# Neural correlates of attention and streaming in a perceptually multistable auditory illusion

Anahita H. Mehta

UCL Ear Institute, University College London, 332, Grays Inn road, London WC1X 8EE, UK

Ifat Yasin

Department of Computer Science, University College London, 66-72 Gower Street, London WC1E 6BT, UK

Andrew J. Oxenham

Department of Psychology, University of Minnesota, 75 East River Parkway, Minneapolis, Minnesota 55455, USA

Shihab Shamma

Institute for Systems Research, 2203 A.V. Williams Building, University of Maryland, College Park, Maryland 20742, USA

Abbreviated title: Neural correlates of attention and streaming

---

[1] Author Shihab Shamma also at École Normale Supérieure, 29 Rue d'Ulm, Paris 75005.

[2] Author Anahita H. Mehta is currently at Department of Psychology, University of Minnesota, Minneapolis, USA, MN 55455. Email: mehta@umn.edu

In a complex acoustic environment, acoustic cues and attention interact in the formation of streams within the auditory scene. In this study, a variant of the ‘octave illusion’ [Deutsch, D. (1974). “An auditory illusion,” Nature, 251, 307–309] was used to investigate the neural correlates of auditory streaming, and to elucidate the effects of attention on the interaction between sequential and concurrent sound segregation in humans. By directing subjects’ attention to different frequencies and ears, it was possible to elicit several different illusory percepts with the identical stimulus. The first experiment tested the hypothesis that the illusion depends on the ability of listeners to perceptually stream the target tones from within the alternating sound sequences. In the second experiment, concurrent psychophysical measures and EEG recordings provided neural correlates of the various percepts elicited by the multistable stimulus. The results show that the perception and neural correlates of the auditory illusion can be manipulated robustly by attentional focus and that the illusion is constrained in much the same way as auditory stream segregation, suggesting common underlying mechanisms.

PACS numbers: 43.25.Nm, 43.66.Ba

# I. INTRODUCTION

Making sense of the acoustic environment is a complex but crucial task of the auditory system. The perceptual organization of complex acoustic environments into coherent auditory scenes is often discussed in terms of auditory grouping (the binding of simultaneous acoustic features into a single auditory “object”) and streaming (the organizing of successive sounds into a unitary percept) (e.g. Bregman, 1990; Darwin and Carlyon, 1995). The acoustical properties that influence auditory grouping and streaming have been explored using several behavioral as well as electrophysiological methods (Alain, 2007; Billig et al., 2013; Denham et al., 2010; Elhilali et al., 2009; Gutschalk et al., 2005; Micheyl et al., 2007, 2013a, 2013b; Pressnitzer and Hupé, 2006; Shamma and Micheyl, 2010). With few exceptions (e.g. Darwin et al., 1995; Shinn-Cunningham et al., 2007), most studies have investigated simultaneous grouping and sequential streaming separately, despite the fact that sound sources in the natural environment overlap as well as unfold over time.

In addition to acoustic cues, attention, expectation, and other “top-down” influences have been found to play a key role in how complex auditory scenes are processed and perceived (Carlyon et al., 2001; Cusack et al., 2004; Elhilali and Shamma, 2008; Moore and Gockel, 2012; Winkler et al., 2012). Perceptually ambiguous or multistable stimuli can be useful in elucidating the mechanisms of perception, and in dissociating the neural responses to physical stimuli from the neural correlates of perception (Leopold and Logothetis, 1999; Schwartz et al., 2012).

The experiments described here address the interaction between sequential and concurrent sound segregation, as well as the interactions between acoustic and attentional manipulations on the perception of sound sequences. The experiments exploit a stimulus paradigm similar to the one used to elicit Deutsch’s “octave illusion” (Deutsch, 1974). In this illusion, most listeners report hearing an alternating pattern of low and high tones, with all the low tones lateralized to one side and all the high tones lateralized to the other side, even though the actual stimulus has alternating low and high tones in both ears (Figure 1).

Deutsch's (1974) stimuli and the robustness of the elicited percepts have been investigated over a wide range of parameters. The illusion has been shown to be robust to changes in tone duration (Zwicker, 1984), intensity (Deutsch, 1978), frequency separation (Brancucci et al., 2009), and timbre (McClurkin and Hall, 1981) and can also be elicited by aperiodic stimuli such as band-pass noise (Brännström and Nilsson, 2011). It was noted by Deutsch and Roll (1976), and later confirmed by Brancucci et al. (2009), that the illusion is not dependent on the tones being in an exact octave relationship, although Brancucci et al. (2009) noted that the illusion became less compelling at musical intervals much less than an octave. Since its introduction, this stimulus has been shown to be multistable, in that it can be perceived in multiple ways (Brancucci et al., 2014; Brancucci and Tommasi, 2011; Chambers et al., 2002; Deutsch and Gregory, 1978). Deutsch and Roll (1976) suggested that most listeners report hearing the tone frequencies that were presented to their “dominant” ear (usually the right), through suppression of the non-dominant ear. Such suppression was postulated not to occur for sound localization, but instead the localization of the tone heard was thought to depend upon the physical location of the higher frequency tone at any given time interval, regardless of the ear of presentation. Subsequently, it was noted that not all participants perceived the illusion in the same fashion (the pattern may differ from high tones in right ear and low tones in left ear to the opposite pattern) and that the percept also depended on the length of stimulus presentation, such that the illusion did not occur for very short presentations of the stimulus (Christensen and Gregory, 1977; Deutsch and Gregory, 1978).

Bregman and Steiger (1980) had suggested that in the case of the classic octave illusion (see Fig. 1), the auditory system treated the 800-Hz tone as a harmonic of the 400-Hz tone and thus localized the percept of the tone at the ear receiving the “more reliable higher harmonic”. Chambers et al. (2002) suggested that the octave illusion percept was based on dichotic fusion, which meant that the percept was made of the tones from both the ears fusing to form a percept that varied very slightly in overall perceived frequency or pitch. Although the Chambers et al. (2002) study highlighted the aspect of bilateral grouping of tones, it has since been established that the tones perceived do not sound like a fused auditory image, and in fact correspond to the pitch of the component high- and low-frequency tones (Deutsch, 2004).

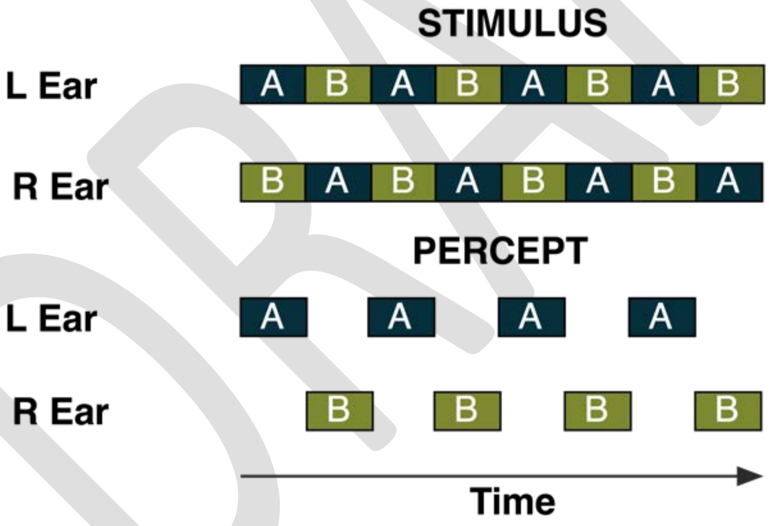

*FIGURE 1: (Color online) The stimulus pattern used in the original experiment of Deutsch (1974) describing the octave illusion, together with the percept most commonly obtained. Boxes labelled ‘A’ indicate tones at 400 Hz, and boxes labelled ‘B’ indicate tones at 800 Hz.*

Although the illusion has received considerable attention, it remains unclear whether it can be manipulated via directed attention or experimenter instructions. Such manipulation could be useful in exploring the neural bases of the illusion, as different percepts might be elicited with physically identical sounds. Some neuroimaging studies have been carried out in an effort to understand the neural bases of the illusion (Brancucci et al., 2009, 2011; Lamminmäki et al., 2012; Lamminmäki and Hari, 2000). However, in all these neuroimaging studies, subjects' spontaneous percepts were either tested beforehand, with the recordings obtained while the subjects were listening passively (Brancucci et al., 2012; Lamminmäki et al., 2012; Lamminmäki and Hari, 2000), or the response measures for a task-based study were mainly focused on the participants' *subjective* reports of their percept (Brancucci et al., 2014). None of these studies attempted to record the neural responses while simultaneously actively manipulating, or objectively measuring, the participants' percepts.

