An auditory illusion reveals the role of streaming in the temporal misallocation of perceptual objects

Anahita H. Mehta[1], Nori Jacoby[2], Ifat Yasin[3], Andrew J. Oxenham[4], Shihab A. Shamma[5,6]

[1]UCL Ear Institute, University College London, London, UK, WC1X 8EE
[2] The Center for Science & Society, Columbia University, New York, USA, NY 10027
[3]Department of Computer Science, University College London, London, UK, WC1E 6BT
[4]Department of Psychology, University of Minnesota, Minneapolis, USA, MN 55455
[5]Electrical and Computer Engineering Department and Institute for Systems Research, University of Maryland, College Park, Maryland, USA, MD 20742
[6]École Normale Supérieure, 75005 Paris, France

Running head:

Auditory illusions and streaming

Changes of affiliation:

Anahita H. Mehta is now at the Department of Psychology, University of Minnesota.

Corresponding author:

Anahita H. Mehta

N625 Elliott Hall, 75 East River Parkway, Minneapolis, MN 55455

mehta@umn.edu

Keywords:

Auditory streaming, Octave illusion, Attention, EEG

Abstract

This study investigates the neural correlates and processes underlying the ambiguous percept produced by a stimulus similar to Deutsch's "octave illusion", in which each ear is presented with a sequence of alternating pure tones of low and high frequencies. The same sequence is presented to each ear, but in opposite phase, such that the left and right ears receive a High-Low-High… and a Low-High-Low… pattern, respectively. Listeners generally report hearing the illusion of 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. The current explanation of the illusion is that it reflects an illusory feature conjunction of pitch and perceived location. Using psychophysics and EEG measures, we test this and an alternative hypotheses involving synchronous and sequential stream segregation, and investigated potential neural correlates of the illusion. We find that the illusion of alternating tones arises from the synchronous tone pairs across ears rather than sequential tone streams within one ear, suggesting that the illusion involves a misattribution of time across perceptual streams, rather than a misattribution of location within a stream. The results provide new insights into the mechanisms of binaural streaming and synchronous sound segregation.

Introduction

Illusions can be intriguing and entertaining, but can also provide important insights into the functioning and underlying mechanisms of perception (1–5). The “octave illusion,” first reported by Diana Deutsch (6), was originally elicited with a stimulus configuration consisting of two pure tones, spaced an octave apart, presented in an alternating low-high tone pattern with different phases at the two ears, such that if the sequence in the left ear started with a low tone, the sequence in the right would start with a high tone. The result was an unexpected illusory percept, where listeners perceived all the low tones in one ear at half the presentation rate, alternating with the high tones in the other ear, also at half the rate (see Figure 1-A).

The stimulus used to elicit the octave illusion has been studied in different contexts and the robustness of the percept has been investigated across a variety of parameters. It has been demonstrated that the percept of this illusion is robust to changes in tone duration (7) and spectral shape (8), and can also be elicited by quasi-periodic stimuli like band-pass noise (9). It was also noted by Deutsch and Roll (10), and later confirmed by Brancucci *et al*. (11), that the illusion is not dependent on the tones being in an exact octave relationship. Indeed, Brancucci *et al*. (11) reported that the illusory percept was present for all musical intervals tested that were larger than a perfect fourth (roughly a ratio of 4:3 or a frequency difference of 33%). Despite the fact that it is not dependent on the octave relationship, we continue to refer to the phenomenon as the “octave illusion” for historical reasons.

To explain the illusion, Deutsch (1) proposed a dual-mechanism model that consists of one mechanism for pitch determination and another for sound localization. The outputs of these mechanisms converge to elicit the illusory percept. The model is based on the assumption that the perceived pitch corresponds to the frequency of the tone presented to the listeners’ “dominant” ear (usually the right), whereas the perceived location of the tone corresponds to the location of the higher-frequency tone (10), so that the final illusory percept is a combination of the output of the two mechanisms, in a feature-combination operation (12). Although some authors have questioned this interpretation (13,14), the most recent studies have verified the basic observations and interpretations of the illusion (12,15).

A number of neuroimaging studies have been carried out using stimuli related to the octave illusion (16–20). Lamminmäki and Hari (17) aimed to find the neurophysiological basis of the ‘where’ mechanism of Deutsch’s dual-mechanism model. The stimuli were 400- and 800-Hz pure tones presented to the left (L) or right (R) ears as follows: L400/R400, L400/R800, L800/R400 and L800/R800. The aim of their study was to find out whether the lateralization of the auditory evoked fields using MEG, in particular the N100m peak, co-varied with the sound localization percept. They found that the N100m was stronger in the hemisphere contralateral to the high-pitch sound, in agreement with the established finding that monaural sounds evoke stronger N100m responses in the hemisphere contralateral to the sound (21). However, the MEG measurements were not carried out on the stimulus eliciting the octave illusion itself, and no attempt was made to relate the neural responses to perception, as the measurements were made with listeners in a passive role, with no task, and no indication as to what the listeners perceived on a trial-by-trial basis. Lamminmäki et al. (18) next investigated the neuromagnetic correlates of the “where” aspect of the dual-mechanism model using frequency-tagged stimuli. Each tone in the stimuli was modulated using a unique ‘tagging’ frequency that helps parse out the corresponding neuromagnetic activity for each tone. They found evidence for binaural suppression and right ear dominance for all their stimuli and concluded that the findings of their study were in line with the dual-mechanism model. Again, however, the authors used a passive paradigm, with no subjective or objective measures of perception or attention, and the stimuli were limited to isolated dichotic tone pairs, rather than illusion-inducing sequences. Several other studies have used the illusion to study aspects of the neural correlates of consciousness, by taking advantage of the fact that the same stimulus can spontaneously elicit different percepts in different listeners and across different repetitions (20,22,23).

