

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

Anahita H. Mehta<sup>1</sup>, Nori Jacoby<sup>2</sup>, Ifat Yasin<sup>3</sup>, Andrew J. Oxenham<sup>4</sup>, Shihab A. Shamma<sup>5,6</sup>

<sup>1</sup>UCL Ear Institute, University College London, London, UK, WC1X 8EE

<sup>2</sup>The Center for Science & Society, Columbia University, New York, USA, NY 10027

<sup>3</sup>Department of Computer Science, University College London, London, UK, WC1E 6BT

<sup>4</sup>Department of Psychology, University of Minnesota, Minneapolis, USA, MN 55455

<sup>5</sup>Electrical and Computer Engineering Department and Institute for Systems Research, University of Maryland, College Park, Maryland, USA, MD 20742

<sup>6</sup>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

## 1 Abstract

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

17

18

19

## 20 Introduction

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

29         The stimulus used to elicit the octave illusion has been studied in different contexts and  
30 the robustness of the percept has been investigated across a variety of parameters. It has been  
31 demonstrated that the percept of this illusion is robust to changes in tone duration (7) and  
32 spectral shape (8), and can also be elicited by quasi-periodic stimuli like band-pass noise (9).  
33 It was also noted by Deutsch and Roll (10), and later confirmed by Brancucci *et al.* (11), that  
34 the illusion is not dependent on the tones being in an exact octave relationship. Indeed,  
35 Brancucci *et al.* (11) reported that the illusory percept was present for all musical intervals  
36 tested that were larger than a perfect fourth (roughly a ratio of 4:3 or a frequency difference of  
37 33%). Despite the fact that it is not dependent on the octave relationship, we continue to refer  
38 to the phenomenon as the “octave illusion” for historical reasons.

39         To explain the illusion, Deutsch (1) proposed a dual-mechanism model that consists of  
40 one mechanism for pitch determination and another for sound localization. The outputs of these  
41 mechanisms converge to elicit the illusory percept. The model is based on the assumption that  
42 the perceived pitch corresponds to the frequency of the tone presented to the listeners’  
43 “dominant” ear (usually the right), whereas the perceived location of the tone corresponds to  
44 the location of the higher-frequency tone (10), so that the final illusory percept is a combination

45 of the output of the two mechanisms, in a feature-combination operation (12). Although some  
46 authors have questioned this interpretation (13,14), the most recent studies have verified the  
47 basic observations and interpretations of the illusion (12,15).

48 A number of neuroimaging studies have been carried out using stimuli related to the  
49 octave illusion (16–20). Lamminmäki and Hari (17) aimed to find the neurophysiological basis  
50 of the ‘where’ mechanism of Deutsch’s dual-mechanism model. The stimuli were 400- and  
51 800-Hz pure tones presented to the left (L) or right (R) ears as follows: L400/R400, L400/R800,  
52 L800/R400 and L800/R800. The aim of their study was to find out whether the lateralization  
53 of the auditory evoked fields using MEG, in particular the N100m peak, co-varied with the  
54 sound localization percept. They found that the N100m was stronger in the hemisphere  
55 contralateral to the high-pitch sound, in agreement with the established finding that monaural  
56 sounds evoke stronger N100m responses in the hemisphere contralateral to the sound (21).  
57 However, the MEG measurements were not carried out on the stimulus eliciting the octave  
58 illusion itself, and no attempt was made to relate the neural responses to perception, as the  
59 measurements were made with listeners in a passive role, with no task, and no indication as to  
60 what the listeners perceived on a trial-by-trial basis. Lamminmäki et al. (18) next investigated  
61 the neuromagnetic correlates of the “where” aspect of the dual-mechanism model using  
62 frequency-tagged stimuli. Each tone in the stimuli was modulated using a unique ‘tagging’  
63 frequency that helps parse out the corresponding neuromagnetic activity for each tone. They  
64 found evidence for binaural suppression and right ear dominance for all their stimuli and  
65 concluded that the findings of their study were in line with the dual-mechanism model. Again,  
66 however, the authors used a passive paradigm, with no subjective or objective measures of  
67 perception or attention, and the stimuli were limited to isolated dichotic tone pairs, rather than  
68 illusion-inducing sequences. Several other studies have used the illusion to study aspects of the  
69 neural correlates of consciousness, by taking advantage of the fact that the same stimulus can

70 spontaneously elicit different percepts in different listeners and across different repetitions  
71 (20,22,23).

