 17709239]

Nozaradan S, Peretz I, Missal M, Mouraux A. Tagging the neuronal entrainment to beat and meter. J Neurosci. 2011; 31:10234–40. [PubMed: 21753000]

Oostenveld R, Fries P, Maris E, Schoffelen JM. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci. 2011; 2011:156869. [PubMed: 21253357]

Pannese A, Herrmann CS, Sussman E. Analyzing the Auditory Scene: Neurophysiologic Evidence of a Dissociation Between Detection of Regularity and Detection of Change. Brain Topogr. 2015; 28:411–422. [PubMed: 24771006]

Pascual-Marqui RD. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Methods Find Exp Clin Pharmacol. 2002; 24(Suppl D):5–12. [PubMed: 12575463]

Patel, AD. Language, music, and the brain, Language music and the brain. New York (NY): Oxford University Press; 2008.

Pressnitzer D, Hupe JM. Temporal Dynamics of Auditory and Visual Bistability Reveal Common Principles of Perceptual Organization. Curr Biol. 2006; 16:1351–1357. [PubMed: 16824924]

Repp BH. Hearing a melody in different ways: Multistability of metrical interpretation, reflected in rate limits of sensorimotor synchronization. Cognition. 2007; 102:434–454. [PubMed: 16545791]

Rizzolatti G, Craighero L. the Mirror-Neuron System. Annu Rev Neurosci. 2004; 27:169–192. [PubMed: 15217330]

Schroeder CE, Lakatos P. Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. 2009; 32:9–18. [PubMed: 19012975]

Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P. Dynamics of Active Sensing and perceptual selection. Curr Opin Neurobiol. 2010; 20:172–176. [PubMed: 20307966]

Shamma SA, Elhilali M, Micheyl C. Temporal coherence and attention in auditory scene analysis. Trends Neurosci. 2011; 34:114–123. [PubMed: 21196054]

Stefanics G, Hangya B, Hernadi I, Winkler I, Lakatos P, Ulbert I. Phase Entrainment of Human Delta Oscillations Can Mediate the Effects of Expectation on Reaction Speed. J Neurosci. 2010; 30:13578–13585. [PubMed: 20943899]

Sterzer P, Kleinschmidt A, Rees G. The neural bases of multistable perception. Trends Cogn Sci. 2009; 13:310–318. [PubMed: 19540794]

Sussman-Fort J, Sussman E. The effect of stimulus context on the buildup to stream segregation. Front Neurosci. 2014; 8:93. [PubMed: 24808822]

Sussman E, Horváth J, Winkler I, Orr M. The role of attention in the formation of auditory streams. Percept Psychophys. 2007; 69:136–152. [PubMed: 17515223]

Sussman E, Ritter W, Vaughan HG Jr. An investigation of the auditory streaming effect using event-related brain potentials. Psychophysiology. 1999; 36:22–34. [PubMed: 10098377]

Sussman E, Ritter W, Vaughan HG. Attention affects the organization of auditory input associated with the mismatch negativity system. Brain Res. 1998; 789:130–138. [PubMed: 9602095]

Sussman E, Steinschneider M. Attention effects on auditory scene analysis in children. Neuropsychologia. 2009; 47:771–785. [PubMed: 19124031]

Sussman E, Winkler I, Huotilainen M, Ritter W, Näätänen R. Top-down effects can modify the initially stimulus-driven auditory organization. Cogn Brain Res. 2002; 13:393–405.