A relatively new approach to understanding the octave illusion comes from the perspective of auditory streaming (17,24). Auditory streaming refers to the perceptual organization of sound sequences that may either be perceived as arising from a single source or multiple sources (25). A recent study showed that the octave illusion shares a number of properties with auditory streaming, including i) the requirement of a minimum frequency difference of several semitones between the two tones for the illusion to occur, and ii) a temporal build-up, whereby the illusion is more likely to occur later than earlier in a sequence (22). The study also showed that the illusion was affected by instructions, and that all listeners reported hearing the original sequence in different ways, depending on which of the four tones they were instructed to attend to (e.g., low tone on the left, or high tone on the right). However, although the illusion shares many properties with streaming, there is no obvious way to explain the illusion in terms of the usual heuristics associated with streaming, such as frequency similarity or temporal proximity (26). The aim of the current study was to provide further empirical data on the octave illusion, in particular to address the question of which tones within the stimulus are most salient in the illusory percept. The first experiment provided two behavioral tests of the illusion, and the second experiment combined behavior and EEG to probe the neural correlates of the illusion. Our results suggest that the illusion results from a misattribution of timing relations between two synchronous, spatially separated tones, rather than (as previously believed) a misattribution of spatial relations between two temporally alternating tones.

Experiment 1

Rationale

The aim of this experiment was to investigate which physical tones contribute most to the illusory percept outlined in Figure 1-A. One tone of the alternating percept can be made the focus of attention by using instructions and/or a sequence of preceding cue tones. It has been assumed that the other tone forming the illusion is the tone in the same ear as the target, alternating in time. This experiment provides two direct empirical tests of that assumption.

Method

Participants

Fifteen listeners (six male and nine female, aged 21–30 years) participated in experiment 1. All listeners had normal hearing, defined as audiometric hearing thresholds no higher than 15 dB Hearing Level (HL) at octave frequencies from 0.25 to 4 kHz, with no history of hearing or neurological disorders. Listeners provided written informed consent and were compensated for their participation. The experiment was carried out at University College London. The University College London Ethics Committee approved the procedure for the experiment. All the participants used were naïve and had not taken part in any other related experiments.

All 15 listeners completed both paradigms described below. The whole experiment took about 2 hours. For each paradigm, there were 5 blocks with 12 test trials (60 trials per paradigm in total). The experiment was blocked according to paradigm. Seven participants completed paradigm 1 before paradigm 2, while the others were tested in the reverse order.

Paradigm 1: Stimuli and procedures

Participants were cued, using a precursor sequence (see Figure 1-B), to attend to one of the four tones within the main sequence. The precursor sequence consisted of three low- or high-frequency tones presented either to the left or right ear prior to the main sequence, in order to indicate the side and frequency to which participants should attend. The side and frequency of the precursor tones were selected at random with equal *a priori* probability on each trial. Following a silent interval of 500 ms, the main sequence of each trial began, as shown in Figure 1-B, with alternating low (1000-Hz) and high (2996-Hz) tones, marked Lo and Hi, respectively. A frequency separation larger than an octave was used because this has been shown to be effective in inducing the illusion (11) and it avoids some potentially confounding influences of using an exact octave (27). Each tone was 100 ms in duration, including 10-ms raised-cosine onset and offset ramps, and tones were separated by 50-ms silent intervals. All tones were presented at 65 dB SPL. The sequence was presented for a total of 6 s (20 repetitions of the alternating synchronous tones as seen in Figure 1-B). During the main sequence of each trial, the tones in one of the two tone sequences at the uncued frequency were sinusoidally amplitude modulated at a rate of 34.47 Hz and with a depth of 75%. For example, in Figure 1-B, the low tones in the right ear are cued, and the high tones that alternate with the cued tones are amplitude modulated. The modulation was randomly assigned on each trial to the tones that were either synchronous or alternating with the cued tones with equal *a priori* probability. For example, on a trial where the precursor tones were low tones in the right ear, the modulated tones could either be the alternating high tones in the right ear or the synchronous high tones in the left ear.

The listeners’ task was to report whether the illusion consisted of modulated tones or unmodulated (pure) tones. No feedback was provided, as there was no correct answer. In the schematic presented in Figure 1-B, if the listener perceived the illusion with one of the tone sequences being amplitude modulated, it would mean that the percept arose from the tones that alternated with the target tones. If instead the listener reported hearing no amplitude modulation in the illusion, it would suggest that the percept was determined from the (unmodulated) tones that were synchronous with the target tones.

Before the main experiment, listeners completed thirty trials in which they were asked to indicate whether a sequence of tones was amplitude modulated or not. A one interval, yes-no task was used, where the stimulus was a diotic sequence of three Lo or Hi tones. 50% of the trials contained modulated tones while the others contained pure tones. Trials were randomized for the presence of modulation as well as carrier frequency (low or high). The tone parameters were identical to the ones for the main experiment. The listeners received visual feedback after each trial. This block was conducted to ensure that all listeners could distinguish between modulated and unmodulated tones. The performance of all the listeners was at ceiling for this task, indicating that they could clearly distinguish between modulated and unmodulated tones.

All stimuli were generated in MATLAB (MathWorks Inc. Natick, MA, USA) and were presented at a sampling rate of 44.1 kHz, using the Psychophysics Toolbox extension in MATLAB (28,29) through Sennheiser HD 215 headphones. All testing took place in a sound treated test booth.

Paradigm 2: Stimuli and procedures

The stimuli for this paradigm were similar to those for paradigm 1, and the generation and presentation methods were identical. Listeners were again cued to attend to one of the four streams through a sequence of three low or high precursor tones either in the left or right ear. In this paradigm, the tones in one of the two tone sequences at the uncued frequency were gradually faded out and back in (see Figure 1-C). For instance, in Figure 1-C, the listener is cued to the low tones in the right ear and the synchronous high tones (tones presented synchronously with the cued tone sequence) in the left ear are faded out and in. The fade was achieved by decreasing the level of each successive tone in the tone sequence by 6 dB until the level was 18 dB below the level of the other tones, and then increasing the level of each successive tone by the same amount. Which of the two tones at the uncued frequency was faded in and out was selected randomly with equal *a priori* probability on each trial.

The listeners' task was to report whether illusion was perceived with or without a fading in and out in loudness of one of the alternating tones. Again, no feedback was provided, as there was no correct answer. In the example in Figure 1-C, if the listener perceived the illusion with a fading in and out of one of the alternating tones, it would indicate that the illusory percept arose from the tones that were synchronous with the cued tones. If the listener reported not hearing the fading in and out within the illusion, it would mean that the percept was determined from the tones that alternated with the cued tones. Demonstrations for both paradigms are available in the supplementary information.

