HHS Public Access

Author manuscript

Brain Topogr. Author manuscript; available in PMC 2015 May 01.

Published in final edited form as:

Brain Topogr. 2015 May ; 28(3): 411–422. doi:10.1007/s10548-014-0368-4.

Analyzing the Auditory Scene: Neurophysiologic Evidence of a Dissociation Between Detection of Regularity and Detection of Change

Alessia Pannese,

The Italian Academy for Advanced Studies, Columbia, University, 1161 Amsterdam Avenue, New York, NY 10027, USA; Department of Biological Sciences, Columbia University, New York, NY 10027, USA

Christoph S. Herrmann, and

Experimental Psychology Lab Center for Excellence “Hearing4all”, European Medical School, Carl von Ossietzky Universität Oldenburg, 26111 Oldenburg, Germany

Elyse Sussman

Dominick P. Purpura Department of Neuroscience, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA; Department of Otorhinolaryngology-HNS, Albert Einstein, College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA

Alessia Pannese: ap2215@caa.columbia.edu; Elyse Sussman: elyse.sussman@einstein.yu.edu

Abstract

Detecting regularity and change in the environment is crucial for survival, as it enables making predictions about the world and informing goal-directed behavior. In the auditory modality, the detection of regularity involves segregating incoming sounds into distinct perceptual objects (stream segregation). The detection of change from this within-stream regularity is associated with the mismatch negativity, a component of auditory event-related brain potentials (ERPs). A central unanswered question is how the detection of regularity and the detection of change are interrelated, and whether attention affects the former, the latter, or both. Here we show that the detection of regularity and the detection of change can be empirically dissociated, and that attention modulates the detection of change without precluding the detection of regularity, and the perceptual organization of the auditory background into distinct streams. By applying frequency spectra analysis on the EEG of subjects engaged in a selective listening task, we found distinct peaks of ERP synchronization, corresponding to the rhythm of the frequency streams, independently of whether the stream was attended or ignored. Our results provide direct neurophysiological evidence of regularity detection in the auditory background, and show that it can occur independently of change detection and in the absence of attention.

Correspondence to: Alessia Pannese, ap2215@caa.columbia.edu; Elyse Sussman, elyse.sussman@einstein.yu.edu.

Present Address: A. Pannese, Swiss Centre for Affective Sciences, University of Geneva, Geneva, Switzerland

Electronic supplementary material: The online version of this article (doi:10.1007/s10548-014-0368-4) contains supplementary material, which is available to authorized users.

Conflict of interest: The authors declare no competing financial interests.

Keywords

Auditory scene analysis; Stream segregation; Mismatch negativity; Event-related potentials

Introduction

Natural auditory environments provide a continuous stream of information originating from multiple overlapping sources, which sensory systems have evolved to filter, analyze and organize (Bregman and Campbell 1971). The process of extracting potentially meaningful patterns from the continuous flow of acoustic input has been referred to as auditory scene analysis (Bregman 1990). An example of this is the cocktail party problem (Cherry 1953), where the listener is faced with the task of listening to one sound stream (e.g., a friend talking) in the presence of competing background sounds (e.g., other voices). Solving the cocktail party problem involves two distinct challenges: (1) integrating and segregating the mixture of sounds that enters the ears to provide neural representations of the distinct sources; and (2) filtering out the unattended background while selectively listening to the stream of interest (Bregman 1978; Winkler et al. 2005). A fundamental issue still unresolved is how the auditory system processes the unattended background when there are multiple sources: as an undifferentiated whole, or as structured into distinct streams. Although we have speculated that stream segregation is an automatic process (e.g., Sussman et al. 1999; Sussman 2005), some recent results indicate that attention to a subset frequency of a sound mixture precludes segregation of the unattended sounds, when the unattended sounds have more than one potential frequency stream (Sussman et al. 2005). This was indicated by the absence of a neurophysiologic indicator of change detection for the unattended sound streams. These results could not resolve the issue of whether attention to one of many frequency streams precluded segregation for the background, unattended sounds, or whether attention mediated the change detection process. Thus, the question is still open, and we hypothesize that whereas the two processes (regularity detection and change detection) are interdependent upon each other when a change is detected (Sussman 2007), they are also distinct. Specifically, we hypothesize that regularity detection is necessary but not sufficient for change detection. If this is true, then indices of regularity detection and change detection can be dissociated. Hence, we addressed the question: can the processes of regularity detection and change detection be dissociated?

This question is interesting because attention is a limited resource. Therefore, some aspects of sound processing, such as those related to differentiating input by structuring it into separate streams based on regularity in their spectro-temporal characteristics, should occur irrespective of the direction of attention. This could facilitate the ability to focus attention to different streams when there are multiple simultaneous sources, and explain the ease with which we can follow events around a crowded room. However, the issue of whether or not attention is necessary for regularity detection and stream segregation to occur is unresolved, with some studies providing evidence for (Carlyon et al. 2001; Sussman et al. 1999), others against (Winkler et al. 2003; Sussman et al. 2007), and others suggesting that attention has little impact on frequency-based segregation, but rather modulates the build-up process of streaming, the integration of sounds over time (Snyder et al. 2006; Elhilali et al. 2009). Our

approach for investigating this problem involves measuring the mismatch negativity (MMN) component of event-related brain potentials (ERPs) and the spectral content of the time-locked neural activity. The MMN is a neurophysiological index of change detection (Näätänen et al. 1978; Sussman et al. 2005), thus providing an indirect index of regularity detection and stream segregation because sound change detection is predicated on first extracting the ongoing regularities in the input (Sussman 2007). Change detection can indirectly index regularity detection by the presence of MMN, but the reverse has not been proven. If there is no index of change detection does that mean that the regularity was not formed to serve as the basis of change, or that the regularity was formed but change detection was mediated by other processes (e.g., attention). That is, when MMN is not elicited, it is not possible to know at what level the system of change detection broke down. One interpretation is that the validity of the MMN as a proxy for regularity detection is limited because the two processes (regularity detection and change detection) are subserved by different neural substrates. The MMN originates within the auditory cortex, whereas the information necessary for regularity detection and stream segregation is already available subcortically, as early as in the cochlear nucleus (the first relay station of the auditory pathway) (Pressnitzer et al. 2008). Moreover, there is a temporal distinction between the two processes, shown by the segregation of sounds to streams occurring prior to the change detection process indexed by MMN (Sussman 2005; Yabe et al. 2001). Thus, although the presence of MMN can indicate that stream segregation has occurred by signaling change detection from a previously established regularity, its absence is inconclusive. Likewise, although attention has been shown to modulate the MMN, it is not known whether attention affects only the detection of regularity (which then enables within stream change detection and the output MMN), or whether attention affects change detection directly (by modulating the information that is used in the change detection process) (Sussman 2007; Sussman et al. 2013). In the present study we seek evidence for dissociation between the detection of regularity and the detection of change. By combining the use of MMN as an index of change detection, with spectral analysis of the EEG as an indicator of regularity detection, we test the hypothesis that there is a dissociation between regularity detection and change detection. This will be observed if attention modulates the change detection process (indexed by MMN) without precluding the perceptual organization of the auditory background into distinct streams (indexed by spectral analysis).