This study investigated sequences of simultaneous alternating high and low pure tones, as shown in Fig. 1, in the context of auditory streaming. Although the illusion has not been explicitly studied in this context before, several of its properties suggest that it may reflect the same underlying mechanisms as streaming. For instance, it has been reported that the illusion may not be as strong when the frequency separation between the two tones becomes too small (Brancucci et al., 2009), as is also observed in streaming studies for both frequency (e.g. van Noorden, 1975) and pitch (e.g. Vliegen and Oxenham, 1999). Furthermore, the combination of sequential organizing with simultaneously presented sounds provides an opportunity to study the interaction between simultaneous grouping and sequential streaming. The study described here also investigated whether the perception and neural correlates of the illusion can be manipulated via selective attention, based on priming and experimenter instructions. More specifically, the experiments described in this study test the following hypotheses: (1) that the illusion occurs within the parameter ranges that induce auditory streaming, (2) that priming listeners with auditory cues affects their perception of the octave illusion, and (3) that the corresponding neural activity obtained via EEG reflects the perception of the illusion, as manipulated by the priming cues.

# II. METHODS

## A. Participants

Fifteen participants (eleven female and four male, aged 20-29 years) took part in the first experiment, which involved only behavioral measures. Ten participants (four female and six male, aged 20-29 years) took part in the second experiment, which involved simultaneous behavioral and EEG measurements. All participants tested were naïve to the aims of the study, and there was no overlap of participants between the experiments. All participants had normal hearing, defined as audiometric hearing thresholds no more than 15 dB Hearing Level (HL) at octave frequencies from 250 Hz to 4 kHz, with no history of hearing or neurological disorders. Participants provided written informed consent and were compensated for their participation. Experiment 1 was carried out at University College London and experiment 2 was carried out at the University of Maryland. The University College London Ethics Committee and the University of Maryland Institutional Review Board approved the procedures for experiments 1 and 2, respectively.

## B. Experiment 1: Stimuli and procedure

Experiment 1 tested the first two of the hypotheses listed above, that (1) the stimulus parameters of the octave illusion correspond to those of stream segregation, and (2) priming cues affect listeners' perception of the illusion. In this experiment, alternating sequences of low and high tones were presented to each ear in opposite phase, such that the sequence in the left ear could be a High-Low-High… pattern while the sequence in the right ear could be a Low-High-Low… pattern (see Fig. 2). Participants were cued to attend to a particular ear (R or L) and frequency (termed Hi or Lo), as indicated by a priming sequence of three pure tones that were presented either to the left or right ear and were either all low or all high in frequency (i.e., all R/Lo, R/Hi, L/Lo, or L/Hi). All tones were 100 ms in duration, with 10-ms raised-cosine onset and offset ramps. Within the priming and the main sequence, the tones were separated from each other by a 50-ms silent period. The silent period between the priming sequence and the test sequence was 500 ms. The sequences were generated in MATLAB (MathWorks Inc. Natick, MA, USA) and were presented at a sampling rate of 44.1 kHz. The experiment was presented using the Psychophysics Toolbox extension in MATLAB (Brainard, 1997; Pelli, 1997) through Sennheiser HD 215 headphones (Old Lyme, CT, USA).

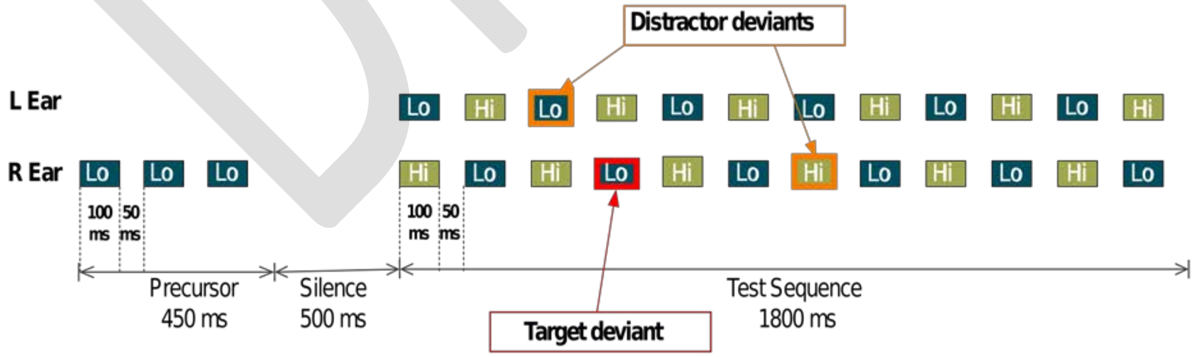

*FIGURE 2: (Color online) Schematic representation of the stimuli. Boxes labelled 'Lo' and 'Hi' indicate pure tones of low and high frequencies. Each ear receives an alternating sequence of Hi-Lo tones. The example trial shown in the figure has a precursor sequence of low frequency tones in the right ear indicating the attended stream. The amplitude deviant in the Right-Low tones thus becomes the target deviant among the other distractor deviants.*

Each ear of a listener was presented with alternating sequences of 12 pure tones per trial – six high and six low tones in each ear (see Fig. 2). Each of the four tone streams (R/Lo, R/Hi, L/Lo, and L/Hi) could have one deviant tone (amplitude increase by 7 dB on one of the tones) that occurred either early, mid, or late in the particular stream. The reason for having deviants was to use an objective measure of segregation (Micheyl and Oxenham, 2010; Thompson et al., 2011). Each stream had a randomized arrangement of the location of the targets and distractor deviants. The deviants did not occur simultaneously in more than one stream. It was ensured that an equal number of early, mid, and late deviants were presented across the test blocks. Depending on the priming sequence, the deviant in the primed stream was the target deviant, and the deviants in the other streams were termed distractor deviants. An example trial is shown in Figure 2, where the priming sequence is for the right ear and low tones (R/Lo), so the target is the deviant in the R/Lo stream and the distractors are deviants in any of the other streams (as indicated in Figure 2). The participants were required to detect the target deviant while ignoring all other distractor deviants. They responded via button press at the end of each trial to indicate whether a target deviant had been presented. All the deviants used in the test sessions were 7 dB higher than the other tones in the sequence, based on listeners achieving a sensitivity index (*d'*) of 1.0 or higher in pilot experiments with that increment level. Each stream had a 0.5 probability of including a deviant, hence listeners could not simply count the number of deviants and respond accordingly. In addition, the positions of the distractor and target deviants were randomized.

