

1 **Effects of gamification and active listening on short-term sound localization training in**
2 **virtual reality.**

3 Mark A. Steadman^{1a}, Chung Eun Kim¹, Jean-Hugues Lestang², Dan F. M. Goodman² and
4 Lorenzo Picinali¹

5 ¹ Dyson School of Design Engineering, Imperial College London, South Kensington Campus,
6 London SW7 2AZ, UK

7 ² Department of Electrical and Electronic Engineering, Imperial College London, South
8 Kensington Campus, London SW7 2AZ, UK

9 (Dated 2nd October 2017)

10 Running title: Sound localization in virtual reality

^a Current address: Department of Bioengineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

Sound localization in Virtual Reality

11 Headphone-based virtual audio systems typically use non-individualized head-related transfer
12 functions (HRTFs) to create the illusion of spatialized sound. Listeners are therefore provided with
13 unfamiliar spatial cues leading to poor sound localization. In this study, a smartphone-based
14 system was developed to investigate the effects of short-term training on virtual sound localization
15 accuracy. Participants underwent multiple training sessions in which visual positional feedback
16 was provided in a virtual environment, interleaved with localization accuracy evaluation sessions.
17 Different versions of the training software were developed to investigate the effects of introducing
18 game-design elements ('gamification') and relative sound source motion using head tracking
19 ('active listening') on improvements in localization accuracy. The results demonstrate that
20 adaptation to a non-individualized HRTF can be facilitated using a small number of short (12
21 minute) training sessions, and is retained across multiple days. This adaptation is not HRTF-
22 specific, as the learning effect generalizes to a second HRTF not used in the training, regardless of
23 the training paradigm used. The introduction of game-design elements and the use of active
24 listening had no significant effect on the efficacy of localization training.

25 I. INTRODUCTION

26 Binaural 3D sound systems aim to accurately reproduce the waveform at the listener's eardrum
27 that would normally be produced by an external sound source. This is generally achieved by
28 filtering a sound using the head-related transfer function (HRTF). In practice, such systems often
29 make many compromises. For example, the impulse responses that comprise the HRTF are
30 measured at several discrete locations for each ear, and must be interpolated to produce an estimate
31 of the complete transfer function.

32 The HRTF also depends on idiosyncratic physical characteristics of the listener. For example,
33 the size of the head alters the interaural time difference (ITD) for a given sound source location.
34 Therefore, to accurately reproduce waveforms at each ear the HRTF for a specific listener would
35 need to be known. Although some work has been done on estimating the HRTF from easily
36 obtainable anthropometric data (e.g. Kahana *et al.*, 1999; Katz, 2001), this often involves making
37 simplifications in order to make numerical calculations tractable. The most accurate estimations
38 require the use of specialized equipment, meaning that consumer-oriented systems must use
39 generic, non-individualized HRTFs such as those measured from artificial anthropometric models
40 (e.g. Gardner and Martin, 1995).

41 It is generally thought that the differences between an individual's true HRTF and these
42 generic HRTFs have a detrimental effect on the perceived realism of virtual sound sources. It has
43 been noted, for example, that the listeners are able to localize a sound that is spatialized using their
44 own HRTF with a similar accuracy to free field listening, albeit with poorer elevation judgments
45 and increased front-back confusions (Morimoto and Aokata, 1984; Wightman and Kistler, 1989).
46 These errors are typically exacerbated where non-individualized HRTFs are used (Wenzel *et al.*,

Sound localization in Virtual Reality

47 1993; Møller *et al.*, 1996). Furthermore, it has been suggested that the use of non-individualized
48 HRTFs results in an auditory perception with reduced ‘presence’ (Väljamäe *et al.*, 2004).

49 It seems that achieving accurate perceptions of virtual auditory sources is limited by the
50 similarity of the listener’s HRTF and the generic HRTF used in a given binaural 3D sound system.
51 Indeed, efforts have been made to ‘match’ listeners to a best-fitting HRTF from a database (Katz
52 and Parseihian, 2012). Whilst this is a promising approach, it does not take advantage of the brain’s
53 ability to adapt to changes in sensory input. There is increasing evidence of the adult brain being
54 more plastic than classically thought (e.g. Fuchs and Flügge, 2014). It has been demonstrated that
55 this plasticity can result in a decrease in sound localization error over time when the HRTF is
56 altered by physically altering the shape of the ears using molds (Hofman *et al.*, 1998; Van Wanrooij
57 and Van Opstal, 2005; Carlile and Blackman, 2014). However, this process occurs through passive
58 learning over the course of several days or weeks.