Materials and Methods

Subjects

Twelve adult volunteers participated in the study (5 females; 21–35 years, *M* = 29 years, SD = 4). Procedures were approved by the Internal Review Board and Committee for Clinical Investigations of the Albert Einstein College of Medicine, where the study was conducted. Participants gave informed consent after the experimental protocol was explained, and were compensated for their participation. All procedures were carried out in accordance with the Declaration of Helsinki. All participants passed a hearing screening (20 dB HL or better bilaterally at 500, 1,000, 2,000, and 4,000 Hz).

Stimuli

Complex tones (fundamental plus 4 harmonics), 50 ms in duration (including 5 ms linear ramps for onset and offset) were equated for intensity at 72 dB SPL using a Bruel & Kjaer (2209) sound level meter. Tones were created using Adobe Audition® software, and presented binaurally via E-a-rtones® 3A insert earphones with NeuroStim (Compumedics Inc., Texas, USA) software and hardware.

Three sets of complex tones were presented, each occupying a distinct frequency range: (1) The high-frequency range (H) included two tones (F0: H_1 = 2,489 Hz, and H_2 = 2,637 Hz); (2) The middle-frequency range (M) included three tones (F0: M_1 = 880 Hz, M_2 = 932 Hz, M_3 = 988 Hz); and (3) the low-frequency range (L) included three tones (F0: L_1 = 311 Hz, L_2 = 330 Hz, L_3 = 349 Hz). The tones were presented in the following alternating pattern: L_1, M_1, H_1, M_2, L_2, M_3, H_1, M_1, L_3, M_2, H_1, M_3, L_1, M_1, H_1, M_2 … etc. (Fig. 1). Within both the low- and middle-frequency ranges, a repeating three-tone rising pattern (L_1, L_2, L_3 and M_1, M_2, M_3, respectively) occurred 93 % of the time (*standard*) with occasional reversals of the within-stream pattern (L_3, L_2, L_1 and M_3, M_2, M_1, respectively) occurring randomly 7 % of the time within each range (*deviant*). High- and low-frequency tones appeared every 4th tone (once every 360 ms), and middle-frequency tones appeared every second tone (once every 180 ms). Thus, when the streams segregated, the rate of repetition of the within-stream middle frequency pattern was twice as fast as the rate of repetition of the three-tone pattern within the low-frequency range. The purpose of introducing different repetition rates was to elicit distinct peaks of ERP synchronization, which allowed disambiguation between the two streams. If no segregation occurred, the within-stream three-tone patterns did not emerge, and the unattended background was heard as an undifferentiated sequence of low- and middle-frequency tones alternating with a galloping rhythm (MLM_MLM etc.).

The two tones in the high-frequency range were presented pseudo randomly, with H_1 occurring 79 % (*standard*) and H_2 21 % (*deviant*) of the time. Occasionally, H_2 occurred twice in succession (7 %, *target*). Pattern reversal deviants in the low- and middle-frequency ranges, and target H_2 in the high-frequency range, never occurred consecutively, either within or across different frequency ranges.

Stimuli were presented in two conditions. In the *3-streams condition*, stimuli from the high-, middle-, and low-frequency ranges were presented as described in the previous paragraph and illustrated in Fig. 1. In the *2-streams condition*, stimuli from the middle-frequency range were omitted, leading to identical timing and structure for the low- and high-frequency tones as was in the 3-streams condition. In each of the conditions, 15 separately randomized sequences were presented, separated by short breaks. Each stimulus block was 108 s in duration, or 27 min duration per condition. The 3-streams condition presented 18,000 tones (4,500 L, 9,000 M, 4,500 H), with 1,200 tones (300 L, 600 M, 300 H) in each of the 15 runs. The 2-streams condition consisted of 9,000 tones (4,500 L, 4,500 H) overall, presented in 15 runs of 600 tones (300 L, and 300 H). In total, each condition had 105 deviants in the low-frequency range, and 105 targets in the high-frequency range. The 3-streams condition had 210 deviants in the middle-frequency range.

Procedure

Subjects were comfortably seated in an electrically-shielded and sound-attenuated booth (IAC, Bronx, NY). Their task was to attend to the high-frequency range (ignoring the other sounds), and to indicate through a button press on a keypad whenever they heard a target: the occurrence of two consecutive within-stream higher-pitched tones. Prior to recording, a practice session was administered in which subjects were first presented with the high-frequency tones by themselves at the rhythm they occur when mixed in the sequence, and asked to detect the two deviants-in-a-row targets. Once the task was clear, and the high-frequency targets in the absence of distracting sounds were correctly identified by button press, subjects were asked to perform the same task on samples of the full sequence (2- or 3-streams, depending on the counterbalanced order to which they had been assigned). Practice lasted until subjects could perform the task accurately (on average, this required three 2-min blocks of sounds). The practice sequences were not used in the main experiment. Total session time, including cap placement and breaks, was approximately 2 hours.

Data Recording

Electroencephalogram (EEG) was acquired using a 32-channel electrode cap in the 10–20 international system. Data were recorded from electrodes FPZ, FZ, CZ, PZ, OZ, FP1, FP2, F7, F8, F3, F4, FC5, FC6, FC1, FC2, T7, T8, C3, C4, CP5, CP6, CP1, CP2, P7, P8, P3, P4, O1, O2, LM, and RM. Vertical eye movements were recorded through an electrooculogram (VEOG), measured from FP1 and an electrode placed below the left eye. Horizontal eye movements were recorded with F7 and F8 electrodes (HEOG). The reference electrode was placed on the tip of the nose. For all electrodes, impedance was kept below 5 Ω. The EEG signal was amplified with a Neuroscan Synamps amplifier (Compumedics Corp., Texas, USA) at a gain of 1000, and digitized with a sampling rate of 500 Hz (bandpass 0.05–100 Hz), on a dedicated PC computer using Neuroscan Scan 4.1 software.

Data Analysis

EEG data were analyzed off-line using Neuroscan Edit 4.5 software. To observe and measure the MMN, data were low-pass filtered at 15 Hz, grouped into epochs of 900 ms (from 100 ms pre-stimulus onset to 800 ms post-stimulus onset. This size epoch allowed observation of the ERPs components evoked by consecutive deviants within the high-tone frequencies that occurred once every 360 ms). Electrical activity that exceeded ±75 μV after baseline-correction on the entire EEG epoch was rejected. This resulted in approximately 22 % rejected epochs. An eye-movement correction algorithm was applied to the data from three of our subjects due to excessive blinking. For each subject, epochs retained for further analysis were sorted by stimulus type (standard and deviant), frequency range (low, middle or high), and condition (3- and 2-streams). The *standard* stimulus was designated in the low- and middle-frequency ranges as the first tone of the ascending three-tone pattern (e.g., L_1 for the low-frequency range, M_1 for the middle-frequency range), and the *deviant* as the first tone of the descending three-tone pattern (e.g., L_3 for the low-frequency range, M_3 for the middle-frequency range). For the high-frequency range, the *standard* evoked response was the frequently occurring high tone (H_1). The epoch for the *target* was calculated from the first of the two consecutive higher-pitched (H_2) tones. Grand-mean waveforms were created

by averaging all subjects' data for each stimulus type, each frequency-range, and each condition separately.