Four different frequency separations between the high and low tones were used: 1, 6, 15 and 20 semitones (ST). The frequency of the low tone was fixed at 1000 Hz while the frequency of the high tone varied between trials. For the 1 and 6 ST conditions, both tones were presented at 70 dB SPL. For the 15 and 20 ST conditions, the higher tone was presented at 70 dB SPL, and the level of the lower frequency tone was adjusted according to the ISO 226:2003 equal-loudness contours to be the same loudness as the higher (70 dB SPL) tone (1.5 dB lower). Within a block, the order of presentation of trials was randomized for the four frequency separations and the four probe types. Participants received visual feedback at the end of each trial.

Each participant undertook an initial session with 5 repetitions of the test sequence without any priming tone sequence at the maximum frequency separation of 20 ST. For each trial, their unbiased percept (i.e., when they were not provided with instructions on what to attend to within the sound sequences) was noted. For this, the participants were asked to simply listen to the sound sequence and report what they heard. The subjective percepts were collected as free responses. Participants were not informed of what the expected percept was and new naïve participants were recruited for each experiment. All participants who were tested in the experiments spontaneously reported either the percept of Right-Low and Left-High or the percept of Right-High and Left-Low. None of the participants reported any of the other irregular percepts described by Deutsch (1981). Hence, the un-cued spontaneous percepts were classified as either one of the two possible percepts. At no point in the experiment were the listeners told how the illusion was thought to occur or what the stimulus configuration was. Participants were presented with all the frequency separations (1, 6, 15 and 20 ST) with the different priming sequences (high and low frequency priming tones in the right and left ear) to check if the priming sequence had an effect on their percept of the illusion. For example, the participants were primed to L/Hi tones and their percept at the end of the test sequence was noted. The participants carried out 40 trials of this subjective block where each trial had one of the four cues. Following this block, they carried out one practice block of the deviant detection task.

For the actual test conditions, each participant completed 20 blocks of the task. Each block consisted of 96 trials. For each frequency separation, there were three deviant (where the deviants were in early/mid/late positions) and three non-deviant trials per block. The order of all trials was fully randomized. Each block took approximately 10 minutes, depending on the participants' response times. The testing was broken up into two sessions of approximately two hours each.

### C. Experiment 2: Stimuli and procedure

The second experiment combined EEG and psychophysical measurements to investigate the perception and neural representation for a stimulus similar to that used in experiment 1. The primary difference was that EEG was only carried out for one frequency separation, where the frequencies of the low and high tones were fixed at 1000 and 3000 Hz, respectively (~19 ST). The level of each tone was 70 and 68.5 dB SPL respectively to ensure equal loudness. As in experiment 1, each ear was presented with an alternating sequence of 12 pure tones per trial (see Figure 2). One amplitude deviant was placed on at least three of the four types of tones (R/Lo, R/Hi, L/Lo and L/Hi) either at the start, middle, or at the end of the sequence. The sequences were again generated in MATLAB at a sampling rate of 44.1 kHz. The stimuli were presented using E-prime (Psychology Software Tools, Inc. Sharpsburg, PA, USA) through Etymotic Research ER-2 insert transducers (Etymotic Research, Elk Grove Village, IL, USA) in a sound-treated room. Depending on the priming sequence, one of the deviants would be the target deviant and others would be distractor deviants for that particular trial. Each stream had a 0.5 probability of including a deviant. The participants were required to detect the amplitude deviants in the stream of sounds that they were cued to (target deviants) and ignore the others, responding via button press at the end of each trial. Feedback was given at the end of each trial. Each listener was presented with 160 trials per priming condition during the test session.

EEG was acquired continuously using a 64-channel BrainVision system consisting of a Brain-Vision™ recorder (Version 1.01b) and a Brain-Vision professional BrainAmp™ integrated amplifier system (Brain Products GmbH, Germany). The signal was digitally sampled at an A/D rate of 1000 Hz (32-bit resolution). Participants were fitted with an electrode cap fitted with 64 silver/silver chloride scalp electrodes positioned in an electrode ‘Easy Cap’ (Falk Minow Services, Herrsching-Breitbrunn, Germany). Electrode impedance was monitored and maintained at a minimum (typically below 5 kΩ).

### D. EEG Analysis

EEG pre-processing, epoching, and averaging was carried out using the EEGLAB toolbox (Delorme and Makeig, 2004). Data was down-sampled and then filtered using a zero-phase-shift bandpass filter from 0.1 Hz to 30 Hz. Baseline was corrected to -100 ms before stimulus onset, followed by artefact rejection at +/- 150 microvolts. Independent component analysis (ICA) was used to remove artefacts related to eye movements and blinks.

The EEG signal for each attention condition (R/Lo, R/Hi, L/Lo, and L/Hi) was separated into epochs 2850 ms long (corresponding to the length of one stimulus sequence including a 100-

ms baseline). These were then grouped separately for correct trials (where the target deviants were correctly detected) and for incorrect trials (targets were either not detected or with false positive behavioral results). As most of the participants had *d'* values greater than 1.0, there were more correct epochs than incorrect epochs. Hence, for the second half of the analysis between correct and incorrect trials, a random subset of the correct trials was chosen to equal the number of incorrect trials in that condition.

The EEG activity was averaged individually for each of the four primed attention conditions: attend to R/Lo, R/Hi, L/Lo, and L/Hi (separately for correct and incorrect trials). Next, the responses were averaged within each pair of conditions that involved attention to tones that were presented synchronously. For example, for priming conditions of R/Lo and L/Hi, the evoked response waveform would show the same effect of attention, as the R/Lo and L/Hi tones are synchronous. In other words, the responses to the R/Lo and L/Hi conditions were averaged, as were the responses to the L/Lo and R/Hi conditions. Finally, the responses to the two pairs of conditions (R/Lo-L/Hi and L/Lo-R/Hi) were subtracted from each other in order to cancel out the common (in-phase) 6-Hz activity (as the tone presentation rate in each ear was 6 Hz) and hence to potentially enhance the relative level of the 3-Hz activity (due to attention to alternate tones). Spectral analysis using a short-time Fourier transform was carried out on the resultant waveforms in order to examine the power spectrum of the EEG waveforms.

# III. RESULTS

## A. Experiment 1: Behavioral results

Subjective reports obtained from participants when listening to a sequence with a large frequency separation (greater than 6 ST) between the low and high tones indicated that the spontaneous percept for the majority of participants (10/15) was of the high tone in the right ear alternating with the low tone in the left ear (R/Hi-L/Lo). The remaining five participants reported hearing the low tone in the right ear, alternating with a high tone in the left ear (R/Lo-L/Hi). No other perceptual configuration was reported by any of the 15 subjects. Next, for the subjective reports of the cued percepts, all 15 subjects reported perceiving the illusion for all the trials as predicted. For example, in the condition where the cue was L/Lo, all subjects spontaneously responded to report the low tone in the left ear and the high tone in the right ear. This observation was reliable and consistent across all subjects. It should be noted that at no point were the listeners told what the expected percept could be (the instructions for the free report were simply “What do you hear?”).

For the 15- and 20-ST frequency separations, the subjective reports after priming indicated that the priming sequence was indeed able to effectively manipulate the percept. For example, participants with the spontaneous perception of R/Lo-L/Hi reported hearing the reversed percept of R/Hi-L/Lo if the priming sequence was either high tones in the right ear or low tones in the left ear. In contrast, the subjective reports for the two smaller frequency separations (1 and 6 ST) suggested that participants perceived a fused stream and that they were not able to precisely locate the ear in which they heard the low and high tones.