Results

The response for each trial was scored according to whether it corresponded to the tones that were synchronous or alternating with the cued tones. For example, if the listener responded to the trial in Figure 1-B as ‘No modulation perceived’, the response would be marked as a synchronous (opposite ear) tone heard, whereas if the modulation was reported, the response would be marked as an alternating (same ear) tone heard. No significant effects of cueing condition (R/Lo, L/Lo, etc.) were observed for either paradigm [Paradigm 1: $F(3,56)=1.28$, $p=0.269$; Paradigm 2: $F(3,56)=2.36$, $p=0.168$], so the results were collapsed across all four conditions. For both the paradigms, the responses across all four conditions were pooled and the proportion of responses corresponding to the synchronous and alternating tones was calculated. These proportion scores were then converted to a scaled score between -1 and +1 by subtracting 0.5 (to make the average zero in the case where synchronous and alternating responses were equal), and multiplying by 2 (to scale from -1 to 1). Thus, if a listener always heard the tone that alternated with the cued tone, the score would be -1, whereas if the synchronous tone was always heard, the score would be +1.

Individual results from the 15 participants, averaged across the four conditions for each of the two paradigms, are shown in Figure 1-D. Most responses were positive, indicating that changes were heard more clearly when they occurred simultaneously with, and in the opposite ear to, the cued tone. A one-sample t-test confirmed that the mean scores for both paradigms were significantly greater than zero [Paradigm 1: $t(14) = 4.36$, $p<0.001$; Paradigm 2: $t(14) = 3.13$, $p<0.001$].

Discussion

The results from both paradigms were consistent in suggesting that listeners’ perception of the alternating tone-sequence in the non-cued ear corresponded to the tones in the non-cued ear that were synchronous with the cued tones and *not* to the alternating tones in the cued ear, as has been previously assumed. This surprising result suggests that it is a perceptual temporal misalignment between the synchronous tones that is responsible for the perception of "alternating" tones, rather than a spatial misattribution of the alternating tones in the same ear as the cue tones, as has generally been assumed. The fundamental question of which tones contribute to the perception of the illusion has been studied in several contexts indirectly (11,13,16) and directly by Deutsch and Roll (10). However, the paradigm used by Deutsch and Roll to study this question did not elicit the octave illusion itself, which makes the interpretation of their results less clear. Experiment 2 followed up on this surprising finding, by combining a further perceptual test with EEG correlates of the illusion.

Experiment 2

Rationale

The aim of this experiment was to provide a further test of the surprising conclusion of Experiment 1 that the tones forming part of the illusion were the ones that were synchronous with the target tones, and not, as previously believed, the tones that were alternating with the target tones. In this experiment, EEG was combined with behavior, and the tones of the illusory stimulus were differentially tagged via amplitude modulation to obtain a direct measure of which tones were most prominent neurally, and hence most likely to be perceptually salient (18,30,31).

The different tones within each sequence were amplitude modulated at different rates, in order to identify their responses in the EEG signal. The hypothesis of this experiment was that the modulation rate corresponding to the contralateral tones synchronous with the cued tones would show an increase in amplitude, relative to the tones that were alternating with the cued tones. For example, if the listener were cued to the low tones in the right ear, then the neural response to the modulation frequency of the synchronous high tones in the left ear should be larger than the neural response to the modulation frequency of the high tones in the right ear.

Participants

Thirteen listeners (six male and seven female, aged 21-30 years) participated in experiment 2. All listeners were naïve and had not taken part in any other related experiments. All participant recruitment procedures and inclusion criteria were the same as for Experiment 1.

Stimuli and procedures

All stimuli were presented using Presentation (Neurobehavioral Systems Inc. Berkeley, CA, USA) through Etymotic Research ER-2 insert earphones (Etymotic Research, Elk Grove Village, IL, USA) in a sound-treated room. The stimulus paradigm was similar to that used in experiment 1, with low and high tone frequencies of 1000 and 2996 Hz, respectively. A schematic diagram of a single sample trial is shown in Figure 2-A. At the start of each trial, a precursor consisting of three low (1000-Hz) tones was presented to either the left or right ear. Each tone was 203.1 ms long with a silent gap of 50 ms between each of the three tones. The precursor was followed by a 1000-ms silent gap before the beginning of the test sequence.

In the test sequence, each ear was presented with a sequence of high and low tones as before. In Figure 2-A, the low tones are indicated by the boxes marked ‘Lo’ and the high tones are marked ‘Hi’. The high tones in each ear were sinusoidally amplitude modulated using modulation frequencies of either 34.47 Hz or 44.31 Hz (indicated by the blue or red outlined boxes), at a modulation depth of 80%. Each tone in the main sequence was also 203.1 ms long and separated by 50-ms silent gaps. To maximize the number of trials per illusory percept, only low precursor conditions were chosen, as this allowed us to test both configurations of the illusory percept (either R/Lo alternating with L/Hi or vice versa). In a previous study (22), we found no difference between the cueing conditions; therefore fewer cuing conditions were chosen for this study.

Each test sequence consisted of 40 tone pairs. The total duration of the test sequence was 10.124 s. The task was to detect a deviant among one of the cued low-frequency tones. The deviants had a 5-dB increase in level, relative to the 70 dB SPL level of the other tones. Depending on the priming sequence, one of the deviants would be the target deviant and others would be distractor deviants for that particular trial. For example, if the precursor low tones were presented to the left ear, a deviant in the left low tone sequence would be the target. Each tone sequence had a 0.5 probability of including a deviant. The targets and deviants were randomly distributed between the 10th and 35th tone. The number of distractor deviants could range from 0 to 3. There was only one target deviant, if present, per trial.

The total EEG stimulus set was counterbalanced for the cued ear and the tagging modulation rate by dividing the set into four conditions. In conditions 1 and 2, listeners were cued to the low-frequency tones in the left and right ear, respectively, while the high-frequency tones in the left ear were modulated at 34.47 Hz, and the high-frequency tones in the right ear were modulated at 44.31 Hz. In conditions 3 and 4, listeners were cued to the low-frequency tones in the left and right ear, respectively, while the high-frequency tones in the left ear were modulated at 44.31 Hz and the high-frequency tones in the right ear were modulated at 34.47 Hz. Two control conditions (conditions 5 and 6) were included to establish a baseline for the tagged frequencies. The control stimuli had only low-frequency unmodulated tones in one ear and only high-frequency modulated tones presented synchronously in the opposite ear (Lo = 1000 Hz with no modulation; Hi = 2996 Hz tagged with modulation frequencies of 34.47 Hz or 44.31 Hz) with the same parameters as in conditions 1-4. All tones in the main sequence were also 203.1 ms long and were separated by 50-ms silent gaps (Figure 3-A). Listeners were cued by a low-frequency tone sequence on either side and were asked to indicate whether amplitude deviants in the cued stream were present or absent (same as conditions 1-4). The control stimuli did not elicit the octave illusion; their purpose was to establish a baseline for the EEG amplitude of the tagged frequencies.