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

92

93

94

## 95 Experiment 1

96 **Rationale**

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

102

103 **Method**104 *Participants*

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

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

117

118 ***Paradigm 1: Stimuli and procedures***

119 Participants were cued, using a precursor sequence (see Figure 1-B), to attend to one of  
120 the four tones within the main sequence. The precursor sequence consisted of three low- or  
121 high-frequency tones presented either to the left or right ear prior to the main sequence, in order  
122 to indicate the side and frequency to which participants should attend. The side and frequency  
123 of the precursor tones were selected at random with equal *a priori* probability on each trial.  
124 Following a silent interval of 500 ms, the main sequence of each trial began, as shown in Figure  
125 1-B, with alternating low (1000-Hz) and high (2996-Hz) tones, marked Lo and Hi, respectively.  
126 A frequency separation larger than an octave was used because this has been shown to be  
127 effective in inducing the illusion (11) and it avoids some potentially confounding influences of  
128 using an exact octave (27). Each tone was 100 ms in duration, including 10-ms raised-cosine  
129 onset and offset ramps, and tones were separated by 50-ms silent intervals. All tones were  
130 presented at 65 dB SPL. The sequence was presented for a total of 6 s (20 repetitions of the  
131 alternating synchronous tones as seen in Figure 1-B). During the main sequence of each trial,  
132 the tones in one of the two tone sequences at the uncued frequency were sinusoidally amplitude  
133 modulated at a rate of 34.47 Hz and with a depth of 75%. For example, in Figure 1-B, the low  
134 tones in the right ear are cued, and the high tones that alternate with the cued tones are  
135 amplitude modulated. The modulation was randomly assigned on each trial to the tones that  
136 were either synchronous or alternating with the cued tones with equal *a priori* probability. For  
137 example, on a trial where the precursor tones were low tones in the right ear, the modulated  
138 tones could either be the alternating high tones in the right ear or the synchronous high tones  
139 in the left ear.

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

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

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

160

161 ***Paradigm 2: Stimuli and procedures***

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

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

181

182 **Results**

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

197           Individual results from the 15 participants, averaged across the four conditions for each  
198 of the two paradigms, are shown in Figure 1-D. Most responses were positive, indicating that  
199 changes were heard more clearly when they occurred simultaneously with, and in the opposite  
200 ear to, the cued tone. A one-sample t-test confirmed that the mean scores for both paradigms  
201 were significantly greater than zero [Paradigm 1:  $t(14) = 4.36$ ,  $p<0.001$ ; Paradigm 2:  $t(14) =$   
202  $3.13$ ,  $p<0.001$ ].

203

## 204 **Discussion**

205           The results from both paradigms were consistent in suggesting that listeners’ perception of the  
206 alternating tone-sequence in the non-cued ear corresponded to the tones in the non-cued ear

207 that were synchronous with the cued tones and *not* to the alternating tones in the cued ear, as  
208 has been previously assumed. This surprising result suggests that it is a perceptual temporal  
209 misalignment between the synchronous tones that is responsible for the perception of  
210 “alternating” tones, rather than a spatial misattribution of the alternating tones in the same ear  
211 as the cue tones, as has generally been assumed. The fundamental question of which tones  
212 contribute to the perception of the illusion has been studied in several contexts indirectly  
213 (11,13,16) and directly by Deutsch and Roll (10). However, the paradigm used by Deutsch and  
214 Roll to study this question did not elicit the octave illusion itself, which makes the interpretation  
215 of their results less clear. Experiment 2 followed up on this surprising finding, by combining  
216 a further perceptual test with EEG correlates of the illusion.

217

## 218 Experiment 2

### 219 **Rationale**

220 The aim of this experiment was to provide a further test of the surprising conclusion of  
221 Experiment 1 that the tones forming part of the illusion were the ones that were synchronous  
222 with the target tones, and not, as previously believed, the tones that were alternating with the  
223 target tones. In this experiment, EEG was combined with behavior, and the tones of the illusory  
224 stimulus were differentially tagged via amplitude modulation to obtain a direct measure of  
225 which tones were most prominent neurally, and hence most likely to be perceptually salient  
226 (18,30,31).