59 This timescale is likely to be impractical for a consumer-oriented system, in which it will
60 generally be undesirable to rely on adaptation periods longer than a few hours. The possibility of
61 accelerating this process has therefore received some interest. Several studies have demonstrated
62 that active learning through positional feedback has the potential to achieve adaptation over such
63 timescales (Zahorik *et al.*, 2006; Parseihian and Katz, 2012; Mendonça *et al.*, 2012). However, it
64 is not clear from the available evidence to what extent adaptation is complete at the end of such
65 short training sessions and whether there is room for further improvement. Also, it is not clear
66 whether improvements in sound localization performance reported in these studies are driven by
67 adaptation to a specific HRTF, or something more general.

Sound localization in Virtual Reality

68 One line of inquiry that could help to optimize this perceptual learning process is the use of
69 ‘gamification’, whereby popular game design elements are utilized in a non-gaming context. The
70 efficacy of videogames to facilitate changes in sensitivity to various stimulus features has been
71 well explored in the visual domain (e.g. Riesenhuber, 2004; Green and Bavelier, 2007; Li *et al.*,
72 2009). However, studies focusing on accelerated perceptual learning with video games in the
73 auditory domain are comparatively sparse (e.g. Honda *et al.*, 2007; Lim and Holt, 2011). The
74 assumption is that the popular game design principles increase the behavioral relevance of the
75 stimuli by providing incentives, which influences processing of low-level stimulus features
76 (Ahissar and Hochstein, 1993). It therefore seems plausible that training paradigms designed to
77 facilitate perceptual learning in an auditory task could be optimized using gamification.

78 The aims of this experiment were therefore firstly to develop a training paradigm that can be
79 used to facilitate and measure adaptation to non-individualized HRTFs. The question of whether
80 improvements are due to HRTF-specific adaptation was addressed by using a control HRTF, which
81 was not used during training. Secondly, the effect of gamification (the introduction of game-like
82 performance-related feedback to the user) was investigated. It was subsequently hypothesized that
83 active listening (the ability of the listener to move the head relative to a spatialized sound source)
84 could play a key role in the adaptation process. A second experiment was therefore carried out
85 using the same system to test this. Finally, with a view to making the results easily translatable to
86 consumer-oriented systems, this paradigm was developed on a commercially available smartphone
87 platform.

88 **II. METHODS**

89 **A. Experimental Design**

90 This study comprised two experiments, both of which utilized the same experimental setup to
91 measure virtual sound localization accuracy. Localization accuracy was evaluated at multiple
92 timepoints following brief localization training sessions. The first experiment investigated two
93 types of training paradigms, gamified and non-gamified. In the second experiment, participants
94 used a modified version of the gamified training paradigm in which they could move their heads
95 relative to the spatialized sources during playback.

96 **B. Participants**

97 A total of 27 adult participants (aged 18 to 38) were recruited for this study. Of these, 16 took
98 part in the first experiment investigating the effect of gamification. These were randomly divided
99 into the two groups, the first of which were assigned to the non-gamified training paradigm (n=9),
100 and the second to the gamified version (n=7). The remaining 11 participants took part in the second
101 experiment incorporating active listening. All participants were asked to complete a questionnaire,
102 which revealed no reported cognitive or auditory deficits.

103 **C. Procedure**

104 Participants were seated on a swivel chair in the center of a quiet room. A virtual environment
105 was presented using a head-mounted display and auditory stimuli were presented over headphones.
106 During both training and evaluation phases, participants initiated a trial by orienting towards a
107 button in the virtual scene and activating it using a handheld controller. Doing so initiated playback
108 of a randomly selected auditory stimulus spatialized at a random location. Source locations were

Sound localization in Virtual Reality

109 uniformly distributed over the upper hemisphere by setting $\theta=2\pi u$ and $\phi=\sin^{-1}v$, where θ and ϕ are
110 the azimuth and elevation angles respectively and u and v are random variates uniformly
111 distributed on the interval $[0, 1]$.

112 In the first experiment, participants were required to orient towards the virtual button
113 throughout playback of the stimulus before orienting towards the perceived direction of the source
114 and indicating their response using the handheld controller. This ensured that sources were
115 presented from a consistent relative direction. If participants moved their head by more than 2°
116 during stimulus playback, the trial was cancelled. In the second experiment, there was no
117 requirement to maintain a fixed orientation and the stimulus was repeated until a response was
118 given. This enabled the listener to affect relative motion of the sound source by turning the head.

119 During training, the correct position of the target was indicated visually after participants gave
120 their response by creating objects in the virtual scene. The object was either a plain, spherical
121 object or an animated spherical robot for the non-gamified and gamified versions respectively. If
122 the target was outside the field of vision, the direction was indicated using an arrow.