The EEG measurements were preceded by a series of behavioral tests. In the first block of ten trials, listeners heard the illusory sequence with no precursor tones and no modulation. 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. Next, listeners were presented with another block of ten trials, where their perceptual responses to the stimulus with low-frequency pure tones and high-frequency modulated tones were recorded. Finally, listeners were presented with a block of ten trials in which the full stimulus was presented (precursor plus main sequence, as in the EEG experiment). Half the trials had the cue presented on the left, and the other half had the cue presented on the right. Again, listeners were asked to report their percepts. For all three blocks of trials, the listeners were naïve to the stimuli and were not told what the expected response was.

In the main EEG portion of the experiment, the stimuli were presented in either ‘test’ blocks (conditions 1-4) or ‘control’ blocks (conditions 5-6). Within each of the blocks, the trials were randomized for cueing sequence type (cues could be low tones in the Right or Left ear) and tagging frequency. Each block included 120 trials and each listener was tested using 4 test blocks and 2 control blocks. Hence, 480 test trials and 240 control trials were conducted for each listener – 120 per condition. For each trial, the listeners were asked to focus on the cued stream (as determined by the precursor). At the end of each trial, the listener had to report via a button press if a target deviant was present or absent. The next trial was initiated 1 s after the response.

EEG signals were acquired continuously using a 64-channel BioSemi active-electrode EEG system (BioSemi Inc., Amsterdam, Netherlands). They were digitally sampled at an A/D rate of 2048 Hz (64-bit resolution). Listeners were fitted with an electrode cap fitted with 64 silver/silver-chloride scalp electrodes. Electrode impedance was monitored and typically maintained below 5 kΩ.

Data Analyses

Behavioral data analyses

The value of the discriminability index, *d'*, was calculated as: $d' = z(H) - z(F)$, where H is the hit rate or the proportion of “target heard” responses when the target was present and F is the false alarm rate or the proportion of “target heard” responses when the target was not present.

EEG analyses

EEG pre-processing, separating the EEG data according to conditions, and averaging were carried out using the EEGLAB toolbox (32). Data were down-sampled and then filtered using a zero-phase band pass filter from 0.1 Hz to 70 Hz. EEG amplitude was measured relative to a 500-ms pre-stimulus baseline. Independent component analysis (ICA) was used to remove artifacts related to eye movements and blinks (33). The EEG data were separated according to the six conditions (four test and two control) and were averaged across a select subset of channels from the left, right and central electrode positions over the temporal and parietal regions, similar to the ones used in previous studies (20). The data were analyzed in terms of relative spectral strength of the tagged frequencies across conditions and for differences in the EEG waveform.

The EEG signal epoch was calculated from the onset of the test sequence to the end of the test sequence, thereby excluding any EEG signals related to the precursor, the silent period in between, and the motor response at the end of the trial. In addition, the responses to the first and last tone pairs were excluded in order to reduce the influence of sequence onset and offset responses. For a given tone sequence for each listener, EEG data from each tone were Fourier transformed using a Fast Fourier transform. Data from all runs of a given condition were then combined for statistical analysis.

Results

Behavioral results

Subjective reports for the illusory stimulus without any modulation or cue sequence indicated that the spontaneous percept for nine of the 13 listeners was of the high tone in the right ear alternating with the low tone in the left ear (R/Hi-L/Lo). The remaining four participants reported hearing the low tone in the right ear, alternating with the high tone in the left ear (R/Lo-L/Hi). No other perceptual configuration was reported (12). For the cued modulated and unmodulated sequences, all 13 listeners reported perceiving the illusion for all the trials as predicted. For example, in the condition where the cue was L/Lo, all listeners consistently reported hearing the low tone in the left ear and the high tone in the right ear.

The behavioral results for the deviant detection task revealed high average performance (mean *d'* = 1.83), but also showed no difference in performance between the two cueing conditions [F(1,24)=2.3, p=0.2], indicating that listeners could perform the task equally well for both cued percepts (Left Low and Right Low).

EEG results

In analyzing the EEG responses, we focused on the change in the *ratio* of the amplitudes of the FFT components at the two tagged frequencies, 34.47 and 44.31 Hz. Figure 2-C indicates the natural logarithmic transform of these ratios. This is because the baseline amplitudes for the two tagged frequencies differed (Figure 3-B). Hence, the ratio of the test amplitudes indicates the relative change in amplitude due to the different test conditions. A 2-way ANOVA with Cued Ear (L/R) and Synchronous Frequency (34.47/44.31 Hz) as factors was carried out on this logarithmic transform. A significant effect of the frequency synchronous with the target was observed [F(1,12)=32.2, *p*<0.0001]. This outcome indicates that there was a difference in the amplitudes of the tagged frequencies when they were synchronous to the attended tone stream compared to the amplitudes of the tagged frequencies that were not synchronous. No significant effect of cued ear was observed [F(1,12)=0.067, p=0.8] and no significant interaction was present [F(1,12)=0.05, p=0.827]. As shown in Figure 2-C, the EEG amplitude of the tagged frequency synchronous with the cued frequency tone was higher than the tagged frequency alternating with the cued tone, irrespective of whether the cue was in the Left or Right ear.

Discussion

We found that the uncued tones that were synchronous with the cued tone sequence (but were heard as alternating with it) elicited stronger responses in the EEG, as measured through their tagged modulation frequency, than the alternating tones. This can clearly be seen from the peak amplitudes (Fig. 2B) as well as the change in ratios (Fig. 2C). There was no effect of which ear was cued, in line with previous experiments that found that the illusion can be elicited in either configuration (R/Lo heard with L/Hi or vice versa) based on the appropriate precursor sequence (22). These results provide further support for the proposal that the illusion arises from the synchronous tone pairs (either R/Lo-L/Hi or R/Hi-L/Lo) in the stimulus.