227 The different tones within each sequence were amplitude modulated at different rates,  
228 in order to identify their responses in the EEG signal. The hypothesis of this experiment was  
229 that the modulation rate corresponding to the contralateral tones synchronous with the cued

230 tones would show an increase in amplitude, relative to the tones that were alternating with the  
231 cued tones. For example, if the listener were cued to the low tones in the right ear, then the  
232 neural response to the modulation frequency of the synchronous high tones in the left ear should  
233 be larger than the neural response to the modulation frequency of the high tones in the right  
234 ear.

235

## 236 **Participants**

237 Thirteen listeners (six male and seven female, aged 21-30 years) participated in  
238 experiment 2. All listeners were naïve and had not taken part in any other related experiments.  
239 All participant recruitment procedures and inclusion criteria were the same as for Experiment  
240 1.

241

## 242 **Stimuli and procedures**

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

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

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

270 The total EEG stimulus set was counterbalanced for the cued ear and the tagging  
271 modulation rate by dividing the set into four conditions. In conditions 1 and 2, listeners were  
272 cued to the low-frequency tones in the left and right ear, respectively, while the high-frequency  
273 tones in the left ear were modulated at 34.47 Hz, and the high-frequency tones in the right ear  
274 were modulated at 44.31 Hz. In conditions 3 and 4, listeners were cued to the low-frequency  
275 tones in the left and right ear, respectively, while the high-frequency tones in the left ear were

276 modulated at 44.31 Hz and the high-frequency tones in the right ear were modulated at 34.47  
277 Hz. Two control conditions (conditions 5 and 6) were included to establish a baseline for the  
278 tagged frequencies. The control stimuli had only low-frequency unmodulated tones in one ear  
279 and only high-frequency modulated tones presented synchronously in the opposite ear ( $L_o =$   
280 1000 Hz with no modulation;  $H_i = 2996$  Hz tagged with modulation frequencies of 34.47 Hz  
281 or 44.31 Hz) with the same parameters as in conditions 1-4. All tones in the main sequence  
282 were also 203.1 ms long and were separated by 50-ms silent gaps (Figure 3-A). Listeners were  
283 cued by a low-frequency tone sequence on either side and were asked to indicate whether  
284 amplitude deviants in the cued stream were present or absent (same as conditions 1-4). The  
285 control stimuli did not elicit the octave illusion; their purpose was to establish a baseline for  
286 the EEG amplitude of the tagged frequencies.

287 The EEG measurements were preceded by a series of behavioral tests. In the first block  
288 of ten trials, listeners heard the illusory sequence with no precursor tones and no modulation.  
289 For each trial, their unbiased percept (i.e., when they were not provided with instructions on  
290 what to attend to within the sound sequences) was noted. For this, the participants were asked  
291 to simply listen to the sound sequence and report what they heard. The subjective percepts were  
292 collected as free responses. Participants were not informed of what the expected percept was.  
293 Next, listeners were presented with another block of ten trials, where their perceptual responses  
294 to the stimulus with low-frequency pure tones and high-frequency modulated tones were  
295 recorded. Finally, listeners were presented with a block of ten trials in which the full stimulus  
296 was presented (precursor plus main sequence, as in the EEG experiment). Half the trials had  
297 the cue presented on the left, and the other half had the cue presented on the right. Again,  
298 listeners were asked to report their percepts. For all three blocks of trials, the listeners were  
299 naïve to the stimuli and were not told what the expected response was.

300 In the main EEG portion of the experiment, the stimuli were presented in either ‘test’  
301 blocks (conditions 1-4) or ‘control’ blocks (conditions 5-6). Within each of the blocks, the  
302 trials were randomized for cueing sequence type (cues could be low tones in the Right or Left  
303 ear) and tagging frequency. Each block included 120 trials and each listener was tested using  
304 4 test blocks and 2 control blocks. Hence, 480 test trials and 240 control trials were conducted  
305 for each listener – 120 per condition. For each trial, the listeners were asked to focus on the  
306 cued stream (as determined by the precursor). At the end of each trial, the listener had to report  
307 via a button press if a target deviant was present or absent. The next trial was initiated 1 s after  
308 the response.