123 The size of the object varied adaptively according to the participant's performance. The initial
124 target size was set such that responses were recorded as a 'hit' if there was less than 25° deviation
125 from the target center in any direction. After achieving 3 consecutive hits, the target size decreased
126 by 10%. After five misses at a given target size, the target size reverted to the previous one until
127 reaching the initial size. The radius of the target object was therefore given by $r=0.9^{L-1}d\sin \theta$, where
128 L is the current difficulty level, d is the target distance and θ is the allowed angle error for a correct
129 response.

Sound localization in Virtual Reality

130 Evaluation sessions were carried out using much the same procedure as the training in the first
131 experiment, except no positional feedback was given. To ensure consistency across participants,
132 target stimuli were positioned systematically. Initially, 12 orientations were defined comprising
133 azimuths at 45° intervals (beginning at 0°) at 0° elevation, and at 90° intervals (beginning at 45°)
134 at 45° elevation. For each evaluation, 4 stimuli were presented corresponding to each of these
135 orientations, giving a total of 48 trials per session, which were presented in a random order. To
136 minimize the chance that participants were simply learning target/response pairs rather than
137 adapting to the new HRTF, each target deviated randomly from the corresponding orientation by
138 up to 20°. For 3 of the 4 stimuli, the same HRTF that was used in the training sessions was used
139 to spatialize the sound. For the other stimulus, a second HRTF was used for the spatialization,
140 which acted as a control condition. In this way, HRTF-specific adaptations could be disambiguated
141 from those that generalize across more than one HRTF.

142 **TABLE I:** Sequence of experiment sessions over the 3 days.

<i>Day 1</i>	<i>Day 2</i>	<i>Day 3</i>
Tutorial	Evaluation 5	Evaluation 7
Evaluation 1	Training 4	Training 7
Training 1	Training 5	Training 8
Evaluation 2	Training 6	Training 9
Training 2	Evaluation 6	Evaluation 9
Evaluation 3		
Training 3		
Evaluation 4		

143 Both experiments comprised 9 training sessions of 12 minutes split over 3 days. On day 1,
144 participants were required to complete a tutorial in which they initiated trials in the same way as
145 described above and located targets visually. No auditory stimuli were presented during this phase.
146 The sequence of sound localization evaluation and training sessions is outlined in Table I.

Sound localization in Virtual Reality

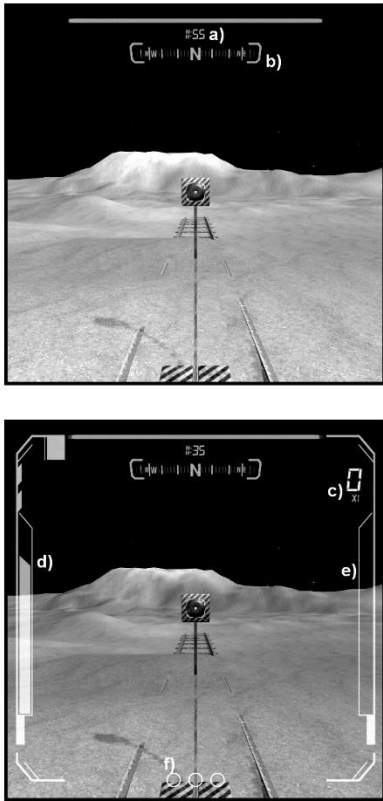


Figure 1: Screenshots of the application as seen by the participant undergoing non-gamified (upper) and gamified (lower) sound localization training. HUD elements are labelled in light text and correspond to a) time remaining, b) orientation (compass), c) player score, d) player health, e) stimulus playback indicator and f) consecutive hit counter.

147 **D. Materials and stimuli**

148 The virtual environment was rendered stereoscopically on a smartphone-based head-mounted
149 display. Participants interacted with the phone using a handheld controller connected via
150 Bluetooth. Head-tracking data was transmitted via wireless Ethernet connection to a separate PC
151 that handled spatial audio rendering. Sound playback and real-time binaural spatialization were
152 implemented using the LIMSI Spatialization Engine (Katz *et al.*, 2010), a real-time binaural
153 audio spatialization platform based on Cycling74's Max/MSP. Binaural audio was presented via a
154 Focusrite Scarlett 2i2 USB audio interface using Sony MDR 7506 closed-back headphones.

Sound localization in Virtual Reality

155 A virtual moon-like environment was designed to be acoustically neutral to minimize the
156 potential mismatch between the anechoic stimuli and the perceived acoustic properties of the
157 virtual space. The scene was also populated with some landmarks and a compass to facilitate
158 orientation, as it has been shown that a lack of visual frame of reference is detrimental to sound
159 localization (Shelton and Searle, 1980). In the gamified version of the task, performance-related
160 feedback was delivered to the participant using an HUD (head-up display), which displayed player
161 score, health and the number of consecutive hits, as shown in Figure 1.