General discussion

The octave illusion is a compelling example of non-veridical auditory perception of a relatively simple repeating stimulus. As demonstrated in a previous study (22), many properties of the octave illusion, including its dependence on frequency separation and its build-up over time, are shared with auditory streaming. The current study further investigated the illusion and its potential underlying mechanisms by providing behavioral and EEG tests of which tones within the sequence contribute most to the illusion. The most interesting and unexpected aspect of the results was that the synchronous tones in the stimulus contribute to the illusory percept of alternating sound sources, showing that the illusory percept probably occurs due to a temporal misattribution of tones that were perceived in their correct physical location, rather than due to a spatial misallocation of tones that were perceived to be in their correct temporal position.

It is known that synchronous tones of different frequencies can be difficult to segregate due to the strong binding cues of temporal coherence (34,35). However, the synchronous tones in the octave illusion clearly sound as two, distinctly lateralized tone streams. We hypothesize that the specific alternating configuration of the synchronous tone pairs, presented separately to the two ears, leads to a unique competitive engagement between the two synchronous tones, causing them to separate perceptually into two streams of their individual frequencies (for example, listeners can perceive synchronous tones L/Hi and R/Lo as two perceptual streams).

The question now arises as to why the two synchronous tones (L/Hi and R/Lo) are heard as temporally misaligned? It is well known that temporal judgements between sounds belonging to different streams are inaccurate and difficult, and in fact, are commonly used as an objective measure or indicator of streaming (36,37), even when the sounds are synchronous (34,38). Furthermore, previous work on temporal order judgements of repeating sequences of short-duration (< 300 ms) stimuli (39–42) suggests it is easy to recognize the identity of the stimuli but difficult to judge their temporal order. In the context of the current illusion, we hypothesize that due to the synchronous tones falling into separate perceptual streams, it becomes difficult for listeners to judge the temporal relationships between these stimuli (38), and that because they are heard as separate, they are by default heard as alternating, in line with the onsets of the tone sequences.

To our knowledge, no current computational model of streaming can predict the outcomes of the current experiments. Such a model would have to take into account the follow key aspects of the results: 1) the illusory percept can be modified by attention, so it cannot be dependent on a hard wired, dominant ear bias; 2) the percept only occurs when the frequencies of the tone pairs are similar (for example, the illusion does not occur when R/Lo and L/Lo are different frequencies); and 3) the tones perceived as alternating tend to be the physically synchronous, rather than alternating, tone pairs.

Competing Interests:

The authors have no competing interests.

Acknowledgements:

This work was supported by a UCL Overseas Research Scholarship (AM), a UCL Graduate Research Scholarship (AM), a UCL Charlotte and Yule Bogue Research Fellowship (AM), NIH grant R01 DC005216 (AJO), NIH grant R01 DC007657 (AJO and SAS), an Army Research Office grant (SAS), an Advanced ERC grant (SAS), and NIH grant R01 DC005779 (SAS).

References

1. Deutsch D. Two-channel listening to musical scales. J Acoust Soc Am. 1975;57(5):1156–60.

2. Deutsch D, Hamaoui K, Henthorn T. The glissando illusion and handedness. Neuropsychologia. 2007;45(13):2981–8.

3. Schwartz J-L, Grimault N, Hupé J-M, Moore BCJ, Pressnitzer D. Multistability in perception: binding sensory modalities, an overview. Philos Trans R Soc B Biol Sci. 2012 Apr 5;367(1591):896–905.

4. Shepard RN, Jordan DS. Auditory Illusions Demonstrating That Tones Are Assimilated to an Internalized Musical Scale. Science. 1984 Dec 14;226(4680):1333–4.

5. Warren RM, Warren RP. Auditory illusions and confusions. WH Freeman; 1970.

6. Deutsch D. An auditory illusion. Nature. 1974 Sep 27;251(5473):307–9.

7. Zwicker T. Experimente zur dichotischen Oktav-Täuschung. Acta Acust United Acust. 1984 May 1;55(2):128–36.

8. McClurkin RH, Hall JW. Pitch and timbre in a two-tone dichotic auditory illusion. J Acoust Soc Am. 1981 Feb 1;69(2):592–4.

9. Brännström KJ, Nilsson P. Octave illusion elicited by overlapping narrowband noises. J Acoust Soc Am. 2011 May 1;129(5):3213–20.

10. Deutsch D, Roll PL. Separate "what" and "where" decision mechanisms in processing a dichotic tonal sequence. J Exp Psychol Hum Percept Perform. 1976;2(1):23–9.

11. Brancucci A, Padulo C, Tommasi L. "Octave illusion" or "Deutsch's illusion"? Psychol Res PRPF. 2009 May 1;73(3):303–7.

12. Deutsch D. The octave illusion and auditory perceptual integration. In: Hearing research and theory. 1981. p. 99–142.

13. Chambers CD, Mattingley JB, Moss SA. The octave illusion revisited: Suppression or fusion between ears? J Exp Psychol Hum Percept Perform. 2002;28(6):1288–302.

14. Chambers CD, Mattingley JB, Moss SA. Reconsidering evidence for the suppression model of the octave illusion. Psychon Bull Rev. 2004 Aug 1;11(4):642–66.

15. Deutsch D. Reply to "Reconsidering evidence for the suppression model of the octave illusion," by C. D. Chambers, J. B. Mattingley, and S. A. Moss. Psychon Bull Rev. 2004 Aug 1;11(4):667–76.

16. Ross J, Tervaniemi M, Näätänen R. Neural mechanisms of the octave illusion: electrophysiological evidence for central origin. NeuroReport. 1996;8(1):303–6.

17. Lamminmäki S, Hari R. Auditory cortex activation associated with octave illusion. Neuroreport. 2000 May 15;11(7):1469–72.

18. Lamminmäki S, Mandel A, Parkkonen L, Hari R. Binaural interaction and the octave illusion. J Acoust Soc Am. 2012 Sep 1;132(3):1747–53.

19. Brancucci A, Franciotti R, D'Anselmo A, Penna S della, Tommasi L. The Sound of Consciousness: Neural Underpinnings of Auditory Perception. J Neurosci. 2011 Nov 16;31(46):16611–8.

20. Brancucci A, Prete G, Meraglia E, di Domenico A, Lugli V, Penolazzi B, et al. Asymmetric Cortical Adaptation Effects during Alternating Auditory Stimulation. PLoS ONE. 2012 Mar 28;7(3):e34367.

21. Hari R. Activation of the Human Auditory Cortex by Various Sound Sequences : Neuromagnetic Studies. In: Williamson SJ, Hoke M, Stroink G, Kotani M, editors. Advances in Biomagnetism [Internet]. Springer US; 1989 [cited 2016 Oct 4]. p. 87–92. Available from: http://link.springer.com/chapter/10.1007/978-1-4613-0581-1_9

22. Mehta AH, Yasin I, Oxenham AJ, Shamma S. Neural correlates of attention and streaming in a perceptually multistable auditory illusion. J Acoust Soc Am. 2016;in press.