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

314

## 315 **Data Analyses**

### 316 *Behavioral data analyses*

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

321

### 322 *EEG analyses*

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

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

340

## 341 **Results**

### 342 *Behavioral results*

343 Subjective reports for the illusory stimulus without any modulation or cue sequence  
344 indicated that the spontaneous percept for nine of the 13 listeners was of the high tone in the  
345 right ear alternating with the low tone in the left ear (R/Hi-L/Lo). The remaining four

346 participants reported hearing the low tone in the right ear, alternating with the high tone in the  
347 left ear (R/Lo-L/Hi). No other perceptual configuration was reported (12). For the cued  
348 modulated and unmodulated sequences, all 13 listeners reported perceiving the illusion for all  
349 the trials as predicted. For example, in the condition where the cue was L/Lo, all listeners  
350 consistently reported hearing the low tone in the left ear and the high tone in the right ear.

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

355

### 356 *EEG results*

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

370 frequency alternating with the cued tone, irrespective of whether the cue was in the Left or  
371 Right ear.

372

### 373 **Discussion**

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

382

### 383 **General discussion**

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

393 than due to a spatial misallocation of tones that were perceived to be in their correct temporal  
394 position.

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

402 The question now arises as to why the two synchronous tones (L/Hi and R/Lo) are heard  
403 as temporally misaligned? It is well known that temporal judgements between sounds  
404 belonging to different streams are inaccurate and difficult, and in fact, are commonly used as  
405 an objective measure or indicator of streaming (36,37), even when the sounds are synchronous  
406 (34,38). Furthermore, previous work on temporal order judgements of repeating sequences of  
407 short-duration (< 300 ms) stimuli (39–42) suggests it is easy to recognize the identity of the  
408 stimuli but difficult to judge their temporal order. In the context of the current illusion, we  
409 hypothesize that due to the synchronous tones falling into separate perceptual streams, it  
410 becomes difficult for listeners to judge the temporal relationships between these stimuli (38),  
411 and that because they are heard as separate, they are by default heard as alternating, in line with  
412 the onsets of the tone sequences.

413 To our knowledge, no current computational model of streaming can predict the  
414 outcomes of the current experiments. Such a model would have to take into account the follow  
415 key aspects of the results: 1) the illusory percept can be modified by attention, so it cannot be  
416 dependent on a hard wired, dominant ear bias; 2) the percept only occurs when the frequencies

417 of the tone pairs are similar (for example, the illusion does not occur when R/Lo and L/Lo are  
418 different frequencies); and 3) the tones perceived as alternating tend to be the physically  
419 synchronous, rather than alternating, tone pairs.

420

#### 421 **Competing Interests:**

422 The authors have no competing interests.

#### 423 **Acknowledgements:**

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

429

#### 430 **References**

- 431 1. Deutsch D. Two-channel listening to musical scales. *J Acoust Soc Am*.  
432 1975;57(5):1156–60.
- 433 2. Deutsch D, Hamaoui K, Henthorn T. The glissando illusion and handedness.  
434 *Neuropsychologia*. 2007;45(13):2981–8.
- 435 3. Schwartz J-L, Grimault N, Hupé J-M, Moore BCJ, Pressnitzer D. Multistability in  
436 perception: binding sensory modalities, an overview. *Philos Trans R Soc B Biol Sci*.  
437 2012 Apr 5;367(1591):896–905.
- 438 4. Shepard RN, Jordan DS. Auditory Illusions Demonstrating That Tones Are  
439 Assimilated to an Internalized Musical Scale. *Science*. 1984 Dec  
440 14;226(4680):1333–4.
- 441 5. Warren RM, Warren RP. Auditory illusions and confusions. WH Freeman; 1970.
- 442 6. Deutsch D. An auditory illusion. *Nature*. 1974 Sep 27;251(5473):307–9.