The mean MMN amplitude was measured for each condition using a 50 ms window centered on the MMN peak latency obtained in the grand-mean deviant-minus-standard difference waveforms (Table 1). For the attended stream, MMN amplitudes were determined from the mastoid (RM), known to reliably show an MMN-related positive peak (inversion). For the unattended streams, MMN amplitudes were determined from the Fz electrode site. When no MMN peak was observed in the unattended frequency ranges, the same window as that found in the attended frequency range of the same condition was applied. The intervals used for the statistical analyses are reported in Table 1. One-sample Student's *t*-tests were calculated to determine the presence of the components, that is, to determine whether the mean amplitude of the waveform in the interval measured was significantly greater than zero. Because the prediction about the (negative) direction of the MMN was made a priori, one-tailed values were reported from the *t* test. Two-tailed paired *t*-tests were used to compare mean amplitudes of the MMNs. Repeated-measures ANOVAs were conducted in order to test main-effects and interactions for factors of condition (3- or 2-streams), frequency range (low, middle, or high), and electrode site (Fz, Cz, or Pz). Greenhouse-Geisser corrections were reported, as appropriate.

For the frequency spectra analysis, fast Fourier transforms (FFT) were computed in MATLAB (The Math-Works, Inc., Natick, Massachusetts, United States). After artifact rejection (±75 μV), data were spline fit to 256 points and averaged to the first tone of the 6-tone cycle for the 2-streams condition and the first tone of the 12-tone cycle for the 3-streams condition (L1–L1 in both conditions). Epochs were 2,000 ms, which included −1,000 ms prestimulus to 1,000 ms poststimulus onset, to capture a full tone cycle in both conditions. Tone cycles with deviants were excluded. Only the standard cycles were averaged together for each individual, in each condition separately, and analyzed for spectral content. Spectra were computed for electrode Cz, where maximal amplitudes for auditory evoked responses are to be expected (the ERP generators in the primary auditory cortex project towards the vertex of the skull) (Näätänen and Picton 1987). One second of ERP data (500 samples) was padded with zeroes corresponding to two seconds (1000 zeroes) in order to achieve better resolution on the frequency axis (1,500 samples, 0.33 Hz frequency resolution). The mean of each ERP was subtracted to reduce the DC peak at 0 Hz. The absolute values of the complex numbers returned by the FFT function were plotted representing amplitude spectra of the ERPs for conditions 2- and 3-streams. We expected to find a spectral peak corresponding to the overall frequency of stimulus presentation: a peak at 11.1 Hz (90 ms onset-to-onset rate) in the 3-streams condition and a peak at 5.6 Hz (180 ms onset-to-onset rate) in the 2-streams condition (Fig 1, left column). Indication of neurophysiological segregation of sounds by frequency stream would be found in the 3-streams condition with spectral peaks at 0.9 and 1.8 Hz, the onset-to-onset pace of the triads (Fig 1, right column). No physiological segregation of background sounds would be indicated by peaks occurring only at 2.8 and 5.6 Hz in the 3-streams condition. In the 2-streams condition, there is only one possible unattended frequency stream when the high tones are attended to perform a task. Physiological segregation would thus be indicated by

the spectral peak at 0.9 Hz (onset-to-onset pace of the triad). In order to determine whether two spectral values differ between conditions, we could not use regular statistics such as t-tests. The spectrum of the averaged ERP was computed as follows: 'spec = abs{fft[mean(ERPs)]}'—this is analogous to computing evoked activity when averaging across trials.[1] If one were to compute spectra for single subjects, one would have to change the formula to 'spec = mean{fft[mean(ERPs)]}'—which is analogous to total activity if averaging across trials. Evoked and total spectra deviate from each other even when averaging across subjects, since computing absolute values is a non-linear function. Therefore, we employed a non-parametric bootstrap procedure to estimate the variability of the spectra. The bootstrap method is a Monte Carlo technique that generates simulated data sets by resampling from empirical data observed in the original experiment (Efron and Tibshirani 1986). We randomly drew a surrogate sample of 12 subjects with replacement out of the existing 12 subjects such that in a surrogate sample each subject could be present multiple times while others would not be present at all. Evoked spectra were then computed for these surrogate samples. We repeated this procedure 1,000 times resulting in a distribution of power spectra. The 68 % confidence intervals (CI) were then computed for the spectral values of each condition by the bootstrap percentile method. This CI corresponds to one standard deviation of the mean in a Gaussian distribution. If the spectral value of one condition lies outside the spectral value of the other condition plus/minus the CI, it can be considered a significant difference.

Behavioral Data

Subjects' target responses were calculated based on hit rate (HR), false alarm rate (FAR), and reaction time for correct responses (RT), measured separately for each subject, and separately for each condition. Responses were considered correct when they occurred within 1,300 ms from the onset of the second of the two higher-pitched tones. HR, FAR, and RT for the two conditions were compared using paired, two-tailed *t*-tests.

Results

Behavioral Results

Participants responded significantly more accurately to the targets in the 2-streams condition (HR = 0.90; SD = 7) compared to the 3-streams condition (HR = 0.83; SD = 7) ($t_{11} = -6.12$, $p<0.001$). Mean FAR did not differ between the 2-streams (0.04; SD = 4) and 3-streams (0.06; SD = 7) conditions ($t_{11} = 1.34$, $p = 0.2$). Mean RT was shorter to targets in the 2-streams condition (395 ms; SD = 78) compared to those in the 3-streams condition (421 ms; SD = 88), although the difference did not reach significance ($t_{11} = 1.94$, $p = 0.078$). The higher HR in the 2-streams conditions suggests an influence of the more complicated background sounds in the 3-streams condition on target performance.

ERP Results

We found that in the attended, high-frequency range, MMN was elicited by each of the two consecutive higher-pitched tone deviants in both the 3- and 2-streams conditions (Fig. 2 top

[1]ABS refers to computing absolute values, FFT to computing a fast Fourier transform, and MEAN to the average of the ERPs.

row; Table 1). Two distinct negative deflections, separated by approximately 360 ms (the corresponding onset-to-onset latency of high-frequency range tones) can be seen in the displayed epoch (Fig. 2 top row).