162 A set of 19 acoustically complex stimuli were developed to provide sufficiently rich cues for
163 sound localization. The stimuli comprised a combination of pink ($1/f$) noise, a short segment of
164 Italian speech produced by a male talker and a 1 kHz tone. Each stimulus used different noise
165 tokens and speech segments. A schematic of the stimulus is shown in Figure 2. An initial 200 ms
166 noise burst is followed by a 1 second fragment of continuous Italian speech with low level pink
167 noise, another 200 ms noise burst and, finally, a 200 ms, 1 kHz tone. Each segment was ramped
168 on and off using a 10 ms raised-cosine ramp. To fit with the aesthetic of the virtual environment,
169 the relative levels were set such that the stimulus resembled a short radio communication. From
170 this set, a single stimulus was used only during evaluation sessions, whilst all other stimuli were
171 used during training.

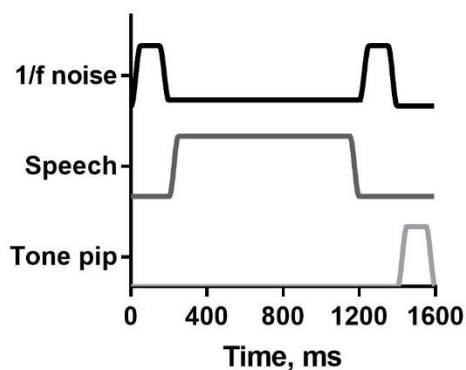


Figure 2: Schematic representation of the target stimulus comprising pink ($1/f$) noise, a segment of Italian speech and a 1 kHz tone.

172 Based on the head tracking data and control signals from the system, sounds were spatialized using
173 HRTFs from the IRCAM Listen database (Warufsel, 2002). Two HRTFs were randomly selected
174 from a subset of this database, which was determined in an earlier study to contain the 7 HRTFs
175 that produced the best subjective spatialization (Katz and Parseihian, 2012). These correspond to
176 participant numbers IRC0008 and IRC0013 in the database. All stimuli were generated and stored
177 in 44.1 kHz, 16-bit format.

178 **III. RESULTS**

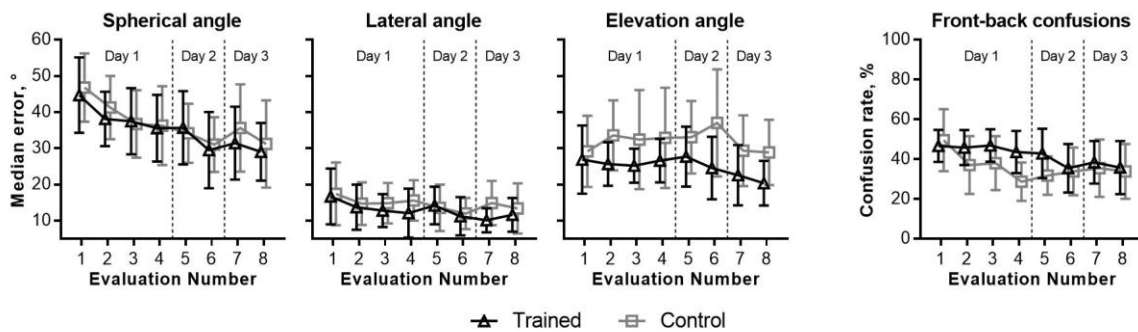
179 **A. Reduction of sound localization error**

180 The first question addressed was whether adaptation to a non-individualized HRTF could be
181 induced using sound localization training with positional feedback, and whether this adaptation
182 was HRTF-specific. To investigate this, the angle errors between target and response during each
183 of the evaluation sessions were initially calculated and are shown in Figure 3 (left). A two-way
184 repeated measures ANOVA was calculated with the evaluation number and HRTF (trained vs
185 control) as the within-participants independent variables and spherical angle error as the dependent
186 variable. For this analysis, data were combined across participants regardless of whether they used
187 the gamified or non-gamified training paradigms. The distribution of errors made by each
188 participant was generally skewed, so the per participant median angle error was the dependent
189 variable in all subsequent analyses. There was a significant main effect of the training, $F(7,$
190 $105)=13.51$, $p<0.001$. Interestingly, there was no significant effect of the HRTF, $F(1, 15)=1.835$,
191 $p=0.196$, nor was there any interaction between the effect of the training and the HRTF, $F(7,$
192 $105)=0.842$, $p=0.555$.