- 443 7. Zwicker T. Experimente zur dichotischen Oktav-Täuschung. *Acta Acust United*  
444 *Acust.* 1984 May 1;55(2):128–36.
- 445 8. McClurkin RH, Hall JW. Pitch and timbre in a two-tone dichotic auditory illusion. *J*  
446 *Acoust Soc Am.* 1981 Feb 1;69(2):592–4.
- 447 9. Brännström KJ, Nilsson P. Octave illusion elicited by overlapping narrowband  
448 noises. *J Acoust Soc Am.* 2011 May 1;129(5):3213–20.
- 449 10. Deutsch D, Roll PL. Separate “what” and “where” decision mechanisms in  
450 processing a dichotic tonal sequence. *J Exp Psychol Hum Percept Perform.*  
451 1976;2(1):23–9.
- 452 11. Brancucci A, Padulo C, Tommasi L. “Octave illusion” or “Deutsch’s illusion”? *Psychol*  
453 *Res PRPF.* 2009 May 1;73(3):303–7.
- 454 12. Deutsch D. The octave illusion and auditory perceptual integration. In: *Hearing*  
455 *research and theory.* 1981. p. 99–142.
- 456 13. Chambers CD, Mattingley JB, Moss SA. The octave illusion revisited: Suppression or  
457 fusion between ears? *J Exp Psychol Hum Percept Perform.* 2002;28(6):1288–302.
- 458 14. Chambers CD, Mattingley JB, Moss SA. Reconsidering evidence for the suppression  
459 model of the octave illusion. *Psychon Bull Rev.* 2004 Aug 1;11(4):642–66.
- 460 15. Deutsch D. Reply to “Reconsidering evidence for the suppression model of the  
461 octave illusion,” by C. D. Chambers, J. B. Mattingley, and S. A. Moss. *Psychon Bull*  
462 *Rev.* 2004 Aug 1;11(4):667–76.
- 463 16. Ross J, Tervaniemi M, Näätänen R. Neural mechanisms of the octave illusion:  
464 electrophysiological evidence for central origin. *NeuroReport.* 1996;8(1):303–6.
- 465 17. Lamminmäki S, Hari R. Auditory cortex activation associated with octave illusion.  
466 *Neuroreport.* 2000 May 15;11(7):1469–72.
- 467 18. Lamminmäki S, Mandel A, Parkkonen L, Hari R. Binaural interaction and the octave  
468 illusion. *J Acoust Soc Am.* 2012 Sep 1;132(3):1747–53.
- 469 19. Brancucci A, Franciotti R, D’Anselmo A, Penna S della, Tommasi L. The Sound of  
470 Consciousness: Neural Underpinnings of Auditory Perception. *J Neurosci.* 2011 Nov  
471 16;31(46):16611–8.
- 472 20. Brancucci A, Prete G, Meraglia E, di Domenico A, Lugli V, Penolazzi B, et al.  
473 Asymmetric Cortical Adaptation Effects during Alternating Auditory Stimulation.  
474 *PLoS ONE.* 2012 Mar 28;7(3):e34367.
- 475 21. Hari R. Activation of the Human Auditory Cortex by Various Sound Sequences :  
476 Neuromagnetic Studies. In: Williamson SJ, Hoke M, Stroink G, Kotani M, editors.  
477 *Advances in Biomagnetism [Internet].* Springer US; 1989 [cited 2016 Oct 4]. p. 87–  
478 92. Available from: [http://link.springer.com/chapter/10.1007/978-1-4613-0581-](http://link.springer.com/chapter/10.1007/978-1-4613-0581-1_9)  
479 [1\\_9](http://link.springer.com/chapter/10.1007/978-1-4613-0581-1_9)