In the detection tasks, the participants' sensitivity to the deviant target was estimated by calculating *d'* for the detection of deviants for all conditions. The value of *d'* here and elsewhere was calculated by subtracting the inverse cumulative standard normal distribution function of the proportion of false alarms (participant responses to trials in which there was no deviant in the target stream, as a proportion of all trials with no deviant in the target stream) from the inverse standard normal cumulative distribution function of the proportion of hits (participant responses to trials in which there was a deviant in the target stream, as a proportion of all trials in which a deviant was present in the target stream): *d'* = z(H) – z(F).

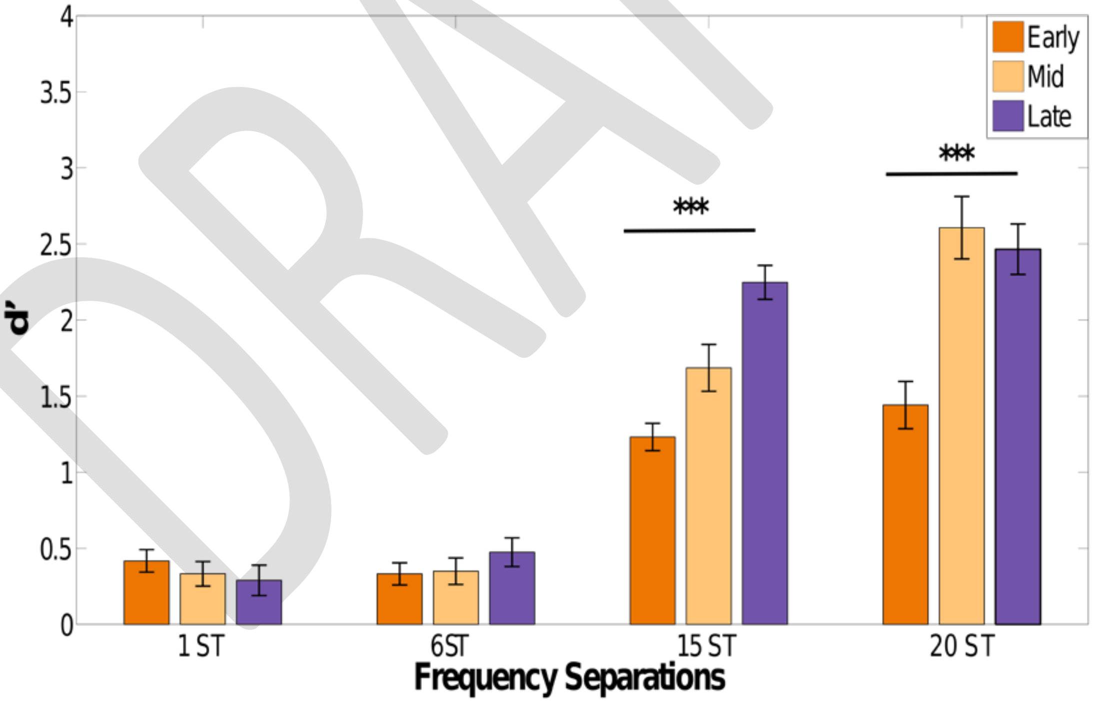

*FIGURE 3: (Color online) Average deviant detection scores across four different frequency separations from behavioral data obtained in Experiment 1. The three bars per frequency separation indicate the detection scores (d') for the early, mid and late deviants respectively. For the higher frequency separations, a significant increase in the detection scores of the late deviants compared to the early deviants was found. Error bars indicate 1 standard error from the mean. Asterisks indicate a significant difference of p<0.001.*

The data, averaged across the different priming conditions, are shown in Fig. 3. The use of *d’* measures in such a deviant detection paradigm has been used previously (e.g., Thompson et al., 2011). However, it should be noted that calculating *d’* measures in such a paradigm make assumptions about equal variance of the distributions of the responses, which may not be justified, as has previously been discussed (Swets, 1986a, 1986b; Verde et al., 2006).

Two predictions can be made if stream segregation plays a role in determining performance in this task. First, segregation is known to increase with increasing frequency separation (Miller and Heise, 1950; van Noorden, 1975); therefore, improved performance would be expected with increasing frequency separation between the two tones. Second, stream segregation tends to build up over time (Anstis and Saida, 1985; Bregman, 1978); therefore, performance should improve over the duration of each sequence, at least for frequency separations at which build-up is expected. The data are consistent with the first prediction, with overall performance increasing with increasing frequency separation from 1 to 20 ST (Fig. 3). The data are also consistent with the second prediction, with better performance observed during the latest than the earliest time periods, at least at the two larger frequency separations (Fig. 3). These trends were confirmed by a repeated-measures ANOVA on the values of *d'*, with three main factors: type of priming cue, frequency separation, and position of deviant. No main effect for type of priming cue was seen [F(3,30)=1.33,p=0.284] which indicates that there was no significant difference in the performance on the task for all four cue conditions. A main effect of frequency separation was observed [F(3,30)=758, $p<0.0001$], along with a main effect of position of deviant [F(2,20)=81.2, $p<0.001$]. Mauchly's test of sphericity indicated that the assumption of sphericity had not been violated for any of the above three factors (cue type: $\chi2(5)=2.86$, p =.723; frequency separation: $\chi2(5)=3.47$, p =.629; deviant position: $\chi2(2)=4.94$, p =.0.08). A significant interaction between frequency separation and deviant position was also observed [F(6,60)=58.2, $p<0.001$]. Post hoc tests for frequency separation indicated that the 1 and 6 ST conditions did not significantly differ from one another (p =0.58), but did differ from the 15- and 20-ST conditions ($p<0.001$), which also did not differ from each other (p=0.07). Post hoc tests also indicated a significant difference in the *d'* scores for early deviants compared to mid and late deviants in the 15- and 20-ST conditions ($p<0.001$).

### B. Experiment 2: Behavioral and EEG results

The behavioral results, averaged across the four conditions (R/Lo, R/Hi, L/Lo, L/Hi) for the single frequency separation (1000 and 3000 Hz), are shown in Fig. 4. Similar to the results obtained in experiment 1, a significant difference was observed between the deviant detection *d'* scores for the early and late target positions [F(1,9)=9.56, $p<0.01$].

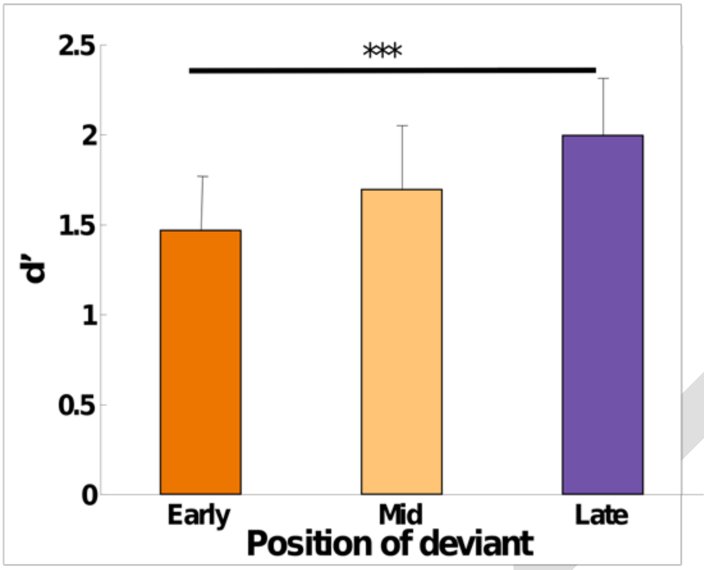

*FIGURE 4: (Color online) Average deviant detection results from behavioral data obtained in Experiment 2 averaged over 10 participants. The three bars per frequency separation indicate the detection scores (d') for the early, mid and late deviants respectively. Data showed a significant difference between d' scores for early and late deviants. Error bars indicate 1 standard error from the mean. Asterisks indicate a significant difference of p<0.001.*