193 Further analyses were carried out to elucidate the factors underlying improvements in
194 localization error. Confusions between the front and rear hemispheres are common in virtual audio
195 systems, and are thought to be resolved most easily through the effect of relative motion of the
196 source and listener on ITD, ILD and spectral contrasts in the HRTF (Wightman and Kistler, 1999).
197 Figure 3 (right) shows the mean per-participant front-back confusion rate for each evaluation
198 session. The rates of front-back confusions were analysed using a two-way repeated measures
199 ANOVA in the same manner as described above. There was a significant main effect of the number

Sound localization in Virtual Reality

200 of training sessions, $F(7, 105)=5.98$, $p<0.001$. There was also a significant effect of the HRTF,
201 $F(1,15)=9.00$, $p=0.009$. Confusion rates were, surprisingly, lower in general for the control HRTF
202 than the trained one after the initial training session, by 5.9% on average. There was also a
203 significant interaction between the effect of the training and the HRTF, $F(7, 105)=3.69$, $p=0.001$.
204 This can be observed in the more pronounced initial improvement after the first training session
205 for the control HRTF, which is not apparent in the trained HRTF.



206 **Figure 3:** Summary of localization errors where the head was fixed during stimulus playback.
207 Points correspond to the mean of the per-participant median error made in each evaluation for
208 sounds spatialized using the trained (Δ) and control (\square) HRTFs.

209 Reductions in overall angle error could indicate improved lateralization, presumably reflecting
210 better ability to utilize interaural time difference (ITD) and interaural level difference (ILD) cues.
211 They could also indicate improvements in elevation judgments, which rely more heavily on
212 spectral cues. In order to investigate the relative contributions of these, target and response
213 coordinates were converted to an interaural-polar coordinate system (Morimoto and Aokata,
214 1984). In this auditory-inspired coordinate system, a lateral coordinate represents the angle of the
215 source from the median plane. An elevation coordinate represents a rotation around the interaural

Sound localization in Virtual Reality

216 axis from the horizontal plane on what is known as a cone of confusion. Thus, errors in judgements
217 based on ITD and ILD cues may be separated from those primarily relying on spectral cues.

218 The median per-participant lateral and elevation angle errors are shown in the two central
219 panels in Figure 3. For each of these derived datasets, a two-way repeated measures ANOVA was
220 again carried out. For lateral angle error, there was a significant main effect of the training, $F(7,$
221 $105)=2.98$, $p=0.007$, reflecting a small reduction in lateral angle error over time. There was also a
222 significant effect of the HRTF, $F(1,15)=9.15$, $p=0.009$, which reflects a marginally lower lateral
223 angle error for the trained HRTF on average ($\mu=1.7\%$). There was no significant interaction
224 between the effect of training and the HRTF, $F(7,105)=1.60$, $p=0.14$.

225 For elevation angle errors, there was no significant main effect of the training, $F(7,105)=1.90$,
226 $p=0.08$. There was, however a significant main effect of the HRTF, $F(1,15)=27.65$, $p<0.001$,
227 reflecting more accurate judgements with the trained HRTF than the control by approximately 7.1°
228 on average. This difference is small in the first evaluation, but largest at the final evaluation due
229 to a reduction in mean error for the trained HRTF that is not apparent in the control. It is notable
230 that the difference between the HRTFs is reversed compared to the front-back confusions, which
231 are both related in that they may be driven by spectral cues. There was no significant interaction
232 between the effect of the training and HRTF, $F(7,105)$, $p=0.48$.

233 A summary of these analyses can be seen in Table II. Taken together, these data indicate that
234 visual positional feedback may be used to decrease localization error of virtual sound sources,
235 which manifests primarily in reductions in front-back confusions and small improvements in
236 lateralization.

Sound localization in Virtual Reality

237 **TABLE II:** Summary of two-way repeated measures ANOVA on changes in various types of
238 localization error indicating the effect of training, HRTF (trained vs. control) and the interaction
239 between them.

	<i>Training</i>	<i>HRTF</i>	<i>Training x HRTF</i>
<i>Overall angle</i>	P<0.001	n.s.	n.s.
<i>Front-back confusions</i>	P<0.001	P=0.009	P=0.001
<i>Lateral angle</i>	P=0.007	P=0.009	n.s.
<i>Elevation angle</i>	n.s.	P=0.001	n.s.