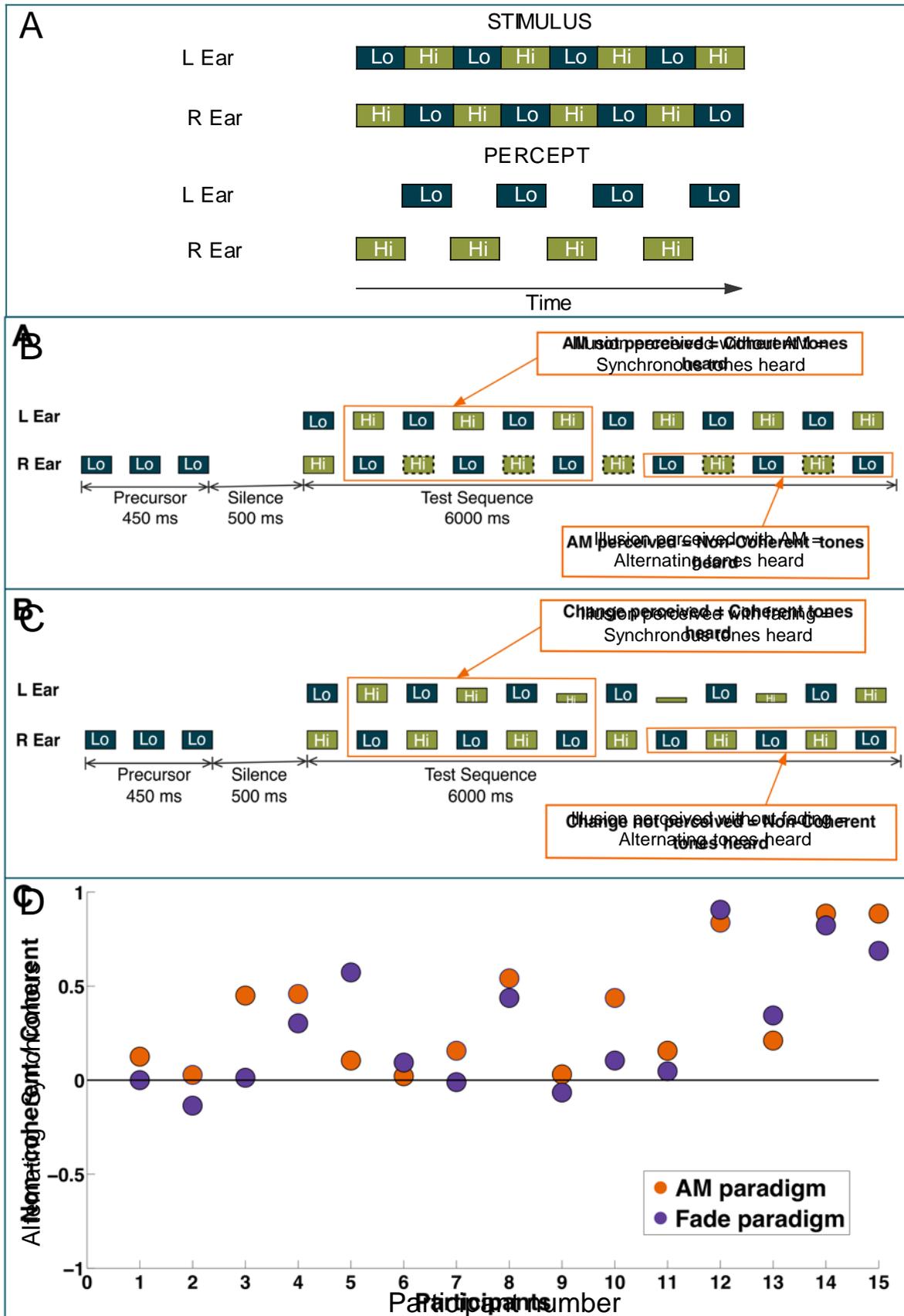
- 480 22. Mehta AH, Yasin I, Oxenham AJ, Shamma S. Neural correlates of attention and  
481 streaming in a perceptually multistable auditory illusion. *J Acoust Soc Am.* 2016;in  
482 press.
- 483 23. Brancucci A, Lugli V, Perrucci MG, Gratta CD, Tommasi L. A frontal but not parietal  
484 neural correlate of auditory consciousness. *Brain Struct Funct.* 2016;221(1):463–  
485 72.
- 486 24. Smith J, Hausfeld S, Power RP, Gorta A. Ambiguous musical figures and auditory  
487 streaming. *Percept Psychophys.* 32(5):454–64.
- 488 25. Moore BCJ, Gockel H. Factors Influencing Sequential Stream Segregation. *Acta*  
489 *Acust United Acust.* 2002 May 1;88(3):320–33.
- 490 26. Bregman AS. *Auditory Scene Analysis: The Perceptual Organization of Sound.* MIT  
491 Press; 1990. 800 p.
- 492 27. Bregman AS, Steiger H. Auditory streaming and vertical localization:  
493 Interdependence of “what” and “where” decisions in audition. *Percept Psychophys.*  
494 1980 Nov 1;28(6):539–46.
- 495 28. Brainard DH. The Psychophysics Toolbox. *Spat Vis.* 1997;10(4):433–6.
- 496 29. Pelli DG. The VideoToolbox software for visual psychophysics: Transforming  
497 numbers into movies. *Spat Vis.* 1997;10(4):437–42.
- 498 30. Gutschalk A, Micheyl C, Oxenham AJ. Neural Correlates of Auditory Perceptual  
499 Awareness under Informational Masking. *PLoS Biol.* 2008 Jun 10;6(6):e138.
- 500 31. Bharadwaj HM, Lee AKC, Shinn-Cunningham BG. Measuring auditory selective  
501 attention using frequency tagging. *Front Integr Neurosci.* 2014;8:6.
- 502 32. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial  
503 EEG dynamics including independent component analysis. *J Neurosci Methods.*  
504 2004 Mar 15;134(1):9–21.
- 505 33. Jung TP, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui V, et al. Removing  
506 electroencephalographic artifacts by blind source separation. *Psychophysiology.*  
507 2000 Mar;37(2):163–78.
- 508 34. Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA. Temporal Coherence in the  
509 Perceptual Organization and Cortical Representation of Auditory Scenes. *Neuron.*  
510 2009 Jan 29;61(2):317–29.
- 511 35. Shamma SA, Elhilali M, Micheyl C. Temporal coherence and attention in auditory  
512 scene analysis. *Trends Neurosci.* 2011 Mar;34(3):114–23.
- 513 36. Vliegen J, Moore BCJ, Oxenham AJ. The role of spectral and periodicity cues in  
514 auditory stream segregation, measured using a temporal discrimination task. *J*  
515 *Acoust Soc Am.* 1999 Aug 1;106(2):938–45.