Sussman ES. Attention Matters: Pitch vs. Pattern Processing in Adolescence. Front Psychol. 2013; 4:1–8. [PubMed: 23382719]

Sussman ES, Bregman AS, Lee WW. Effects of task-switching on neural representations of ambiguous sound input. Neuropsychologia. 2014; 64:218–229. [PubMed: 25281308]

Sussman ES, Steinschneider M. Attention modifies sound level detection in young children. Dev Cogn Neurosci. 2011; 1:351–360. [PubMed: 21808660]

Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J. Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. J Neurosci. 1996; 16:4240–9. [PubMed: 8753885]

Tregellas JR, Davalos DB, Rojas DC. Effect of task difficulty on the functional anatomy of temporal processing. Neuroimage. 2006; 32(1):307–15. [PubMed: 16624580]

Ulanovsky N, Las L, Farkas D, Nelken I. Multiple Time Scales of Adaptation in Auditory Cortex Neurons. Blood Press. 2004; 24:10440–10453.

van Noorden, LPAS. Temporal coherence in the perception of tone sequences. Eindhoven (Netherlands): Institute for Perceptual Research; 1975.

Welch PD. The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. IEEE Trans Audio Electroacoust. 1967; 15:70–73.

Winkler I, Denham S, Mill R, Bohm TM, Bendixen A. Multistability in auditory stream segregation: A predictive coding view. Philos Trans R Soc B Biol Sci. 2012; 367:1001–1012.

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

Winkler I, Van Zuijen TL, Sussman E, Horvath J, Naatanen R. Object representation in the human auditory system. Eur J Neurosci. 2006; 24:625–634. [PubMed: 16836636]

Xiang J, Simon J, Elhilali M. Competing streams at the cocktail party: exploring the mechanisms of attention and temporal integration. J Neurosci. 2010; 30:12084–93. [PubMed: 20826671]

Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, et al. Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. Neuron. 2013; 77:980–991. [PubMed: 23473326]