23. Brancucci A, Lugli V, Perrucci MG, Gratta CD, Tommasi L. A frontal but not parietal neural correlate of auditory consciousness. Brain Struct Funct. 2016;221(1):463–72.

24. Smith J, Hausfeld S, Power RP, Gorta A. Ambiguous musical figures and auditory streaming. Percept Psychophys. 32(5):454–64.

25. Moore BCJ, Gockel H. Factors Influencing Sequential Stream Segregation. Acta Acust United Acust. 2002 May 1;88(3):320–33.

26. Bregman AS. Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press; 1990. 800 p.

27. Bregman AS, Steiger H. Auditory streaming and vertical localization: Interdependence of “what” and “where” decisions in audition. Percept Psychophys. 1980 Nov 1;28(6):539–46.

28. Brainard DH. The Psychophysics Toolbox. Spat Vis. 1997;10(4):433–6.

29. Pelli DG. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spat Vis. 1997;10(4):437–42.

30. Gutschalk A, Micheyl C, Oxenham AJ. Neural Correlates of Auditory Perceptual Awareness under Informational Masking. PLoS Biol. 2008 Jun 10;6(6):e138.

31. Bharadwaj HM, Lee AKC, Shinn-Cunningham BG. Measuring auditory selective attention using frequency tagging. Front Integr Neurosci. 2014;8:6.

32. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. 2004 Mar 15;134(1):9–21.

33. Jung TP, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui V, et al. Removing electroencephalographic artifacts by blind source separation. Psychophysiology. 2000 Mar;37(2):163–78.

34. Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA. Temporal Coherence in the Perceptual Organization and Cortical Representation of Auditory Scenes. Neuron. 2009 Jan 29;61(2):317–29.

35. Shamma SA, Elhilali M, Micheyl C. Temporal coherence and attention in auditory scene analysis. Trends Neurosci. 2011 Mar;34(3):114–23.

36. Vliegen J, Moore BCJ, Oxenham AJ. The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. J Acoust Soc Am. 1999 Aug 1;106(2):938–45.

37. Roberts B, Glasberg BR, Moore BCJ. Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. J Acoust Soc Am. 2002 Nov 1;112(5):2074–85.

38. Micheyl C, Hunter C, Oxenham AJ. Auditory stream segregation and the perception of across-frequency synchrony. J Exp Psychol Hum Percept Perform. 2010;36(4):1029–39.