As described in the methods, the EEG signals from two of the conditions (R/Lo and L/Hi) were averaged and subtracted from the sum of the EEG signals from the other two conditions (R/Hi and L/Lo) to enhance the difference between conditions in which participants attended to different time epochs. The prediction was that high activity at 3 Hz (the repetition rate of the target tones) would indicate enhancement of the attended tones. It was found that for the correct trials (all correct trials as well as the subset of correct trials taken to match the number of incorrect trials; see Fig. 5 and top panel of Fig. 6), activation around 3Hz emerged prominently during stimulus presentation, whereas it was markedly reduced during the trials that were incorrectly responded to (Fig. 6, middle panel).

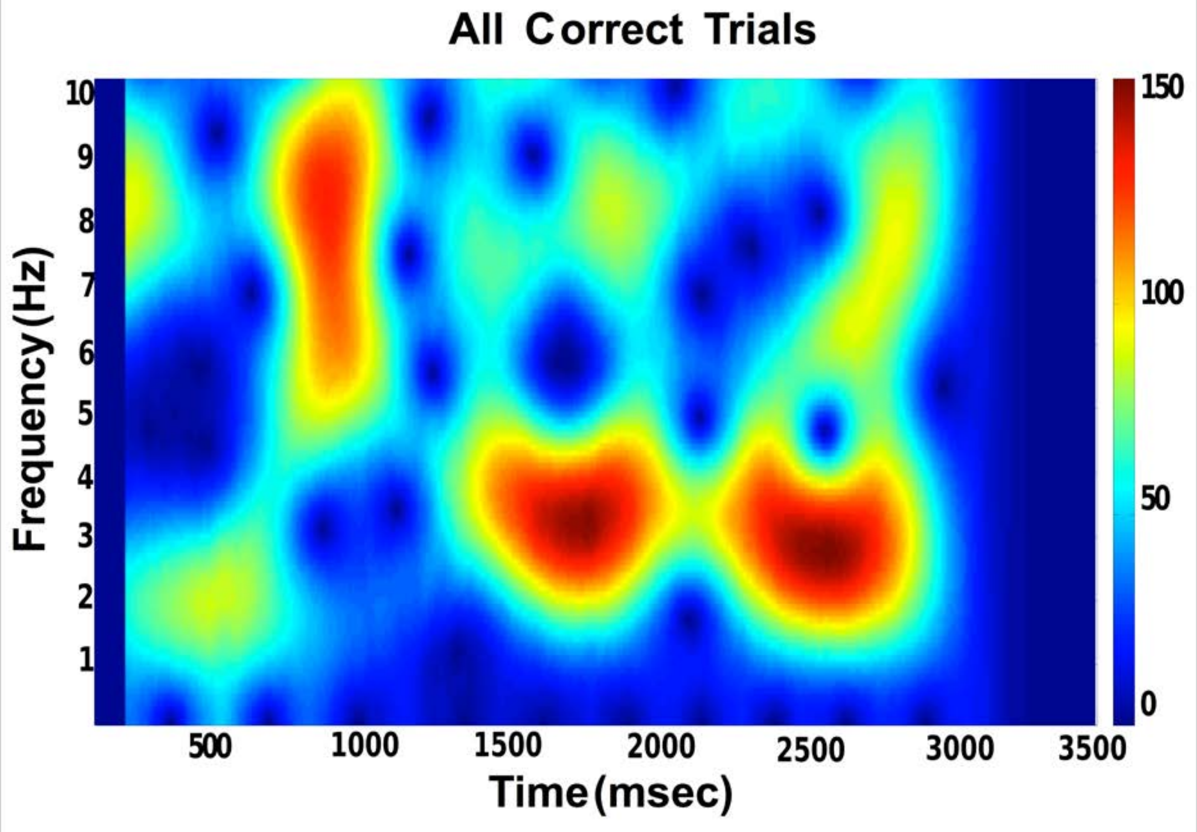

*FIGURE 5: (Color online) Spectral analysis of all correct data indicating a 3 Hz pattern (data combined across all 10 participants and all conditions). The spectral analysis was carried out on the averaged waveform across the four priming conditions as described in the methods section (the averaged waveform of conditions R/Lo and L/Hi were subtracted from the averaged waveforms of R/Hi and L/Lo). The color bar indicates power (μ$V^2$ ).*