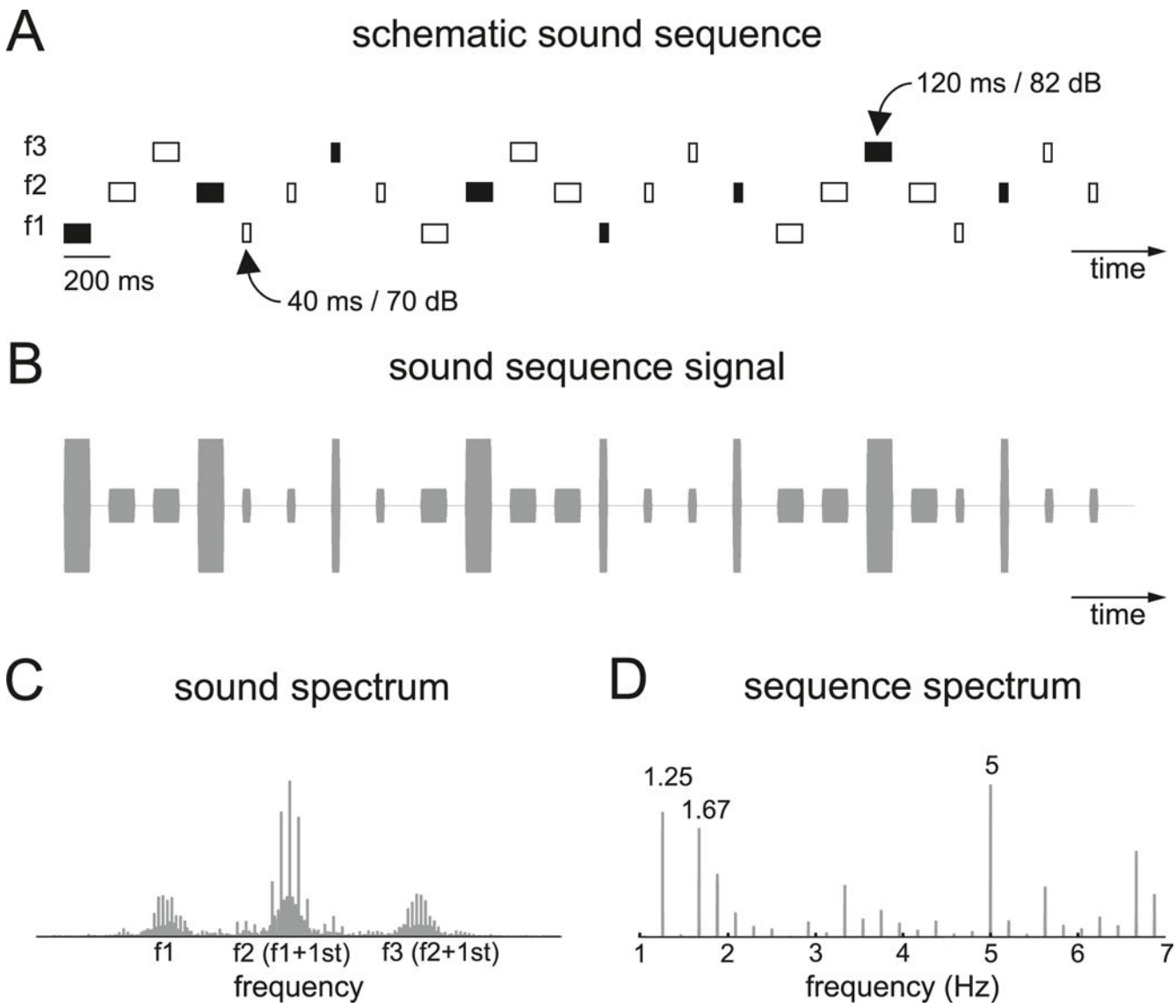

Figure 1. Acoustic features of the sound sequence

A) Schematic illustration of the sound sequence. **B)** Acoustic signal of the sound sequence. **C)** Frequency decomposition of the acoustic signal in the high frequency range. **D)** Frequency decomposition of the acoustic signal in the low frequency range. Note the three main rhythms of interest: 1.25 Hz, corresponding to the duration pattern; 1.67 Hz, corresponding to the intensity pattern; and 5 Hz, corresponding to the tone presentation rate (*See also sound1.mp3*).