Statistically significant MMNs were elicited in the high-frequency range (Table 1). There was no difference between the MMN amplitudes elicited by the successive high tone (first and second) deviants (i.e., the targets) within the attended, high-frequency range in the 2- and 3-streams conditions (2-streams condition: $t_{11} = 0.265$, $p = 0.795$; 3-streams condition: $t_{11} = 1.099$, $p = 0.294$, two-tailed). As expected, attention-related ERP components, N2b and P3b, were also elicited by target deviants in the attended stream ($t_{11} = 2.767$, $p = 0.009$, and $t_{11} = 6.128$, $p < 0.00005$, respectively, in the 3-streams condition; $t_{11} = 3.158$, $p = 0.004$, and $t_{11} = 6.593$, $p < 0.00005$, respectively, in the 2-streams condition; Fig. 3 top row). N2b and P3b components were not observed in the to-be-ignored, unattended middle- and low-frequency ranges, in both the 3- and the 2-streams conditions (Fig. 3 bottom row), as expected when subjects effectively ignore the sounds in accordance with the instructions.

Unattended low frequency deviant sounds elicited significant MMNs in the 2-streams condition (M = −0.975 μV, SD = 1.104 μV, $t_{11} = 3.061$, $p = 0.005$). In contrast, no MMN was elicited by the unattended low frequency deviant sounds in the 3-streams condition (M = 0.063 μV, SD = 0.989 μV, $t_{11} = 0.222$, $p = 0.413$) (Fig. 2 bottom row; Table 1). (The Supplementary Figure displays the ERP responses to the deviant sounds of the middle-frequency tones in the 3-stream condition, which did not reach significance. The clear deflections are overlapping contributions from the high stream tones). The absence of MMN in the unattended low-frequency stream of the 3-streams condition indicates that deviance detection has *not* occurred. This is key to understanding the significance of this study's results, as if stream segregation and deviance detection were interrelated, one would also expect no evidence of sound organization into distinct streams.

Fourier analysis revealed that the spectra of conditions 2- and 3-streams do not differ significantly at 0.9 and 5.6 Hz (Fig. 4). These two frequencies appear in the possible perceptions in both conditions (cf. Fig. 1). As expected, at 2.8 Hz, the spectrum of the 2-streams condition is significantly stronger (1.67 SD) than that of the 3-streams condition. The 11.1 Hz peak is also expected for the 3-streams condition, because it is the overall stimulus repetition rate. The difference at 11.1 Hz (1.43 SD above the 2-streams condition) is in line with the fact that only the 3-streams condition used the fast repetition rate of 90 ms (corresponding to 11.1 Hz). The key finding demonstrating neurophysiological segregation of the background sounds in the 3-streams condition is the significant peak at 1.8 Hz.

This peak reflects the triad onset-to-onset pace of the middle frequency stream, which is not contained in the 2-streams condition. An undifferentiated background for the 3-streams condition would be shown if only the 2.8 Hz (low frequency) and 5.6 Hz (middle frequency) peaks were evident. The difference at 1.8 Hz (1.69 SD above the 2-streams condition) corresponds to the segregated triad in the 3-streams condition. Thus, the presence of, and the significantly larger peak at 1.8 Hz in the 3-streams condition, compared to the 2-streams condition, is evidence for regularity detection of the background sounds (cf. Fig. 1).

Discussion

Our results provide neurophysiologic evidence that regularity detection and change detection can be empirically dissociated, and that the brain extracts regularity even under conditions that affect its ability to detect change. The presence of the MMN and N2b components in the attended frequency range in both 2- and 3-streams conditions indicates that change detection occurred in the attended sounds independently of the complexity of the unattended background. The higher HR and faster RT to targets in the 2-streams condition compared to the 3-streams condition suggest that the complexity of the background influenced target performance. Moreover, the absence of the MMN component in the unattended frequency range of the 3-streams condition indicates that change detection for the unattended sounds did *not* occur. Change detection for unattended sounds occurred only in the 2-streams condition, in which there was only one possible stream after attention segregated out the high frequency sounds to perform the task. In other words, selectively attending the high tones resulted in segregation of the high from the low tones. Hence, for the unattended background sounds, change detection occurred only when the unattended background consisted of a single stream (the 2-streams condition), but not when the background included tones belonging to multiple frequency ranges and multiple regularities (the 3-streams condition, a simplified version of the cocktail party scenario).

Using change detection as a proxy for stream segregation, these MMN results suggest that the unattended auditory background remains unstructured (as was also found in Sussman et al. 2005). However, the results of our frequency spectra analysis challenge this interpretation by showing that the unattended auditory background can be structurally organized: the 0.9 Hz peak representing the triad repetition in the 2-streams condition, and the 1.8 Hz peak representing the triad repetition for the middle frequency stream in addition to the 0.9 Hz representing the triad repetition for the low tones of the 3-streams condition. These results demonstrate that the unattended sounds were neurophysiologically segregated into distinct streams even when within-stream changes were not detected. That is, the match between the ERP synchronization and the within-stream repetition rates indicates that stream segregation occurred within the unattended background sounds independently of the direction of attention, and of the change detection process. These results are thus consistent with previous results showing that stream segregation occurs prior to the change detection process indexed by MMN (Sussman 2005; Yabe et al. 2001), and further highlights a distinction between these processes.

The dissociation found in our data between stream segregation and change detection emphasizes the limited value of the MMN (an index of change detection) as an indicator of the organization of sound into distinct streams. Certainly, the presence of MMN provides a firm indicator that the detected regularity may include stream segregation, but the absence of MMN does not mean that organization of sounds to streams did not occur. As best is known, the MMN arises primarily from neural generators in the superior temporal plane (within auditory cortices), and from feedback from higher auditory and multimodal brain regions (Winer et al. 2005). However, the stream segregation process has more wide-reaching contributions, beginning with sounds reaching the ear that are decomposed based on their frequency components, and with the tonotopic representation continuing in the cochlear

nucleus and other subcortical areas along the afferent pathway, prior to reaching the auditory cortex (Moore 2008). This information is integrated across time to give rise to auditory objects (Nelken 2004; Nelken et al. 1999, 2003). Evidence suggesting a primitive nature of the stream segregation process is found in animal studies showing that aspects of streaming occur under general anaesthesia (Pressnitzer et al. 2008), and in developmental studies on human infants showing that basic stream segregation mechanisms are present in infancy (Demany 1982; McAdams and Bertoncini 1997). Hence, the auditory system is evolutionarily and developmentally well adapted to perceiving the auditory environments as spectrally and temporally structured at very early processing stages, which subserves the ability to navigate the noisy auditory scene with great facility. This organizational principle for negotiating the complex auditory environment is linked to but not fully interdependent with the processes that lead to change detection, according to our current results. In agreement with anatomical and physiological evidence of early perceptual organization, our results indicate that the auditory system extracts the structure of the auditory environment for attended and unattended inputs, and this occurs prior to the change detection process that generates the MMN.