- 516 37. Roberts B, Glasberg BR, Moore BCJ. Primitive stream segregation of tone sequences  
517 without differences in fundamental frequency or passband. *J Acoust Soc Am.* 2002  
518 Nov 1;112(5):2074–85.
- 519 38. Micheyl C, Hunter C, Oxenham AJ. Auditory stream segregation and the perception  
520 of across-frequency synchrony. *J Exp Psychol Hum Percept Perform.*  
521 2010;36(4):1029–39.
- 522 39. Warren RM. Auditory temporal discrimination by trained listeners. *Cognit Psychol.*  
523 1974 Apr;6(2):237–56.
- 524 40. Warren RM, Obusek CJ. Identification of temporal order within auditory sequences.  
525 *Percept Psychophys.* 12(1):86–90.
- 526 41. Garner WR. The accuracy of counting repeated short tones. *J Exp Psychol.*  
527 1951;41(4):310–6.
- 528 42. Norman DA. Temporal confusions and limited capacity processors. *Acta Psychol*  
529 (Amst). 1967;27:293–7.

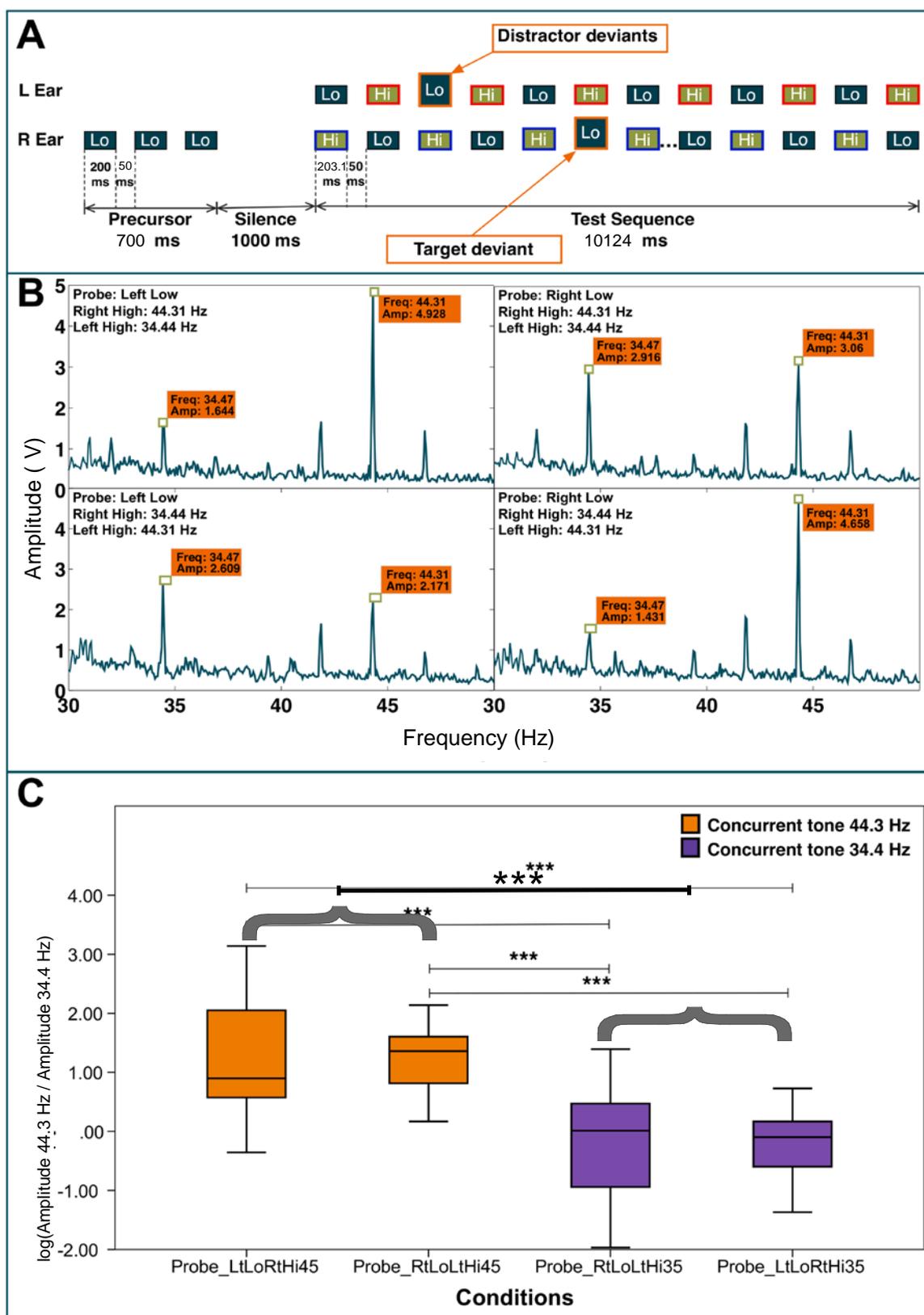
530

531

532 Figures and figure legends



534 **Figure 1:** Stimulus and results for experiment 1. **A:** The stimulus pattern used in the original  
535 experiment of Deutsch (1974) describing the octave illusion, together with the percept most  
536 commonly obtained. Boxes labelled ‘Lo’ indicate low-frequency tones, and boxes labelled ‘Hi’  
537 indicate high-frequency tones. **B:** Schematic diagram illustrating a sample trial of paradigm 1  
538 for experiment 1 where all the high-frequency tones in the right ear are amplitude modulated  
539 (indicated by the dashed lines) **C:** Schematic diagram illustrating paradigm 2 for experiment 1  
540 where some of the high-frequency tones in left ear are reduced in amplitude, indicated by the  
541 reduced height of the green (Hi) boxes. **D:** Individual results from 15 participants in both  
542 paradigms. The orange circles indicate results from the amplitude-modulated tone paradigm  
543 whereas the dark blue circles indicate the results from the fading tones paradigm. The ordinate  