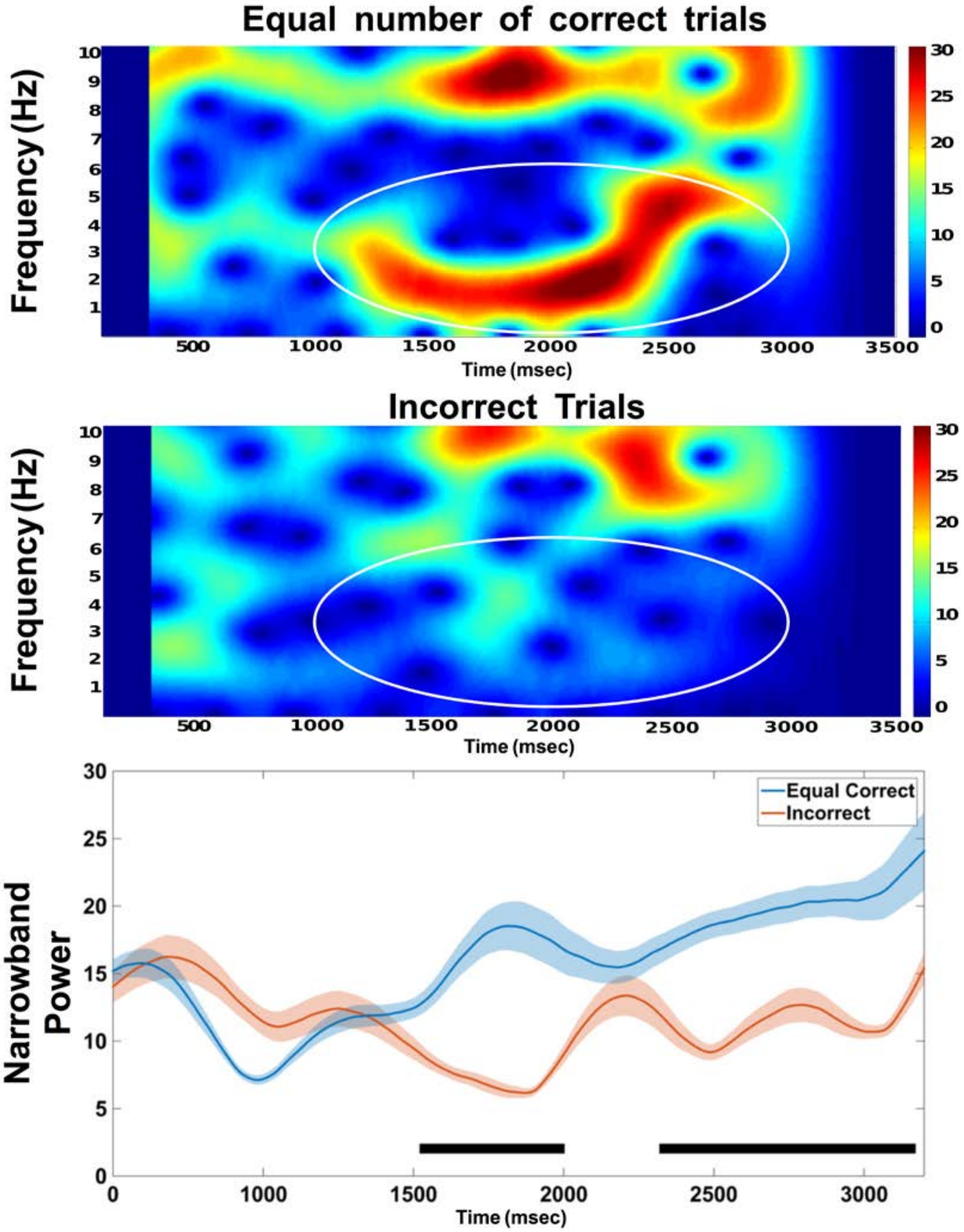

*FIGURE 6: (Color online) Subtracted waveforms Spectral analysis of equal number of correct and incorrect trials indicating a 3 Hz pattern (data combined across all participants and all conditions) for the correct trials (top panel) but not for the incorrect trials (middle panel). Bottom*

*panel shows the narrowband power in the region of 3 Hz for the correct and incorrect trials. The black bars show the temporal windows of significant difference ($p<0.001$). The color bar indicates power ($\mu V^2$ ).*

To further characterize the reliability of the 3-Hz activation in the spectrograms shown in Figures 6A and 6B, the individual narrowband power for an equal number of correct and incorrect trials was analyzed using a repeated-measures analysis. This repeated-measures analysis is derived from the cluster-based statistics approach described by Maris and Oostenveld (2007) and has been previously used with EEG measures (for e.g. Kouider et al., 2015). The analysis was carried out using the FieldTrip toolbox (Oostenveld et al., 2010) and is a non-parametric method similar to bootstrapping but it systematically controls for the problem of multiple comparisons. The black bars in Figure 6c show the time points where significant clusters of the difference between conditions were present (Monte-Carlo P value<0.05).

## IV. DISCUSSION

The present study investigated the percept elicited by a complex stimulus of alternating high and low tones played in opposite presentation phases in the two ears, known as Deutsch's octave illusion (Deutsch, 1974). The hypotheses tested were (1) that the illusion could be understood in terms of the basic principles of auditory streaming, (2) that the perception of the illusion could be manipulated by directed attention by changes in listening instructions provided via auditory priming cues, and (3) that the corresponding neural activity would mirror the changes in the perception of the illusion. The results provide support for all three hypotheses.

### A. Role of stream segregation

The octave illusion is thought to arise from mechanisms involving concurrent and sequential sound segregation. As noted earlier, there are certain key distinct properties of stream segregation that we could tap into to assess the critical role of streaming in establishing the illusion. They include the build-up of segregation over time and the effect of frequency separation on streaming. As previously noted by (Brancucci et al., 2009), the fact that the illusion breaks down at smaller frequency differences, in case of the current experiments, between 6 ST and 15 ST, suggests that it is mediated at least in part by auditory streaming constraints (van Noorden, 1975). The behavioral results from experiment 1 confirm and extend these observations by showing a deterioration in a performance-based task in conditions with a small frequency difference between the low and high tones (of 6 ST or less), suggesting a lack of stream segregation that results in an inability to “hear out” and follow a subset of tones within the complex sequence.

Another key indicator of streaming is a build-up of segregation over time as the sequence unfolds (Anstis and Saida, 1985). A build-up effect was observed when the frequency separation between the tones was large enough for participants to perform well in the deviant detection task (15- and 20-ST conditions). The build-up appears more rapid in the 20- than the 15-ST condition, in line with earlier work showing a very rapid build-up at large separations (Micheyl et al., 2007).

Thus the behavioral results are consistent with the hypothesis that the Deutsch illusion is sub served by the same mechanisms that govern auditory streaming. This is based on two characteristics observed for the octave illusion (frequency-separation dependence and build-up) that are consistent with the characteristics of auditory streaming.

### B. Effects of priming and directed attention on perception and EEG responses

In Deutsch's (1974) octave illusion, the alternating sequence of low and high tones in both ears were heard as a series of low tones in one ear *alternating* with the high tones in the other ear, and both heard at a rate that was *half* that of the actual presentation rate (Fig. 1, Percept). Thus, it appeared as if only the two tones from one ear were being perceived, with one of those two tones mis-located to the opposite ear (Deutsch, 1974; Deutsch and Gregory, 1978; Deutsch and Roll, 1976). Our stimuli in experiment 1 broadly evoke similar percepts but in a manner that could be manipulated by instructing participants, via a priming sequence, to attend specifically to one tone and ear or another. For example, if a participant’s unbiased percept of the illusion involves hearing the low tones in the right ear and high tones in the left ear, the participant can, with apparent ease, perceive the opposite percept of the low tones in the left ear and high tones in the right ear, if cued appropriately. This outcome shows that the illusion is robust but malleable to instructions and attention.

The simultaneously gathered data from EEG activity also indicated that participants were able to attend to the target tones in the correct ear, which were presented at half the rate of the stimulus, i.e., 3 Hz. Thus, consistent with the reported perception, in trials where the participants were able to detect the deviant in the target stream, neural activity at the target repetition rate (around 3 Hz) was enhanced, in phase with the target presentation times. Perhaps as expected, in trials where the participants were not successful in following the target tones (as evidenced by failure to detect the target deviant), activation around 3 Hz was markedly reduced, leading to a significant difference in the activation around 3 Hz between incorrect and correct trials, even when the same numbers of trials were evaluated in both correct and incorrect categories (see Fig. 6). The enhancement of EEG activity associated with the attended stream of tones is consistent with a growing body of literature showing enhanced responses to attended (and detected) streams in a background of other streams (Alain et al., 2001; Alain and Izenberg, 2003; Carlyon, 2004; Carlyon et al., 2001; Cusack et al., 2004; Dyson and Alain, 2004; Gutschalk et al., 2005, 2007, 2008; Hillyard et al., 1973; Zion Golumbic et al., 2013).

### C. The octave illusion as a probe of multistable perception and perceptual organization

Studies of the perception of, and neural responses to, perceptually multistable stimuli can help explain how objects or sources in the environment with conflicting or ambiguous cues are grouped according to specific characteristics to form a coherent representation of our surroundings (Schwartz et al., 2012). Several theories regarding the principles underlying perceptual bistability and multistability have been put forward. Leopold and Logothetis (1999) have suggested that a ‘central, supramodal mechanism’ underlies the perceptual decision making in multistable stimuli. Tong et al. (2006) proposed another model using multistable stimuli in the visual domain with a focus on the idea of distributed competition and have suggested that it is essential to understand the underlying neural mechanism involved in the processing of multistable stimuli, perceptual grouping and the effect of attention on them.