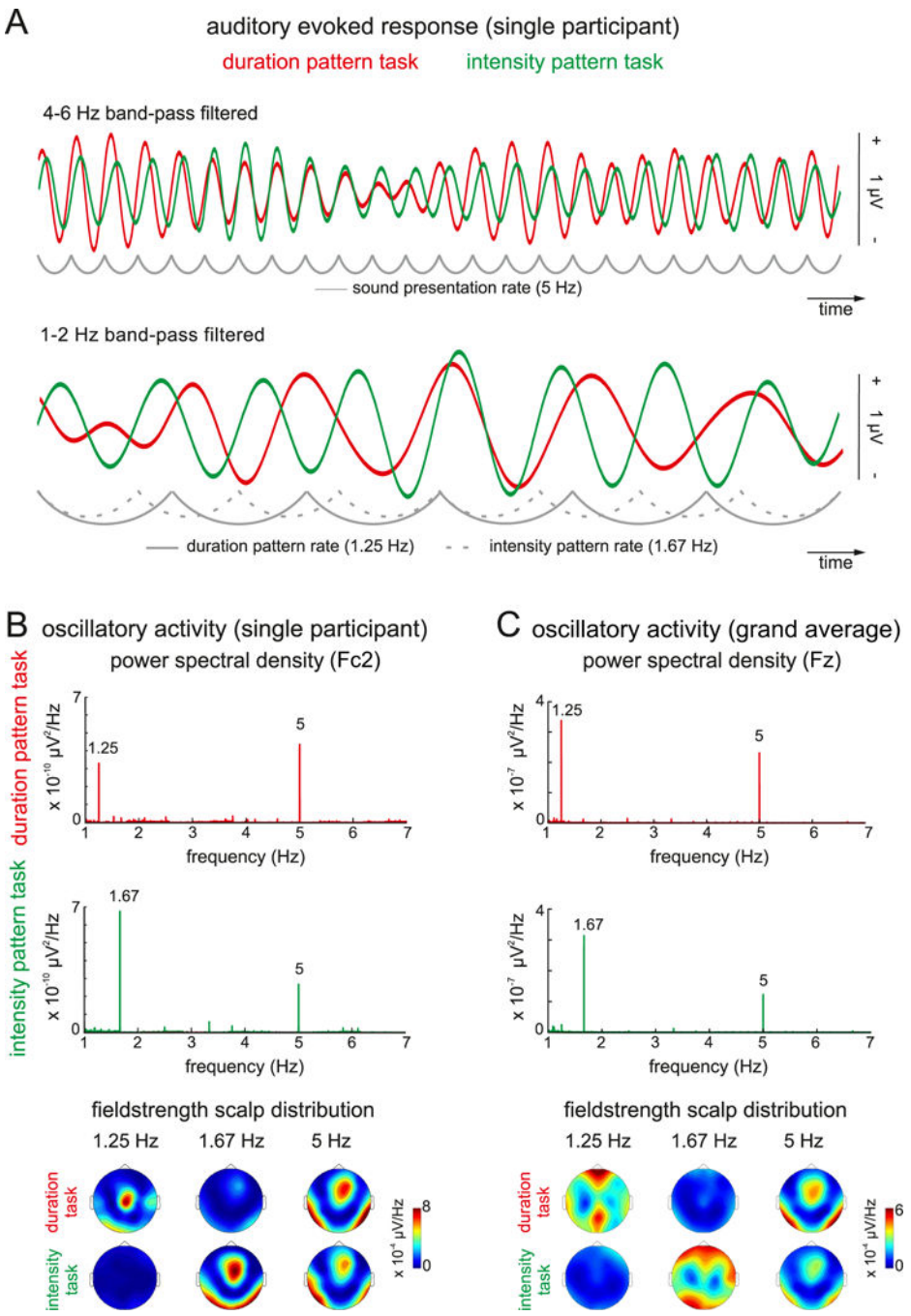

Figure 2. Task dependent neural responses

A) Auditory evoked response to the sound sequence by a representative participant (frontocentral electrode [FCz]) while performing the duration pattern task (DPT; *red*) and the intensity pattern task (IPT; *green*). **B)** Frequency decomposition of the same participant's neural activity. Note the fine-tuning of neural activity to the rhythms of interest and the specificity of the neural representation of the attended pattern rhythm. Below, topographical scalp distribution of fieldstrength at the rhythms of interest. **C)** Same as in B), but for the averaged data from all participants (Fz electrode; participant's data was zero padded when

necessary to match the length of the longest data segment). Note the difference in the fieldstrength scalp distribution between the activity at the rate of the attended pattern vs. activity at the rate of stimulation.