The modulatory effect of attention on the neural responses has been well documented, and attention is known to interact closely with the physical attributes of the stimulus (Posner and Petersen 1990; Fritz et al. 2007). According to “early selection” theories, attention operates through a “gain control” mechanism, involving the enhancement of relevant and suppression of irrelevant inputs (Hillyard et al. 1973). These processes may begin as early as 20 ms post-stimulus (Woldorff et al. 1993). Conversely, “late selection” theories propose that attention modulates neural responses through a distinct set of cortical neurons that process the task-relevant aspects of sounds without affecting the early stages of sound representation (Alho 1992; Nätänen 1992). This functional dichotomy between attention-dependent neurons (which analyze acoustic features of behaviorally relevant sounds) and stimulus-dependent neurons (which transmit acoustic information) has also received empirical support (e.g., Jäncke et al. 2003; Petkov et al. 2004). Complementing “early-” and “late-selection” theories, the recently developed “tuning model” proposes that attention may act directly on the auditory neurons through a mechanism of short-term plasticity that enhances the auditory neurons' selectivity to task-relevant features (Fritz et al. 2003; Jääskeläinen et al. 2007). Empirical support for short-term plasticity and enhancement of task-relevant features comes from studies of entrainment to rhythmic stimuli. The neural responses to multiple rhythms have been studied in the auditory steady state response (e.g., John et al. 2001; Draganova et al. 2002), as well as in the auditory attention literature. Attending to rhythmic auditory streams has been shown to result in the entrainment of ongoing oscillatory activity in the primary auditory cortex, possibly enhancing the representation of the attended stimuli (Lakatos et al. 2013). The attention-related enhancement of the neural responses to rhythmic stimuli has also been shown to reflect the perceptual detectability of the rhythm, and to correlate with behavioral performance (Elhilali et al. 2009). This close match between psychophysical and behavioral measures provides compelling empirical support to the notion that the attention-related enhancement may act through mechanisms of task-induced neural plasticity (Fritz et al. 2005).

Our results are compatible with aspects of all three models. On the one hand, the presence of the MMN to targets in the attended stream occurred regardless of the complexity of the unattended background, and may reflect an enhanced neural representation for attended features of the input (sensory gain control and plasticity, as proposed in the “early selection” and “tuning” theories). This would be consistent with previous findings of attention-related modulation (Bidet-Caulet et al. 2007; Elhilali et al. 2009). On the other hand, the evidence of distinct peaks of neural synchronization corresponding to the distinct streams of the unattended background in the absence of MMN suggests that task-relevant and non-task relevant sensory information may undergo distinct processing (as predicted by the “late selection” theories), with attention affecting only the degree of processing for the unattended inputs. The change detection process may require some degree of attention in complicated situations, to identify sound events, or change events for unattended background sounds, which might in part explain why people have difficulty following a speech stream in a noisy environment.

Our data also suggest that the attention-related enhancement is attenuated in the 3-streams condition (lower 2.8 Hz peak compared to the 2-streams condition). This effect is likely due to the presence of an extra set of tones using cognitive resources available for auditory processing, consistent with previous findings showing that the MMN response can be reduced (or even abolished) depending on the auditory cognitive load (e.g., when attention is selectively focused on a subset of auditory stimuli, as opposed to when it is focused on a visual task) (Sussman et al. 2005). Our results therefore appear compatible with the feature-based modulation model, and suggest a tight interaction between the top-down effect of attention, the bottom-up physical properties of the stimulus, and the behavioral task demands.

Overall, our results are consistent with models of sound organization involving at least two distinct stages: an early frequency-based segregation step, entirely stimulus driven; and a late step of integration of sounds over time, partly under top-down control. By supporting this two-stage model, our results reconcile two seemingly incompatible views on the role of attention on auditory processing: the first holding that attention is required for the change detection process; the second arguing that attention is not necessary for the MMN to occur, although it may affect sound organization at processing stages prior to MMN generation (Sussman et al. 1998, 2002), and depending on context (Sussman and Winkler 2001; Sussman and Steinschneider 2006). Our results suggest that different studies may have produced discrepant evidence because they may have tapped into different stages of sound organization, and are compatible with previous literature proposing that frequency-based stream segregation precedes temporal integration (Yabe et al. 2001; Sussman 2005; Snyder et al. 2006). Furthermore, by showing that unattended tones were grouped in triads, not merely segregated based on tone frequency, our data indicate that both phases—frequency-based segregation and integration of sounds over time— have occurred in the absence of MMN, supporting the hypothesis that both segregation and integration occur prior to MMN generation. By showing that the structure of the auditory environment is extracted prior to the generation of the MMN, our results indicate that the auditory system prioritizes information about regularity over information about change, and are consistent with predictive models of regularity as providing initial hypotheses about the structure of the

external world (Gregory 1980), which are continually tested against (and adjusted based on) the current sensory input (Winkler et al. 2009).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the “Art & Neuroscience” fellowship from the Italian Academy for Advanced Studies in America (A.P.), by the “Columbia Science” fellowship from Columbia University (A.P.), by the National Institutes of Health Grant # R01 DC004263 (E.S.), and by the German Research Foundation Grant SFB/TRR 31 (C.S.H.). A.P. is currently supported by a Marie Curie Actions - BRIDGE Fellowship from the European Union Seventh Framework Programme, at the University of Geneva.

References

Alho K. Selective attention in auditory processing as reflected by event-related brain potentials. Psychophysiology. 1992; 29:247–263. [PubMed: 1626035]

Bidet-Caulet A, Fischer C, Besle J, Aguera PE, Giard MH, Bertrand O. Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. J Neurosci. 2007; 27(35):9252–9261. [PubMed: 17728439]

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

Bregman, A. Auditory scene analysis: the perceptual organization of sound. The MIT Press; Cambridge: 1990.

Bregman AS, Campbell J. Primary auditory stream segregation and perception of order in rapid sequences of tones. J Exp Psychol. 1971; 89:244–249. [PubMed: 5567132]

Carlyon RP, Cusack R, Foxton JM, Robertson IH. Effects of attention and unilateral neglect on auditory stream segregation. J Exp Psychol Hum Percept Perform. 2001; 27:115–127. [PubMed: 11248927]

Cherry EC. Some experiments on the recognition of speech, with one and two ears. J Acoust Soc Am. 1953; 25:975–979.

Demany L. Auditory stream segregation in infancy. Infant Behav Dev. 1982; 5:261–276.

Draganova R, Ross B, Borgmann C, Pantev C. Auditory cortical response patterns to multiple rhythms of AM sound. Ear Hear. 2002; 23:254–265. [PubMed: 12072617]

Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci. 1986; 1:54–75.