The multistable stimulus, used initially by Deutsch (1974), has been studied in various contexts and over a range of parameters using behavioral as well as neuroimaging techniques (Brancucci et al., 2009, 2012, 2014; Brännström and Nilsson, 2011; Deutsch, 1978; Deutsch and Roll, 1976; Lamminmäki et al., 2012; Lamminmäki and Hari, 2000; McClurkin and Hall, 1981). Our findings extend these results over wider parameter variations and, more specifically, focus on the role of auditory streaming and attention in this multistable illusion. In contrast to previous studies, we have sought to actively guide the subjects' percepts of the stimulus, thus inducing different perceptual organizations that could be measured objectively. We have found that the spontaneously experienced and widely reported perceptual organization was quite malleable to instructions or priming tones which obviated the advantages of any of the alternative percepts, as evidenced by equivalent performance across all different conditions. Indeed, the malleability of the percept of this ambiguous stimulus renders it as a highly promising tool with which to study further the perception and neural correlates of auditory stream segregation, as it involves both sequential as well as synchronous sound segregation.

### D. What mechanisms induce the illusion?

We have shown that the "octave illusion" is a robust percept that can be controlled by attention and persists over a wide range of frequencies and rates that closely parallel those observed in studies of streaming and auditory scene analysis. How can the emergence of this percept from these relatively simple stimuli be explained? Many explanations over the years have been based on a dual mechanism model in which the pitch and location of the tones are processed independently and then combined to give the percept (Deutsch, 1981; Lamminmäki et al., 2012). This mechanism and more elaborate versions of it (Chambers et al., 2002) have been shown to be inadequate as new experiments demonstrated the persistence of the illusion regardless of the octave (or any exact frequency) relationship between the tones or their spatial bilateral grouping (Brancucci et al., 2009; Brännström and Nilsson, 2011; Bregman and Steiger, 1980; Chambers et al., 2002; Deutsch, 2004). Furthermore, the dual mechanism model asserts that the pitch percept of the illusion corresponds to the frequency sequence present in the 'dominant' ear of the individual. If this were the case, then directing attention to the non-dominant ear of the listeners should not change the percept of the listeners. However, we find that the percept for all listeners can be manipulated by a simple precursor sequence (as indicated by the results of Experiments 1 and 2). Consequently, since the percept is malleable, it is not possible to explain the illusion based solely on theories involving a dominant ear, or on theories in which localization is dominated by the higher-frequency tone. Our results thus provide new constraints for future theories and models surrounding this long-established illusion.

## V. SUMMARY AND CONCLUSIONS

The purpose of this study was to investigate whether the octave illusion could be used as a potential tool to study the behavioral and neural effects of attention on concurrent as well as sequential stream segregation. Our results suggest that the illusory percepts seem to have common underlying mechanisms with auditory stream segregation. Furthermore, the percept can be manipulated by selective attention, which can be measured objectively using psychophysics as well as EEG. The methods introduced here therefore provide a potentially useful tool in the search for neural bases of auditory steaming and attention.

## ACKNOWLEDGMENTS

Work supported by the UCL Overseas Research Scholarship (AHM), UCL Graduate Research Scholarship (AHM), UCL Charlotte and Yule Bogue Research Fellowship (AHM), NIH grant R01DC07657 (SS & AJO), and Advanced ERC Grant "ADAM" #295603 (SS).

## References

Alain, C. (**2007**). "Breaking the wave: Effects of attention and learning on concurrent sound perception," Hear. Res., **229**, 225–236.

Alain, C., Arnott, S. R., and Picton, T. W. (**2001**). "Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials," J. Exp. Psychol. Hum. Percept. Perform., **27**, 1072–1089.

Alain, C., and Izenberg, A. (**2003**). "Effects of attentional load on auditory scene analysis," J. Cogn. Neurosci., **15**, 1063–1073.

Anstis, S. M., and Saida, S. (**1985**). "Adaptation to auditory streaming of frequency-modulated tones," J. Exp. Psychol. Hum. Percept. Perform., **11**, 257–271.

Billig, A. J., Davis, M. H., Deeks, J. M., Monstrey, J., and Carlyon, R. P. (**2013**). "Lexical Influences on Auditory Streaming," Curr. Biol., **23**, 1585–1589.

Brainard, D. H. (**1997**). "The Psychophysics Toolbox," Spat. Vis., **10**, 433–436.

Brancucci, A., Lugli, V., Perrucci, M. G., Del Gratta, C., and Tommasi, L. (**2016**). "A frontal but not parietal neural correlate of auditory consciousness," Brain Struct. Funct., **221**, 463-472.

Brancucci, A., Lugli, V., Santucci, A., and Tommasi, L. (**2011**). “Ear and pitch segregation in Deutsch’s octave illusion persist following switch from stimulus alternation to repetition,” J. Acoust. Soc. Am., **130**, 2179–2185.

Brancucci, A., Padulo, C., and Tommasi, L. (**2009**). “‘Octave illusion’ or ‘Deutsch’s illusion’?,” Psychol. Res. PRPF, **73**, 303–307.

Brancucci, A., Prete, G., Meraglia, E., di Domenico, A., Lugli, V., Penolazzi, B., and Tommasi, L. (**2012**). “Asymmetric Cortical Adaptation Effects during Alternating Auditory Stimulation,” PLoS ONE, **7**, e34367.

Brancucci, A., and Tommasi, L. (**2011**). “‘Binaural rivalry’: Dichotic listening as a tool for the investigation of the neural correlate of consciousness,” Brain Cogn., **76**, 218–224.

Brännström, K. J., and Nilsson, P. (**2011**). “Octave illusion elicited by overlapping narrowband noises,” J. Acoust. Soc. Am., **129**, 3213–3220.

Bregman, A. S. (**1978**). “Auditory streaming is cumulative,” J. Exp. Psychol. Hum. Percept. Perform., **4**, 380–387.

Bregman, A. S. (**1990**). Auditory Scene Analysis: The Perceptual Organization of Sound, MIT Press. Chapter-1, 1-45.

Bregman, A. S., and Steiger, H. (**1980**). “Auditory streaming and vertical localization: Interdependence of ‘what’ and ‘where’ decisions in audition,” Percept. Psychophys., **28**, 539–546.

Carlyon, R. P. (**2004**). “How the brain separates sounds,” Trends Cogn. Sci., **8**, 465–471.

Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (**2001**). “Effects of attention and unilateral neglect on auditory stream segregation,” J. Exp. Psychol. Hum. Percept. Perform., **27**, 115–127.

Chambers, C. D., Mattingley, J. B., and Moss, S. A. (**2002**). “The octave illusion revisited: Suppression or fusion between ears?,” J. Exp. Psychol. Hum. Percept. Perform., **28**, 1288–1302.

Christensen, I. P., and Gregory, A. H. (**1977**). “Further study of an auditory illusion,” Nature, **268**, 630–631.

Cusack, R., Decks, J., Aikman, G., and Carlyon, R. P. (**2004**). “Effects of Location, Frequency Region, and Time Course of Selective Attention on Auditory Scene Analysis,” J. Exp. Psychol. Hum. Percept. Perform., **30**, 643–656.

Darwin, C. J., and Carlyon, R. P. (**1995**). “Auditory grouping,” Handb. Percept. Cogn. Hear. London, Academic Press., 387-424.

Darwin, C. J., Hukin, R. W., and Al-Khatib, B. Y. (**1995**). “Grouping in pitch perception: Evidence for sequential constraints,” J. Acoust. Soc. Am., **98**, 880–885.

Delorme, A., and Makeig, S. (**2004**). “EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis,” J. Neurosci. Methods, **134**, 9–21.

Denham, S. L., Gyimesi, K., Stefanics, G., and Winkler, I. (**2010**). “Stability of Perceptual Organisation in Auditory Streaming,” In E. A. Lopez-Poveda, A. R. Palmer, and R. Meddis (Eds.), Neurophysiol. Bases Audit. Percept., Springer New York, pp. 477–487.