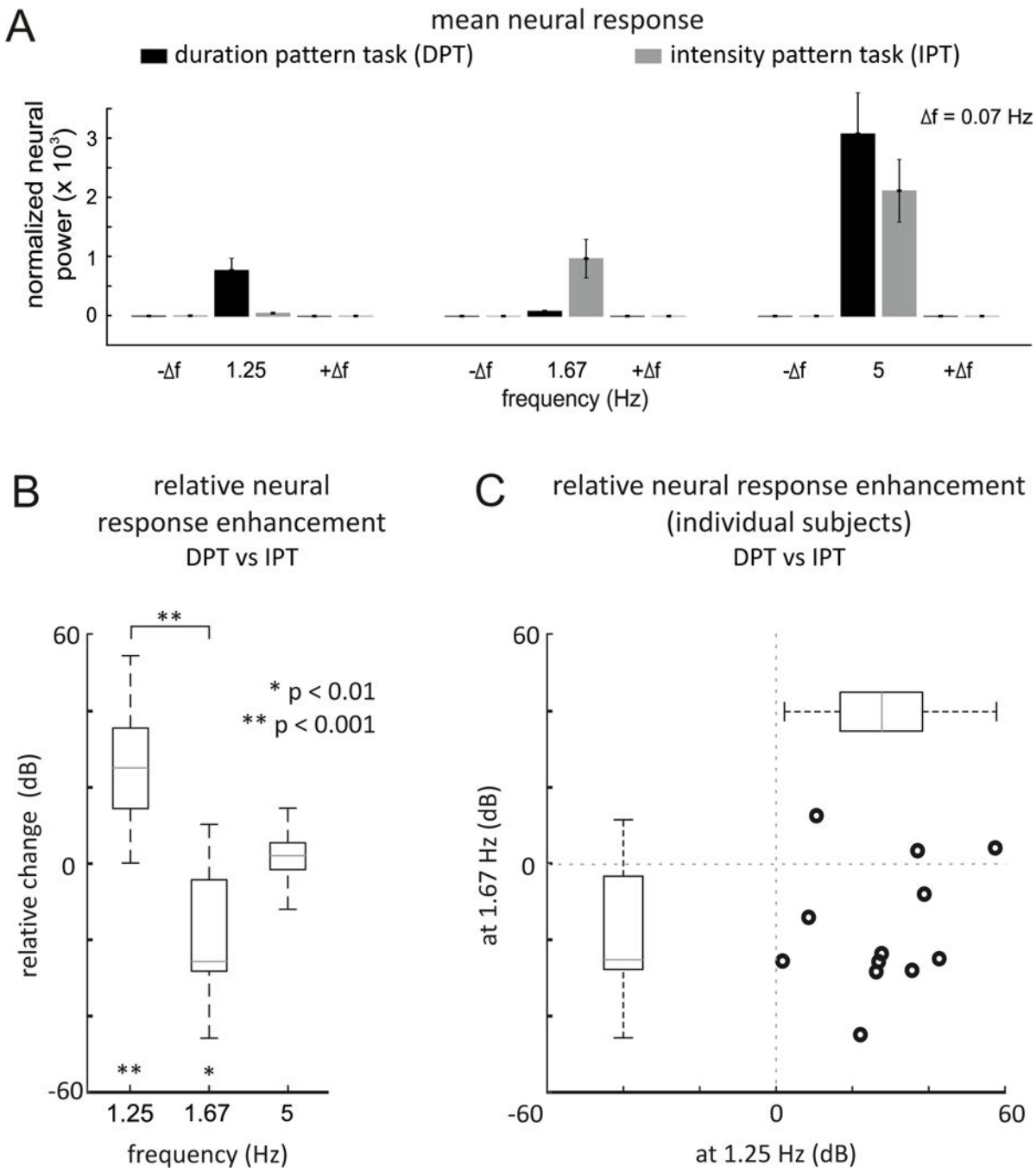

Figure 3. Neural response specificity and relative change

A) Mean neural response averaged across all participants. Error bars represent SEM. Note the double dissociation between neural power at the rates of the attended patterns as a function of task demands [DPT, *black*; IPT, *grey*]), and the exquisite fine-tuning of selective attention to rhythmic patterns, as frequency components only 0.07 Hz away from the rates of interest were not affected. **B)** Relative change (dB) between the neural responses elicited during the DPT vs. IPT at the three frequency components of interest (pooled participants). The boxplots represent the median value (*grey line*), the interquartile range (*full box*) and

the extreme values (*whiskers*). Significant relative enhancements were found at the frequency components corresponding to the rates of the attended sound patterns. **C)** Individual relative neural response enhancement. Note the grouping of individual participant's data in the lower right quadrant, indicating enhanced response at the attended pattern rate. The boxplots from B) are positioned to facilitate the viewing of data distribution.