240 **B. Effects of gamification**

241 Two versions of the training software were used in the first experiment. The first had a minimal
242 interface and provided no performance-related feedback to the participant, except for the trial by
243 trial positional feedback. The second version was ‘gamified’ by incorporating several common
244 game-design elements including player score and explicit level progression. To investigate the
245 effect of this gamification on the efficacy of the training, participants were randomly split into two
246 groups, which were trained using the non-gamified (N=9) or gamified version (N=7). A mixed-
247 design ANOVA was carried out on the per-participant median angle errors for targets spatialized
248 using the trained HRTF only, with evaluation number as a within-participants factor and training
249 type (gamified vs non-gamified) as a between-participants factor. As expected based on the
250 previous analyses, there was a significant main effect of the training, $F(7, 98)=10.73$, $p<0.001$.
251 There was no significant main effect of the training type, $F(1, 14)=3.75$, $p=0.073$, demonstrating
252 that the groups were well matched in terms of localization performance in general. However, there
253 was no significant interaction between training type and number of training sessions, $F(1,$
254 $14)=1.52$, $p=0.881$, so the introduction of explicit performance-related feedback itself was not
255 sufficient to increase the training-induced perceptual learning effect.

256 **C. Effects of active listening**

257 It was hypothesized that active-listening, the ability of the listener to experience and affect
258 relative motion of the source and the head, might be important to induce this HRTF-specific
259 adaptation. A third version of the training software was produced in which target stimuli were
260 played continuously after a trial was initiated, enabling participants (N=11) to move their head
261 whilst listening. This version also incorporated the game-design elements used previously.

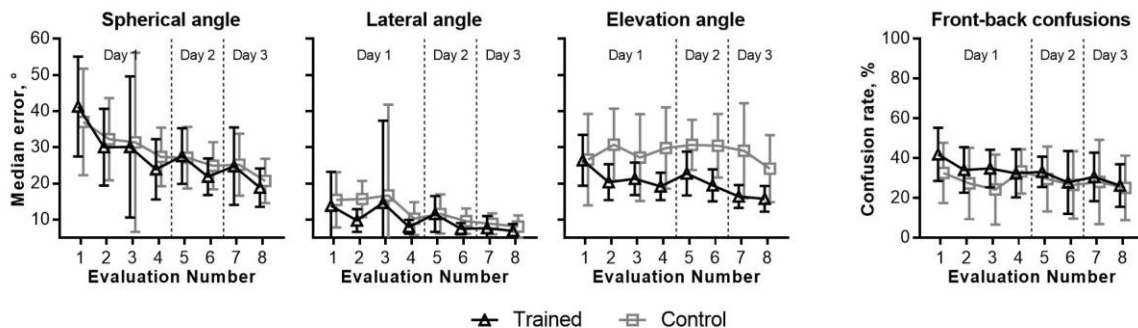
262 **TABLE III:** Summary of two-way repeated measures ANOVA on changes in various types of
263 localization error indicating the effect of training, HRTF (trained vs. control) and the interaction
264 between them for a training paradigm incorporating head-tracking and continuous target stimulus
265 playback.

	<i>Training</i>	<i>HRTF</i>	<i>Training x HRTF</i>
<i>Overall angle</i>	P=0.034	n.s.	n.s.
<i>Front-back confusions</i>	P=0.007	n.s.	n.s.
<i>Lateral angle</i>	n.s.	P=0.004	n.s.
<i>Elevation angle</i>	n.s.	P<0.001	n.s.

266 Localization errors were calculated and analysed in the same way as described previously, the
267 results of which are shown in Figure 4 and summarised in Table III. The overall angle errors were
268 first assessed using a two-way repeated measures ANOVA with the evaluation number and HRTF
269 (trained vs control) as the within-participants independent variables. There was a significant main
270 effect of the training, $F(7, 70)=3.97$, $p=0.034$ (Greenhouse-Geisser corrected), reflecting a general
271 reduction in localization error, primarily on the first day but retained across multiple sessions
272 (Figure 4, left). There was no significant effect HRTF (trained vs control), $F(1, 10)=0.51$, $p=0.49$,
273 nor was there a significant interaction between the HRTF and the effect of the training,
274 $F(7,70)=33.66$, $p=0.20$. These results agree with those of the previous experiment, in that the

Sound localization in Virtual Reality

275 improvements in overall localization accuracy generalized across both HRTFs despite visual
276 feedback being given for only one of them.



277 **Figure 4:** Summary of localization errors where the participant was free to move their head during
278 repeated stimulus playback. Points correspond to the mean of the per-participant median error
279 made in each evaluation for sounds spatialized using the trained (Δ) and control (□) HRTFs.

280 Analysis of the rates of front-back confusions (Figure 4, right) using the same ANOVA as
281 described above likewise showed a significant main effect of the training, $F(7,70)$, $p=0.007$, but
282 no significant effect of the HRTF (trained vs control), $F(1, 10)$, $p=0.065$, nor a significant
283 interaction between the training and HRTF, $F(7, 70)$, $p=0.184$.