39. Warren RM. Auditory temporal discrimination by trained listeners. Cognit Psychol. 1974 Apr;6(2):237–56.

40. Warren RM, Obusek CJ. Identification of temporal order within auditory sequences. Percept Psychophys. 12(1):86–90.

41. Garner WR. The accuracy of counting repeated short tones. J Exp Psychol. 1951;41(4):310–6.

42. Norman DA. Temporal confusions and limited capacity processors. Acta Psychol (Amst). 1967;27:293–7.

Figures and figure legends

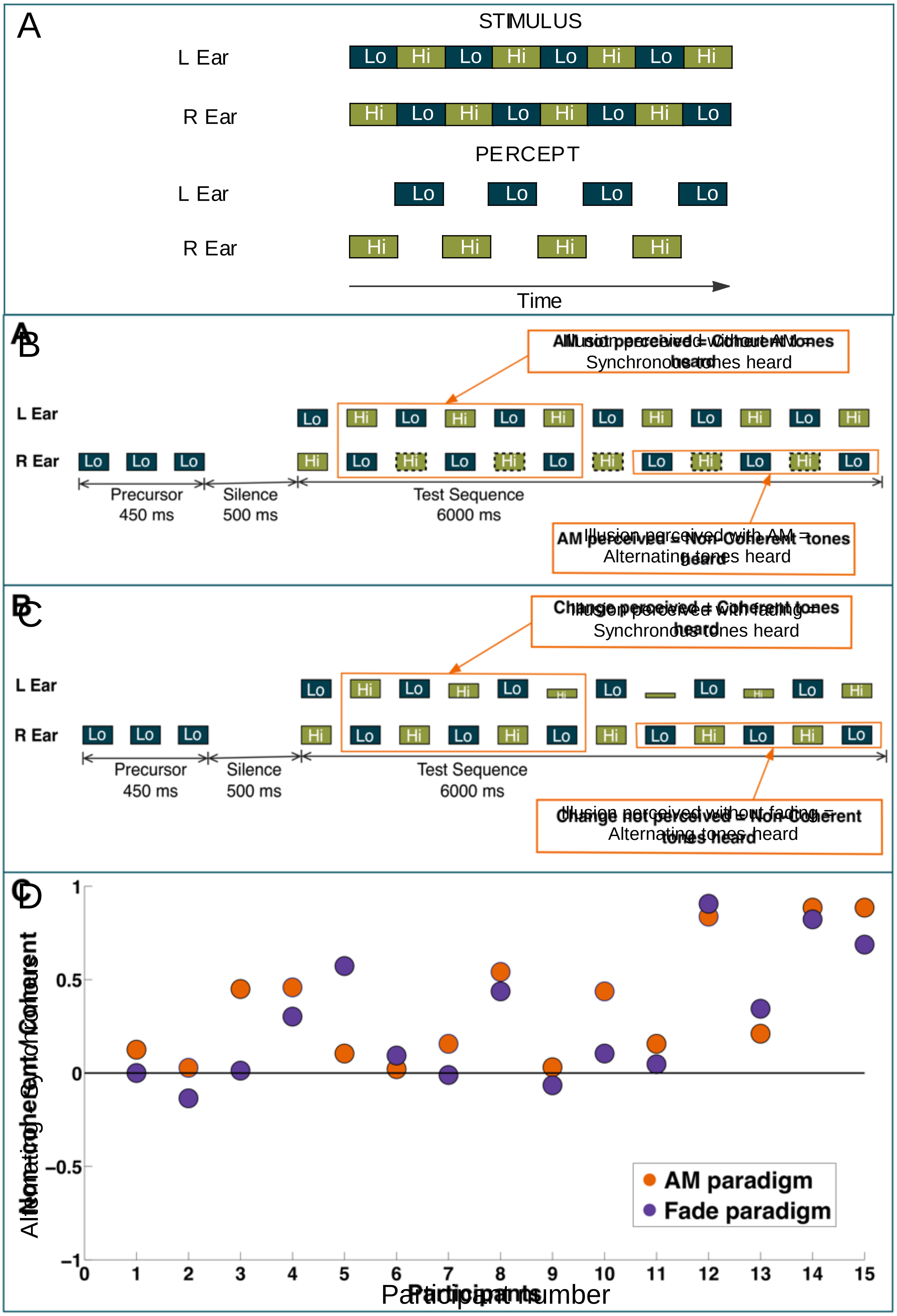
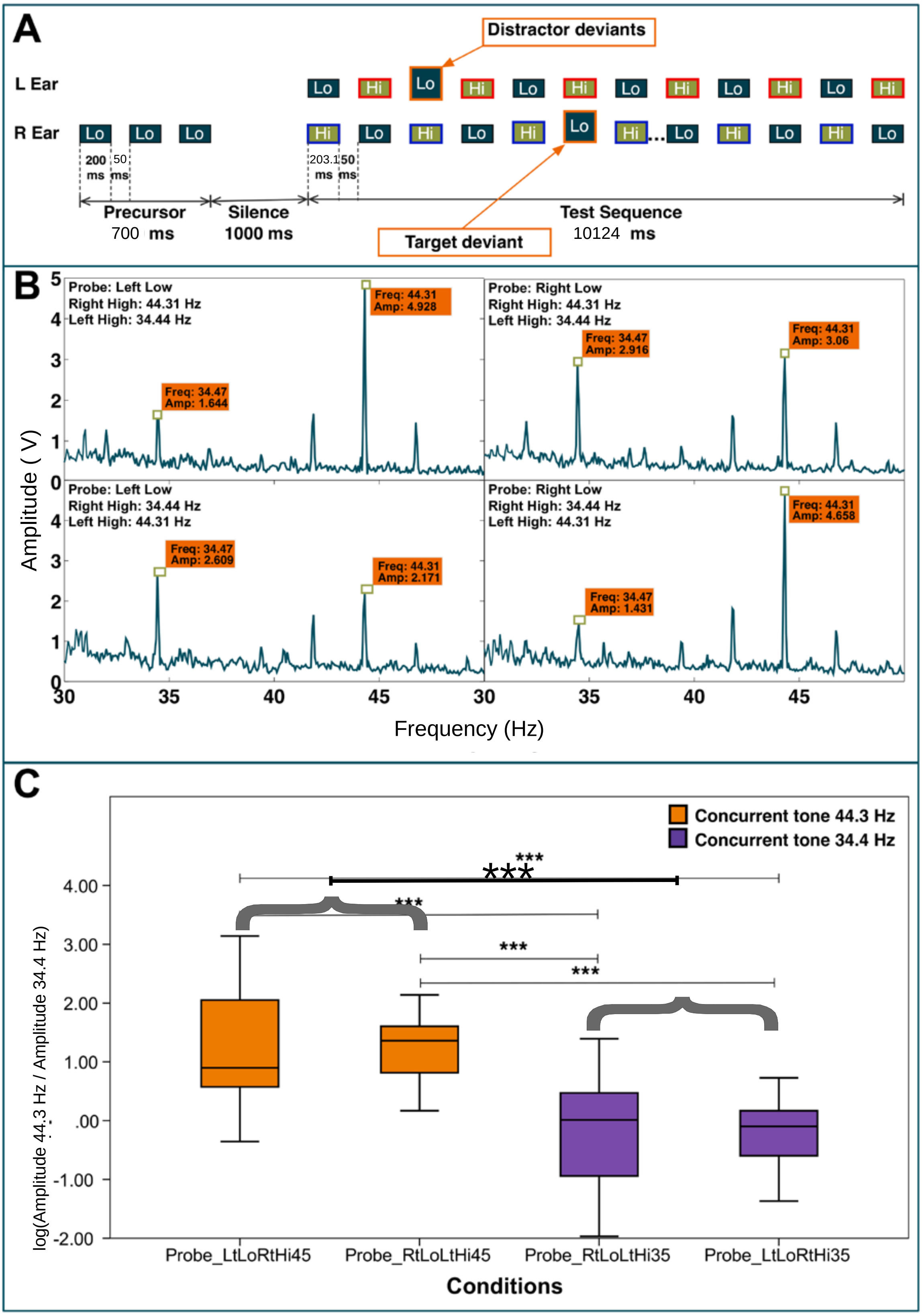

A
STIMULUS
L Ear
Lo Hi Lo Hi Lo Hi Lo Hi
R Ear
Hi Lo Hi Lo Hi Lo Hi Lo
PERCEPT
L Ear
Lo Lo Lo Lo
R Ear
Hi Hi Hi Hi
Time
B
Illusion perceived without AM = Synchronous tones heard
AM not perceived = Coherent tones heard
L Ear
R Ear
Lo Hi
Precursor 450 ms
Silence 500 ms
Test Sequence 6000 ms
Illusion perceived with AM = Alternating tones heard
AM perceived = Non-Coherent tones heard
C
Illusion perceived with fading = Synchronous tones heard
Change perceived = Coherent tones heard
L Ear
R Ear
Precursor 450 ms
Silence 500 ms
Test Sequence 6000 ms
Illusion perceived without fading = Alternating tones heard
Change not perceived = Non-Coherent tones heard
D
Normalized Coherent/Non-Coherent
Alternating-Synchronous
1
0.5
0
−0.5
−1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Participant number
Participants
● AM paradigm
● Fade paradigm

Figure 1: Stimulus and results for experiment 1. **A**. The stimulus pattern used in the original experiment of Deutsch (1974) describing the octave illusion, together with the percept most commonly obtained. Boxes labelled ‘Lo’ indicate low-frequency tones, and boxes labelled ‘Hi’ indicate high-frequency tones. **B:** Schematic diagram illustrating a sample trial of paradigm 1 for experiment 1 where all the high-frequency tones in the right ear are amplitude modulated (indicated by the dashed lines) **C:** Schematic diagram illustrating paradigm 2 for experiment 1 where some of the high-frequency tones in left ear are reduced in amplitude, indicated by the reduced height of the green (Hi) boxes. **D:** Individual results from 15 participants in both paradigms. The orange circles indicate results from the amplitude-modulated tone paradigm whereas the dark blue circles indicate the results from the fading tones paradigm. The ordinate is scaled such that the upper half of the graph (from 0 to +1) indicates when the responses corresponded more to “synchronous” tones being heard and the lower half of the graph (from 0 to -1) indicates when the responses corresponded more to “alternating” tones being heard.