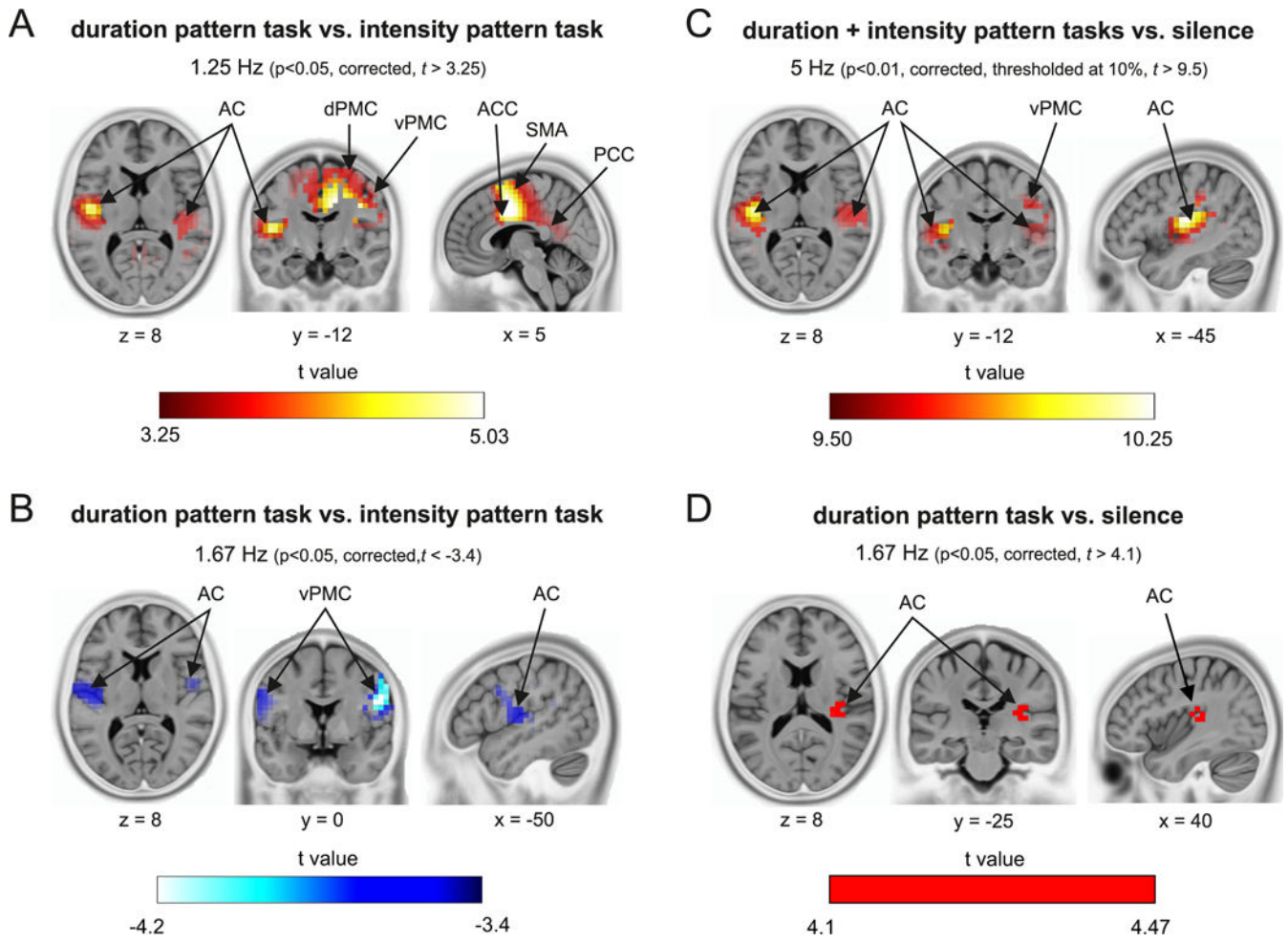

Figure 4. Source reconstruction of oscillatory entrainment

A) Significantly activated areas at the duration pattern rate (1.25 Hz) when performing the DPT vs. IPT. Perceptual organization of the sound sequence according to the duration pattern, driven by task demands, entrained neuronal oscillations in a sensory-motor network including brain regions located at the auditory cortex (AC), dorsal and ventral premotor cortex (dPMC; vPMC), supplementary motor area (SMA) and anterior and posterior cingulate cortex (ACC; PCC). **B)** Significantly activated areas at the intensity pattern rate (1.67 Hz) when performing the DPT vs. IPT. Perceptual organization of the sound sequence according to the intensity pattern, driven by task demands, entrained neuronal oscillations in a sensory-motor network including brain regions located at the AC and the vPMC. **C)** Significantly activated areas at tone presentation rate (5 Hz) when performing the DPT+IPT vs. a silent baseline (results only show the 10% of the most significant activations). Regardless of the performed task and the perceptual organization of the sequence, the strongest oscillatory entrainment appeared mainly in a network comprising AC regions. **D)** Significantly activated areas at the intensity pattern rate (1.67 Hz) when performing the DPT vs. a silent baseline (*red; not graded due to the circumscribed effects*). Note how the presentation rate of louder tones, although outside the focus of attention, nevertheless synchronized brain oscillations at the right AC (*see also* Fig. S1).