284 An interaural coordinate system was again used to investigate errors in lateralization and
285 elevation separately. For lateral angle errors, there was no significant effect of the training, $F(7,$
286 $70)$, $p=0.304$. There was a significant main effect of the HRTF, $F(1, 10)$, $p=0.004$, which reflected
287 generally lower errors for targets spatialized using the trained HRTF, particularly on day 1. The
288 interaction between HRTF and the effect of the training was not significant, $F(7, 70)$, $p=0.25$.
289 Analysis of the elevation errors yielded a similar pattern of results. There was no significant effect
290 of the training, $F(7, 70)$, $p=0.118$, and no significant training/HRTF interaction, $F(7, 70)$, $p=0.085$.

291 There was, however, a significant main effect of the HRTF, $F(1, 10)$, $p < 0.001$ reflecting lower
292 errors for the trained HRTF than for the control.

293 To make a direct comparison of the training paradigm with no continuous target stimulus used
294 in the previous experiment and this training paradigm, which enables active exploration, errors
295 made using the trained HRTF were compared. Only those data from participants using the gamified
296 training paradigm were used in this analysis. A mixed-design ANOVA was carried out on the per-
297 subject median angle errors, with evaluation number as a within-subjects factor and training type
298 (single vs continuous target stimulus) as the between-subjects factor. Whilst there was a significant
299 main effect of the training overall, $F(7, 112) = 8.88$, $p < 0.001$, and a significant effect of the training
300 type, $F(1, 16) = 11.9$, $p = 0.003$, there was no significant interaction, $F(7, 112)$, $p = 0.717$
301 (Greenhouse-Geisser corrected).

302 **D. Timescale of HRTF adaptation**

303 To assess at which point over the course of the experiment the changes in localization
304 performance occurred, data for all participants were combined, regardless of the training paradigm
305 used. These data are summarized in Figure 5 (upper panel). A one-way repeated-measures
306 ANOVA, with evaluation number as the within-subjects factor and overall localization error as the
307 dependent variable unsurprisingly revealed a significant effect of the training, $F(7, 182)$, $p < 0.001$.
308 Bonferroni-corrected pairwise comparisons were made between each evaluation for localization
309 errors where targets were spatialized with the trained HRTF, which are summarized in Table IV.
310 This revealed a significant reduction in error after only a single training session ($p = 0.03$).
311 However, the improvement appears to be retained across the multiple days and improvements are

Sound localization in Virtual Reality

312 ongoing; there was also a significant reduction in errors between the initial evaluation on day 2
 313 and the final evaluation on day 3 ($p < 0.001$).

314 **TABLE IV:** Summary of Bonferroni-corrected pairwise comparisons of average overall angle
 315 error for targets spatialized with the trained HRTF at each evaluation.

	1	2	3	4	5	6	7	8
1	-	0.027	0.432	<0.001	0.002	<0.001	<0.001	<0.001
2		-	1.000	0.208	0.764	<0.001	0.030	<0.001
3			-	1.000	1.000	0.249	1.000	0.038
4				-	1.000	0.122	1.000	<0.001
5					-	0.004	1.000	<0.001
6						-	1.000	1.000
7							-	0.915
8								-

316

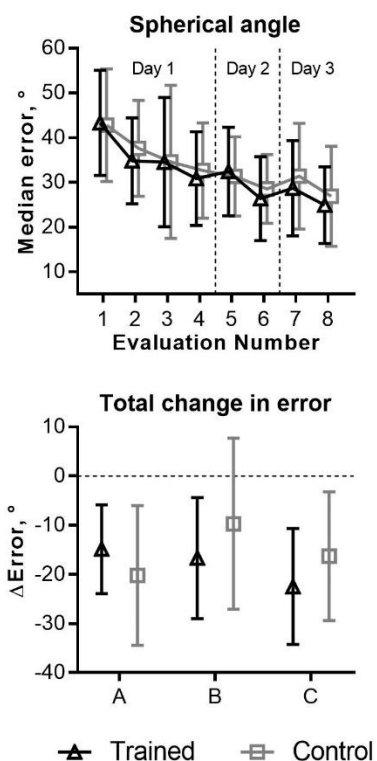


Figure 5: Summary of localization errors pooled across all training paradigms for sounds spatialized using the trained (Δ) and control (\square) HRTFs (upper) and the total change in localization error from the initial to final evaluations for each training paradigm (lower). Groups A and B correspond to non-gamified and gamified training paradigms with the head fixed during stimulus playback, and group C corresponds to the paradigm where head movement (active listening) was encouraged.