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(6):e1000129. [PubMed: 19529760]

Fritz J, Shamma S, Elhilali M, Klein D. Rapid task-related plasticity of spectro-temporal receptive fields in primary auditory cortex. Nat Neurosci. 2003; 6:1216–1223. [PubMed: 14583754]

Fritz JB, Elhilali M, Shamma SA. Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. J Neurosci. 2005; 25:7623–7635. [PubMed: 16107649]

Fritz JB, Elhilali M, David SV, Shamma SA. Auditory attention—focusing the searchlight on sound. Curr Opin Neurobiol. 2007; 17:437–455. [PubMed: 17714933]

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

Hillyard SA, Hink RF, Schwent VL, Picton TW. Electrical signs of selective attention in the human brain. Science. 1973; 182:177–180. [PubMed: 4730062]

Jääskeläinen IP, Ahveninen J, Belliveau JW, Raij T, Sams M. Short-term plasticity in auditory cognition. Trends Neurosci. 2007; 30:653–661. [PubMed: 17981345]

Jäncke L, Specht K, Shah JN, Hugdahl K. Focused attention in a simple dichotic listening task: an fMRI experiment. Brain Res Cogn Brain Res. 2003; 16:257–266. [PubMed: 12668235]

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]

Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE. The spectrotemporal filter mechanism of auditory selective attention. Neuron. 2013; 77:750–761. [PubMed: 23439126]

McAdams S, Bertoncini J. Organization and discrimination of repeating sound sequences by newborn infants. J Acoust Soc Am. 1997; 102:2945–2953. [PubMed: 9373981]

Moore, BCJ. An introduction to the psychology of hearing. 5th. Emerald Group Publishing; Bingley: 2008.

Näätänen, R. Attention and brain function. Lawrence Erlbaum; Hillsdale: 1992.

Näätänen R, Gaillard AWK, Mäntysalo S. Early selective-attention effect on evoked potential reinterpreted. Acta Psychol (Amst). 1978; 42:313–329. [PubMed: 685709]

Näätänen R, 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. Processing of complex stimuli and natural scenes in the auditory cortex. Curr Opin Neurobiol. 2004; 14:474–480. [PubMed: 15321068]

Nelken I, Rotman Y, Bar-Yosef O. Response of auditorycortex neurons to structural features of natural sounds. Nature. 1999; 397:154–157. [PubMed: 9923676]

Nelken I, Fishbach A, Las L, Ulanovsky N, Farkas D. Primary auditory cortex of cats: feature detection or something else? Biol Cybern. 2003; 89:397–406. [PubMed: 14669020]

Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL. Attentional modulation of human auditory cortex. Nat Neurosci. 2004; 7:658–663. [PubMed: 15156150]

Posner MI, Petersen SE. The attention system of the human brain. Annu Rev Neurosci. 1990; 13:25–42. [PubMed: 2183676]

Pressnitzer D, Sayles M, Micheyl C, Winter IM. Perceptual organization of sound begins in the periphery. Curr Biol. 2008; 18:1124–1128. [PubMed: 18656355]

Snyder JS, Alain C, Picton TW. Effects of attention on neuroelectric correlates of auditory stream segregation. J Cogn Neurosci. 2006; 18:1–13. [PubMed: 16417678]

Sussman E. Integration and segregation in auditory scene analysis. J Acoust Soc Am. 2005; 117:1285–1298. [PubMed: 15807017]

Sussman E. A new view on the MMN and attention debate: Auditory context effects. J Psychophysiol. 2007; 21(3–4):164–175.

Sussman E, Steinschneider M. Neurophysiological evidence for context-dependent encoding of sensory input in human auditory cortex. Brain Res. 2006; 1075(1):165–74. [PubMed: 16460703]

Sussman E, Winkler I. Dynamic sensory updating in the auditory system. Brain Res Cogn Brain Res. 2001; 12:431–439. [PubMed: 11689303]

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

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, Winkler I, Huotilainen M, Ritter W, Näätänen R. Top-down effect on the initially stimulus-driven auditory organization. Brain Res Cogn Brain Res. 2002; 13:393–405. [PubMed: 11919003]

Sussman ES, Bregman AS, Wang WJ, Khan FJ. Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. Cogn Affect Behav Neurosci. 2005; 5(1):93–110. [PubMed: 15913011]

Sussman ES, Horvath J, Winkler I, Orr M. The role of attention in the formation of auditory streams. Percept Psychophys. 2007; 69(1):136–152. [PubMed: 17515223]

Sussman E, Chen S, Sussman-Fort J, Dinces E. The five myths of MMN: Redefining how to use MMN in basic and clinical research. Brain Topogr. 201310.1007/s10548-013-0326-6

Winer JA, Miller LM, Lee CC, Schreiner CE. Auditory thalamocortical transformation: structure and function. Trends Neurosci. 2005; 28:255–263. [PubMed: 15866200]

Winkler I, Sussman E, Tervaniemi M, Ritter W, Horvath J, Näätänen R. Pre-attentive auditory context effect. Cogn Affect Behav Neurosci. 2003; 3:57–77. [PubMed: 12822599]

Winkler I, Takegata R, Sussman E. Event-related brain potentials reveal multiple stages in the perceptual organization of sound. Brain Res Cogn Brain Res. 2005; 25:291–299. [PubMed: 16005616]

Winkler I, Denham SL, Nelken I. Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn Sci. 2009; 13(12):532–540. [PubMed: 19828357]

Woldorff MG, Gallen CC, Hampson SA, Hillyard SA, Pantev C, Sobel D, Bloom FE. Modulation of early sensory processing in human auditory cortex during auditory selective attention. Proc Natl Acad Sci USA. 1993; 90:8722–8726. [PubMed: 8378354]

Yabe H, Winkler I, Czigler I, Koyama S, Kakigi R, Sutoh T, Hiruma T, Kaneko S. Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration. Brain Res. 2001; 897:222–227. [PubMed: 11282382]