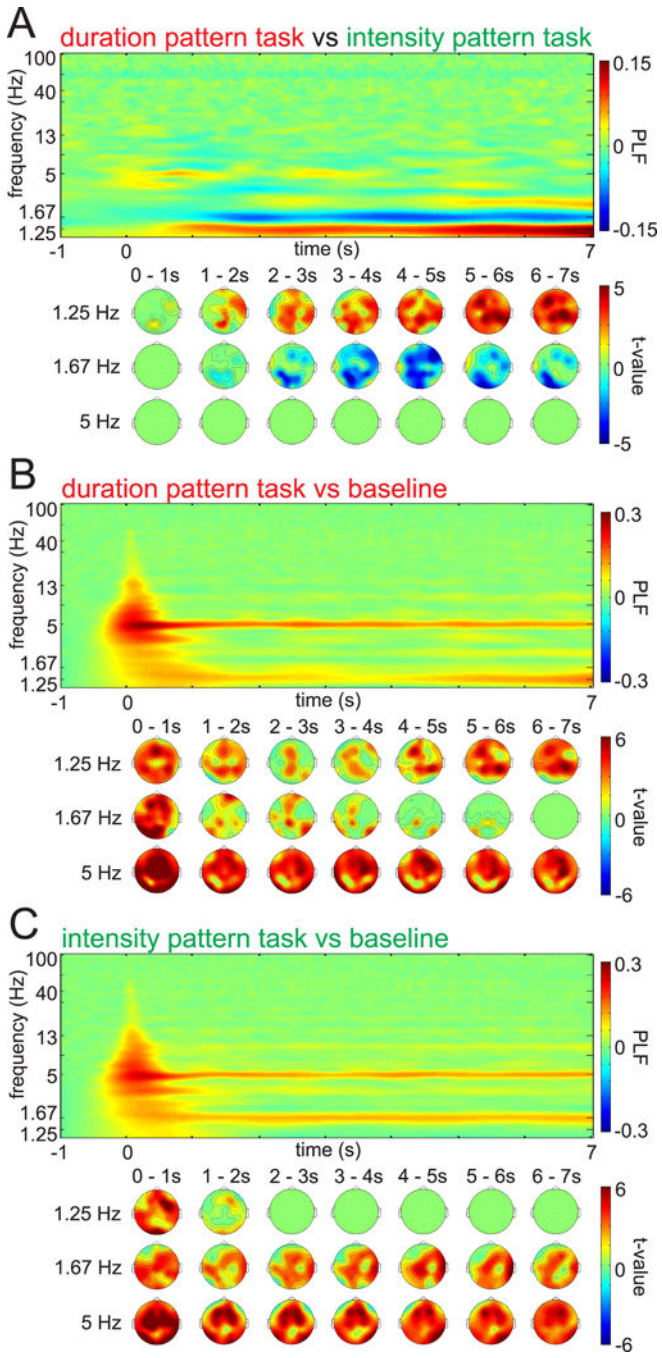

Figure 5. Time course of phase reorganization

A) Time-frequency plot illustrating the difference in phase-locking factor (PLF) when performing the DPT vs IPT for the whole acoustic sequence, pooled across subjects and electrodes. Below, evolution of the significant time-electrode clusters at the rates of the three frequencies of interest, thresholded at the t value corresponding to p<0.05 (*green, p>0.05*). Note that selective entrainment starts after the initial PLF burst and that differences at the intensity pattern rate start later than those at the duration pattern rate. **B)** Same as above but for the DPT against a silent baseline. Note the initial PLF burst spanning a broad frequency

range and the tuning of entrained oscillations at the stimulation rate (5 Hz) and at the perceived sequence organization rate (1.25 Hz). See how the entrainment to the ignored intensity pattern fades over time. **C)** Same as above, but for the IPT vs. silence. Here, the evolution of the clusters in time shows a rapid fading of oscillatory entrainment to the rate of the ignored duration pattern.

Table 1
Anatomical regions exhibiting oscillatory entrainment

This table lists the broad anatomical regions showing oscillatory entrainment to the three main rhythms of interest (presentation rate, 5 Hz; duration pattern rate, 1.25 Hz; intensity pattern rate, 1.67 Hz), for both