544 is scaled such that the upper half of the graph (from 0 to +1) indicates when the responses  
545 corresponded more to “synchronous” tones being heard and the lower half of the graph (from  
546 0 to -1) indicates when the responses corresponded more to “alternating” tones being heard.  
547



548

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

551 no modulation; Hi = 2996 Hz tagged with modulation frequencies of 34.47 Hz or 44.31 Hz).  
552 Listeners were cued to focus on the low-frequency precursor on either side, as indicated by a  
553 cueing sequence, and were asked to detect target amplitude deviants. The schematic diagram  
554 below shows a sample trial where the right ear and left ear high tones are differentially tagged  
555 (red and blue outlines) and the low frequency tone cues are in the right ear. **B:** Amplitude  
556 spectrum of the EEG responses at the tagged frequencies. **C:** The amplitudes of the EEG  
557 responses at the tagged frequencies for each test condition were calculated as the natural  
558 logarithmic transform of the ratio of the amplitude of 44.31-Hz component to the amplitude of  
559 34.47-Hz component. In conditions where the synchronous tone was tagged with 44.31 Hz, the  
560 ratio was found to be significantly higher than in the conditions where the synchronous tone  
561 was tagged with 34.47 Hz. The x-axis conditions indicate the type of cue and tagged frequency.  
562 For example, “ProbeLtLoRtHi44” indicates that the cueing sequence was a low-frequency  
563 sequence in the left ear and the high-frequency tones synchronously presented with the cued  
564 sequence, i.e. RtHi, were tagged with a 44.31-Hz tag, whereas the alternating high tones were  
565 tagged with 34.47 Hz.

566



576 shows the raw spectra of the test signals using the two control sequences as a baseline  
577 measure. The figures indicate that the tone at 44.31 Hz evokes a larger EEG signal than the  
578 tone at 34.47 Hz.  
579

DRAFT