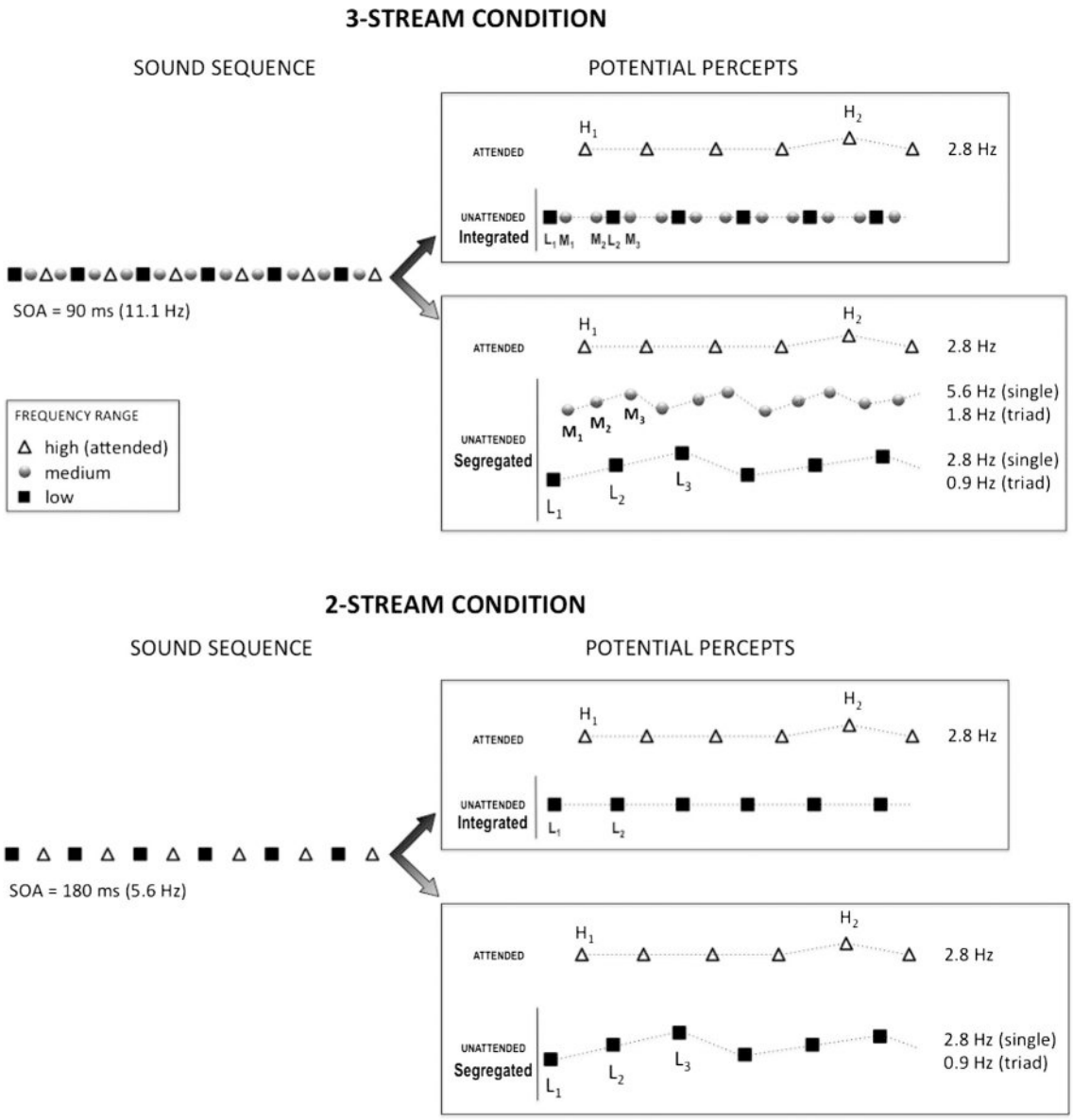

Fig. 1.

Experimental design. Schematic diagram of the stimulus paradigm showing a sample of sound sequences (*left*) from the 3-streams (*top panel*) and 2-streams (*bottom panel*) conditions, along with the corresponding organizations of the sounds (i.e., potential percepts, right). High- (*triangles*), middle- (*circles*, only in the 3-streams condition), and low-frequency range (*squares*) tones were presented in rapid alternation. Within each frequency range, rare deviant tones violated the standard pattern. In the high frequency range, this violation consisted of two successive within-stream higher-pitched (H_2) tones. In the middle- and low-frequency ranges, the violation was a reversal of the ascending triads (e.g., L_3–L_2–L_1). Each frequency range was characterized by a distinct stimulus rate of single tones and triads. A high or low frequency tone occurred once every 360 ms (corresponding to a rate of 2.8 Hz). Middle-frequency tones occurred once every 180 ms

(corresponding to 5.6 Hz). Low frequency triads were repeated once every 1080 ms (corresponding to 0.9 Hz), and middle frequency triads once every 540 ms (corresponding to 1.8 Hz)

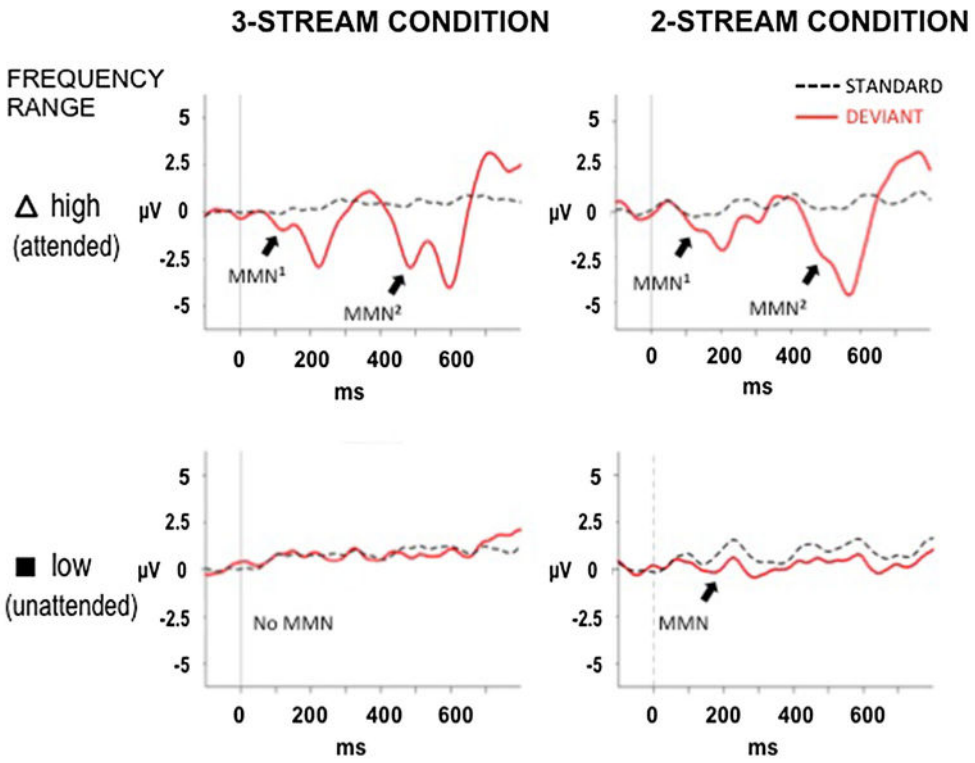

Fig. 2.