hemispheres separately, when subjects performed the DPT and the IPT. The table includes the number of activated voxels within brain region (cluster size) and the statistic of the voxel showing maximal activation (peak voxel). AC = Auditory Cortex; ACC = Anterior Cingluate Cortex; DLPFC = DorsoLateral PreFrontal Cortex; dPMC/SMA = dorsal PreMotor Cortex/Supplementary Motor Area; HG = Heschl's Gyrus; IFG = Inferior Frontal Gyrus; IPL = Inferior Parietal Lobule; MTG = Mid Temporal Gyrus; OT = OccipitoTemporal Area; PCC = Posterior Cingulate Cortex; SMC = SensoryMotor Cortex; STG = Superior Temporal Gyrus; vPMC/BA44 = ventral PreMotor Cortex/BA44.

Condition	Brain Region		Cluster Size	Peak Voxel *t* value
DPT vs IPT 1.25 Hz (p<0.05)	AC	L	117	4,92
		R	98	4,71
	ACC	L	65	4,56
		R	97	5,59
	DLPFC	R	20	4,64
	dPMC/SMA	L	111	4,33
		R	170	5,22
	OT	R	8	3,40
	IPL	L	13	4,17
		R	6	3,79
	PCC	L	54	4,12
		R	90	4,46
	SMC	L	34	4,16
		R	78	4,08
	vPMC/BA44	L	29	4,40
		R	80	5,13
DPT vs IPT 1.67 Hz (p<0.05)	AC	L	53	−3,75
		R	8	−3,87
	DLPFC	L	8	−3,70
		R	8	−4,22
	IPL	R	9	−3,91
	SMC	L	9	−3,61
		R	38	−4,66
	vPMC/BA44	L	18	−3,78
		R	42	−4,55
DPT & IPT vs Silence 5 Hz (p<0.01, only upper 10%)	AC	L	174	12,50
		R	124	11,13
	IPL	L	33	11,85
		R	43	11,92

Condition	Brain Region		Cluster Size	Peak Voxel *t* value
	PCC	L	12	9,85
		R	24	9,80
	SMC	L	17	11,34
		R	53	12,11
	vPMC/BA44	R	31	10,79
DPT vs Silence 1.67 Hz ($p<0.05$)	AC	R	14	4,47