Deutsch, D. (**1974**). “An auditory illusion,” Nature, **251**, 307–309.

Deutsch, D. (**1978**). “Lateralization by frequency for repeating sequences of dichotic 400- and 800-Hz tones,” J. Acoust. Soc. Am., **63**, 184–186.

Deutsch, D. (**2004**). “The Octave Illusion Revisited Again,” J. Exp. Psychol. Hum. Percept. Perform., **30**, 355–364.

Deutsch, D., and Gregory, A. H. (**1978**). “Deutsch’s octave illusion,” Nature, **274**, 721–721.

Deutsch, D., and Roll, P. L. (**1976**). “Separate ‘what’ and ‘where’ decision mechanisms in processing a dichotic tonal sequence,” J. Exp. Psychol. Hum. Percept. Perform., **2**, 23–29.

Dyson, B. J., and Alain, C. (**2004**). “Representation of concurrent acoustic objects in primary auditory cortex,” J. Acoust. Soc. Am., **115**, 280–288.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., and Shamma, S. A. (**2009**). “Temporal Coherence in the Perceptual Organization and Cortical Representation of Auditory Scenes,” Neuron, **61**, 317–329.

Elhilali, M., and Shamma, S. A. (**2008**). “A cocktail party with a cortical twist: How cortical mechanisms contribute to sound segregation,” J. Acoust. Soc. Am., **124**, 3751–3771.

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., and Oxenham, A. J. (**2005**). “Neuromagnetic Correlates of Streaming in Human Auditory Cortex,” J. Neurosci., **25**, 5382–5388.

Gutschalk, A., Micheyl, C., and Oxenham, A. J. (**2008**). “Neural Correlates of Auditory Perceptual Awareness under Informational Masking,” PLoS Biol, **6**, e138.

Gutschalk, A., Oxenham, A. J., Micheyl, C., Wilson, E. C., and Melcher, J. R. (**2007**). “Human Cortical Activity during Streaming without Spectral Cues Suggests a General Neural Substrate for Auditory Stream Segregation,” J. Neurosci., **27**, 13074–13081.

Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. (**1973**). “Electrical signs of selective attention in the human brain,” Science, **182**, 177–180.

Kouider, S., Long, B., Le Stanc, L., Charron, S., Fievet, A.-C., Barbosa, L. S., and Gelskov, S. V. (**2015**). “Neural dynamics of prediction and surprise in infants,” Nat. Commun., **6**, 8537.

Lamminmäki, S., and Hari, R. (**2000**). “Auditory cortex activation associated with octave illusion,” Neuroreport, **11**, 1469–1472.

Lamminmäki, S., Mandel, A., Parkkonen, L., and Hari, R. (**2012**). “Binaural interaction and the octave illusion,” J. Acoust. Soc. Am., **132**, 1747–1753.

Leopold, D. A., and Logothetis, N. K. (**1999**). “Multistable phenomena: changing views in perception,” Trends Cogn. Sci., **3**, 254–264.

Maris, E., and Oostenveld, R. (**2007**). “Nonparametric statistical testing of EEG- and MEG-data,” J. Neurosci. Methods, **164**, 177–190.

McClurkin, R. H., and Hall, J. W. (**1981**). “Pitch and timbre in a two-tone dichotic auditory illusion,” J. Acoust. Soc. Am., **69**, 592–594.

Micheyl, C., Carlyon, R. P., Gutschalk, A., Melcher, J. R., Oxenham, A. J., Rauschecker, J. P., Tian, B., et al. (**2007**). “The role of auditory cortex in the formation of auditory streams,” Hear. Res., **229**, 116–131.

Micheyl, C., Hanson, C., Demany, L., Shamma, S., and Oxenham, A. J. (**2013**). “Auditory stream segregation for alternating and synchronous tones,” J. Exp. Psychol. Hum. Percept. Perform., **39**, 1568–1580.

Micheyl, C., Kreft, H., Shamma, S., and Oxenham, A. J. (**2013**). “Temporal coherence versus harmonicity in auditory stream formation,” J. Acoust. Soc. Am., **133**, EL188-EL194.

Micheyl, C., and Oxenham, A. J. (**2010**). “Objective and Subjective Psychophysical Measures of Auditory Stream Integration and Segregation,” J. Assoc. Res. Otolaryngol., **11**, 709–724.

Miller, G. A., and Heise, G. A. (**1950**). “The Trill Threshold,” J. Acoust. Soc. Am., **22**, 637–638.

Moore, B. C. J., and Gockel, H. E. (**2012**). “Properties of auditory stream formation,” Philos. Trans. R. Soc. B Biol. Sci., **367**, 919–931.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., Oostenveld, R., Fries, P., Maris, E., et al. (**2010**). “FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data.” Comput. Intell. Neurosci., **2011**, e156869.

Pelli, D. G. (**1997**). “The VideoToolbox software for visual psychophysics: Transforming numbers into movies,” Spat. Vis., **10**, 437–442.

Pressnitzer, D., and Hupé, J.-M. (**2006**). “Temporal Dynamics of Auditory and Visual Bistability Reveal Common Principles of Perceptual Organization,” Curr. Biol., **16**, 1351–1357.

Schwartz, J.-L., Grimault, N., Hupé, J.-M., Moore, B. C. J., and Pressnitzer, D. (**2012**). “Multistability in perception: binding sensory modalities, an overview,” Philos. Trans. R. Soc. B Biol. Sci., **367**, 896–905.

Shamma, S. A., and Micheyl, C. (**2010**). “Behind the scenes of auditory perception,” Curr. Opin. Neurobiol., **20**, 361–366.

Shinn-Cunningham, B. G., Lee, A. K. C., and Oxenham, A. J. (**2007**). “A sound element gets lost in perceptual competition,” Proc. Natl. Acad. Sci., **104**, 12223–12227.

Swets, J. A. (**1986**). “Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance,” Psychol. Bull., **99**, 181–198.

Swets, J. A. (**1986**). “Indices of discrimination or diagnostic accuracy: Their ROCs and implied models,” Psychol. Bull., **99**, 100–117.

Thompson, S. K., Carlyon, R. P., and Cusack, R. (**2011**). “An objective measurement of the build-up of auditory streaming and of its modulation by attention,” J. Exp. Psychol. Hum. Percept. Perform., **37**, 1253–1262.

Tong, F., Meng, M., and Blake, R. (**2006**). “Neural bases of binocular rivalry,” Trends Cogn. Sci., **10**, 502–511.

van Noorden (**1975**). Temporal coherence in the perception of tone sequences. Technische Hogeschool Eindhoven. PhD thesis, 17-24.

Verde, M. F., Macmillan, N. A., and Rotello, C. M. (**2006**). “Measures of sensitivity based on a single hit rate and false alarm rate: The accuracy, precision, and robustness of′,Az, andA’,” Percept. Psychophys., **68**, 643–654.

Vliegen, J., and Oxenham, A. J. (**1999**). “Sequential stream segregation in the absence of spectral cues,” J. Acoust. Soc. Am., **105**, 339–346.

Winkler, I., Denham, S., Mill, R., Bőhm, T. M., and Bendixen, A. (**2012**). “Multistability in auditory stream segregation: a predictive coding view,” Philos. Trans. R. Soc. B Biol. Sci., **367**, 1001–1012.

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., et al. (**2013**). “Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a ‘Cocktail Party,’” Neuron, **77**, 980–991.

Zwicker, T. (**1984**). “Experimente zur dichotischen Oktav-Täuschung,” Acta Acust. United Acust., **55**, 128–136.