Event-related potentials (ERPs). Grand-mean ERPs elicited by the standard (*dashed lines*) and deviant (*thick lines*) tones are displayed for the 3-streams (*left column*) and 2-streams (*right column*) conditions at the Fz electrode, the site of best signal-to-noise ratio (SNR). The ERPs elicited while selectively attending the high tones (*top row*, indicated with a *triangle*, to match with Fig. 1) and ERPs elicited by the unattended low tones (*bottom row*, denoted with a *black square*, to match with Fig. 1) are displayed separately

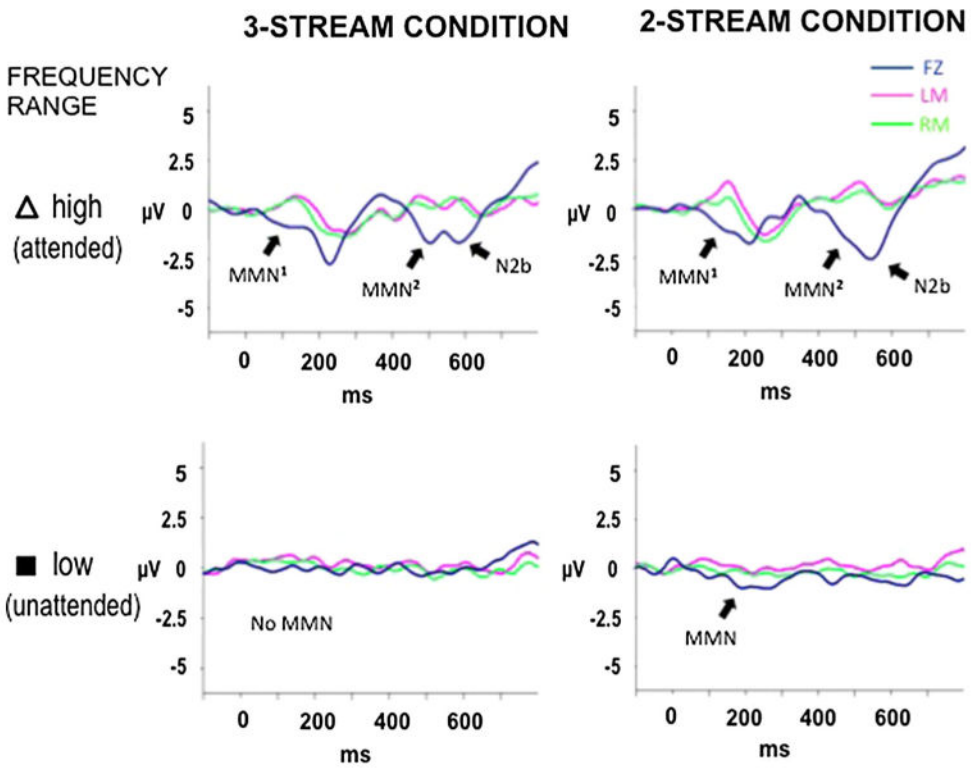

Fig. 3.

Difference waveforms. The grand-mean difference waveforms were derived by subtracting the ERP response to the standard tones from the ERP response to the deviant tones, to delineate the MMN component. Significant MMNs are labeled with arrows. Two consecutive MMNs are observed in both the 3-streams (*left column*) and 2-streams (*right column*) conditions (*top row*, attended high tones). These were the targets (two successive high deviants). The second MMN (elicited by the second of two consecutive deviants) is followed by the N2b component, denoting detection of the target prior to the button press. N2b is an endogenous component that occurs in response to attended target sounds. No N2b was observed to the unattended deviants (*bottom row*, unattended low tones). MMN was significantly elicited by unattended deviants only in the 2-streams condition, when there were no competing unattended sounds (*bottom, right panel*)

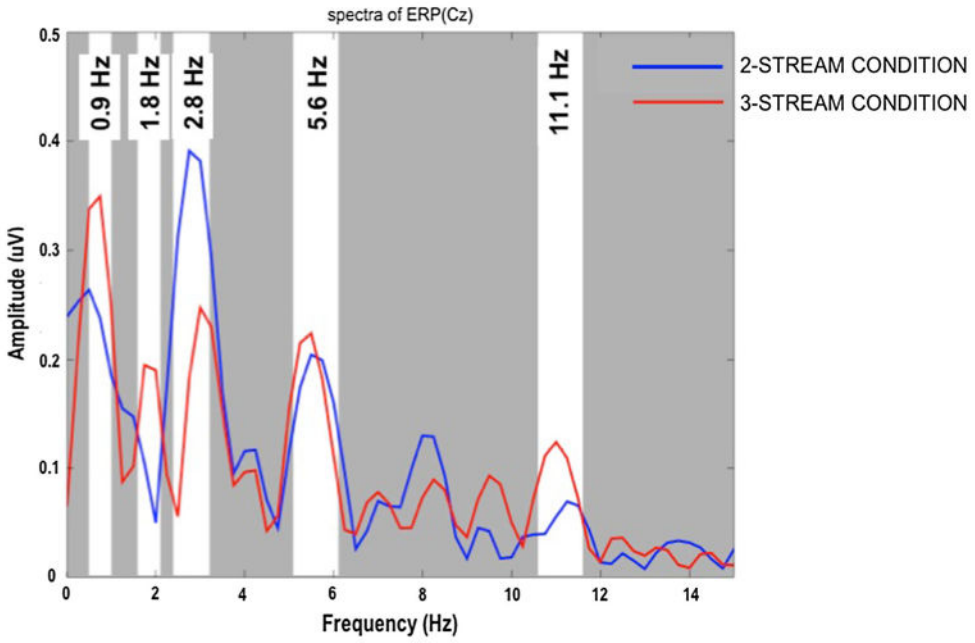

Fig. 4.

Frequency spectra of ERP. Frequency ranges and amplitudes of ERP synchronization at the Cz electrode, obtained using Fast Fourier Transform analysis on the standards for both 2- and 3-streams conditions separately. Peaks of synchronization are visible at 0.9 Hz (corresponding to the frequency of the low-frequency three-tone pattern—once every 1,080 ms), 2.8 Hz (corresponding to the frequency of individual tones within either the high- or the low-frequency range—once every 360 ms), and 5.6 Hz (corresponding to the frequency of individual tones within the middle-frequency range of the 3-streams condition, as well as to the global rate of occurrence of tones, irrespectively of the range, in the 2-streams condition —once every 180 ms in both cases). In the 3-streams condition, additional peaks are visible at 1.8 Hz (corresponding to the rate of repetition of the middle-frequency range triad—once every 540 ms), and 11.1 Hz (corresponding to the global rate of occurrence of tones, irrespectively of the range—once every 90 ms)

Brain Topogr. Author manuscript; available in PMC 2015 May 01.

Table 1

MMN results

	Latency (ms)	Difference (Deviant – Standard)		*t* Stat	*p* (one-tailed)
		M	***SD***		
3-streams condition					
High					
MMN[1]	107–157	−0.98	0.99	3.43	**0.003**
MMN[2]	469–519	−3.11	3.82	2.81	**0.009**
Middle	107–157	−0.41	0.81	1.74	0.06
Low	171–221	0.06	0.99	0.22	0.41
2-streams condition					
High					
MMN[1]	113–163	−0.88	1.32	2.32	**0.020**
MMN[2]	471–521	−2.46	3.33	2.57	**0.013**
Low	171–221	−0.98	1.10	3.06	**0.005**

Mean amplitudes (in μV) and standard deviations recorded in the MMN latency windows for each frequency range (high, middle, and low) and each condition (3- and 2-streams) Statistically significant differences are emphasized in bold