548

549 **Figure 2:** Stimulus and results for experiment 2**. A:** Test stimuli example. Each ear was presented with opposing, alternating frequency sequences of pure tones (Lo = 1000 Hz with

no modulation; Hi = 2996 Hz tagged with modulation frequencies of 34.47 Hz or 44.31 Hz). Listeners were cued to focus on the low-frequency precursor on either side, as indicated by a cueing sequence, and were asked to detect target amplitude deviants. The schematic diagram below shows a sample trial where the right ear and left ear high tones are differentially tagged (red and blue outlines) and the low frequency tone cues are in the right ear. **B:** Amplitude spectrum of the EEG responses at the tagged frequencies. **C:** The amplitudes of the EEG responses at the tagged frequencies for each test condition were calculated as the natural logarithmic transform of the ratio of the amplitude of 44.31-Hz component to the amplitude of 34.47-Hz component. In conditions where the synchronous tone was tagged with 44.31 Hz, the ratio was found to be significantly higher than in the conditions where the synchronous tone was tagged with 34.47 Hz. The x-axis conditions indicate the type of cue and tagged frequency. For example, "ProbeLtLoRtHi44" indicates that the cueing sequence was a low-frequency sequence in the left ear and the high-frequency tones synchronously presented with the cued sequence, i.e. RtHi, were tagged with a 44.31-Hz tag, whereas the alternating high tones were tagged with 34.47 Hz.

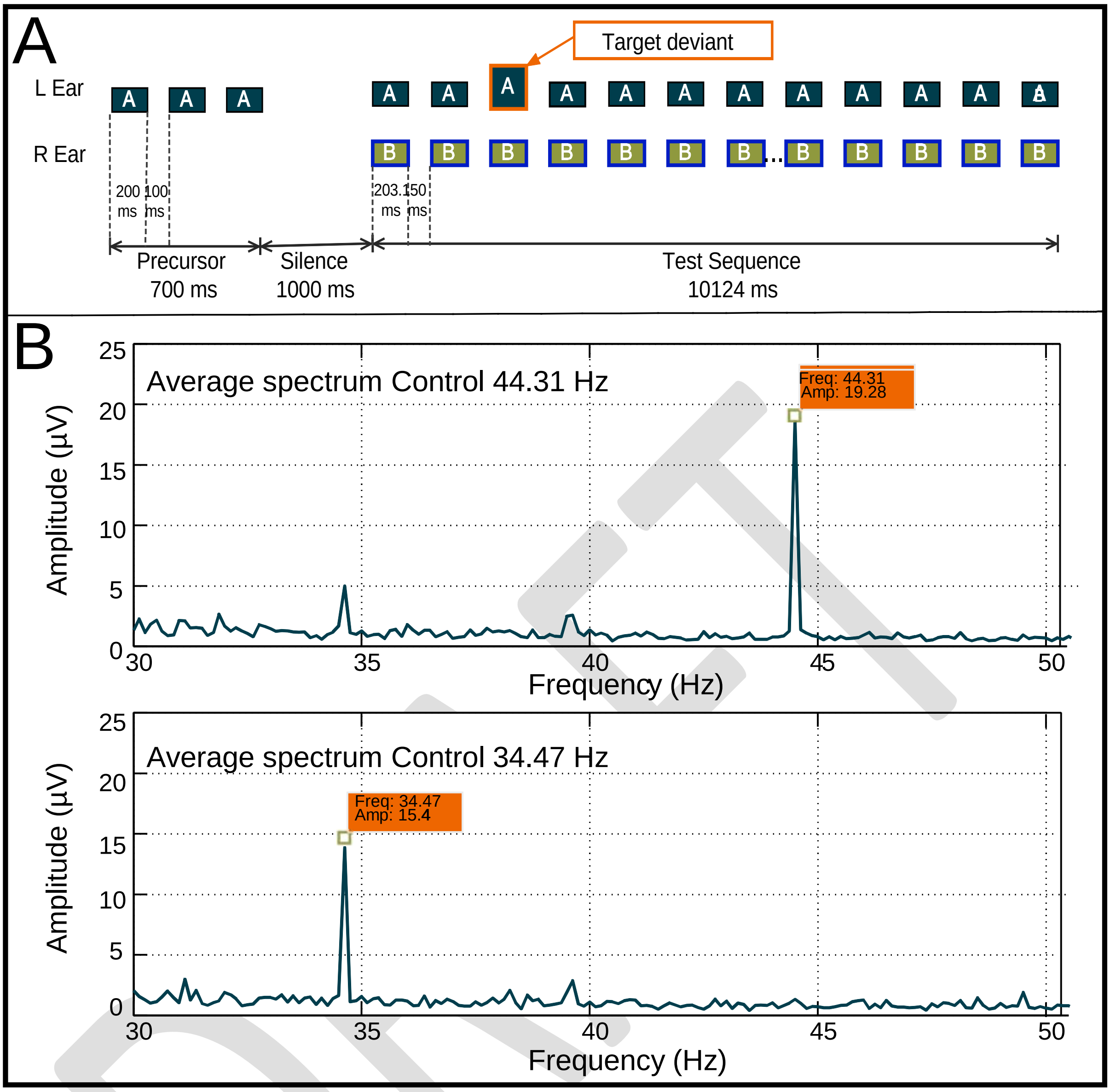


Figure 3: A. Schematic diagram of the stimuli used for the EEG control measurements. Each ear was presented with single-frequency sequences of pure tones (A = 1000 Hz with no modulation; B = 2996 Hz tagged with modulation frequencies of 34.4 Hz or 44.3 Hz). Listeners were cued to focus on the low-frequency precursor on either side, indicated by a priming sequence, and were asked to detect target amplitude deviants. The example shows a condition where the high-frequency tones in the right ear tagged (blue outlines) and the low-frequency tones were cued in the left ear. This stimulus paradigm does not elicit the illusory percept. **B.** Amplitude spectra of tagged frequencies for the control sequences. The figure

shows the raw spectra of the test signals using the two control sequences as a baseline measure. The figures indicate that the tone at 44.31 Hz evokes a larger EEG signal than the tone at 34.47 Hz.