316 **IV. DISCUSSION**

317 This study was designed to investigate the effects of training on virtual sound localization
318 using non-individualised HRTFs. The study was divided into two experiments. The first
319 experiment compared gamified and non-gamified training paradigms. The second experiment used
320 a gamified training paradigm that incorporated active listening. In all cases, sound localization
321 accuracy was measured at multiple time points for sounds spatialized using two HRTFs, one of
322 which was used throughout training and evaluation sessions, another which was used only during
323 evaluation sessions.

324 **A. The effect of training on virtual sound localization accuracy**

325 This study used visual positional feedback in a virtual reality system to decrease virtual sound
326 localization errors. The system was developed using readily available consumer electronics, such
327 that it could easily be implemented in a consumer-facing system. The reduction in errors reflected
328 significantly fewer front-back confusions and, in the first experiment, improvements in
329 lateralization accuracy. The reduction in front-back confusions is consistent with previous studies
330 with training of comparable timescales (Zahorik *et al.*, 2006; Majdak *et al.*, 2010; Parseihian and
331 Katz, 2012). Surprisingly, the analyses indicate no significant effect of training on elevation errors,
332 although the data suggests a general improvement over time particularly for the trained HRTF
333 (Figures 3 and 4).

334 Despite the apparent rapid onset of the adaptation, it is not clear that the process has reached
335 a plateau by the end of the experiments. Indeed, post hoc analysis indicated that mean errors are
336 still decreasing between days two and three. It seems plausible, therefore, that further training
337 sessions would lead to further improvements in localization accuracy at the expense of becoming

338 impractical for consumer applications. Earlier experiments also seem to suggest that the timescale
339 of adaptation is considerably longer (e.g. Kumpik *et al.*, 2010; Majdak *et al.*, 2013) and it may be
340 that this timescale is imposed the rate of plastic changes in the brain, which would be difficult to
341 circumvent.

342 **B. Underlying mechanisms of HRTF adaptation**

343 It was hypothesized that training would lead to HRTF-specific improvements in localization
344 accuracy. The reason for this hypothesis was that improvements in virtual sound localization are
345 often explained as ‘adaptation’ to a given HRTF. To investigate this, this study differed from
346 previous, similar studies by incorporating sounds spatialized using a second non-individualized
347 HRTF during the evaluation phases, which acts as a control. This made it possible to discriminate
348 between improvements that are HRTF-specific and due to the listener learning to use idiosyncratic
349 non-individualized cues for source location, and those due to other factors discussed below.

350 Across both experiments, it was found that any changes in localization accuracy for the trained
351 HRTF were not significantly different to those for the control HRTF. Surprisingly, the only case
352 where a significant interaction was found was in an analysis of the front-back confusions in the
353 first experiment, which reflects a more pronounced early-onset improvement for the control
354 HRTF. It may be that the control HRTF exhibits stronger perceptual contrasts between frontal and
355 rear targets, which participants were able to utilise, the benefit of which disappears as a result of
356 training with the other HRTF.

357 One reason for the finding that training generalized across both HRTFs could be that the two
358 HRTFs used in this study were perceptually similar to each other. The method used to select them
359 from the HRTF database would certainly not guarantee perceptual distinctiveness; the subset they

Sound localization in Virtual Reality

360 were taken from was selected based on effectiveness for producing subjectively more realistic,
361 spatialized percepts for the greatest number of listeners (Katz and Parseihian, 2012). It could be
362 argued that this method would tend to produce a subset that epitomizes stereotypical features and
363 minimizes idiosyncratic variation. To discriminate between changes due to HRTF-specific
364 adaptation and other factors, it would be useful to select HRTFs that are perceptually distinct. This
365 could be done using perceptually-based distance metrics, such as those proposed by So *et al.*
366 (2010).

367 A second possibility is related to the putative mechanisms of adaptation. Earlier research in
368 this area has suggested that the adaptation process involves a re-mapping of spectro-temporal
369 features to source locations (e.g. Van Wanrooij and Van Opstal, 2005). However, training using
370 non-individualized HRTFs typically produces little to no after-effects; learning to localize sound
371 with ‘new ears’ does not result in decreased localization accuracy once the listeners original
372 HRTFs are restored. This led others to suggest that, rather than a re-mapping process, ‘adaptation’
373 could involve the development of a parallel internal auditory-spatial map (Hofman *et al.*, 1998;
374 Trapeau *et al.*, 2016).

375 An alternative possibility, which could account for our finding that the adaptation appears to
376 generalize to more than one HRTF, is that the process involves a re-weighting of acoustic cues for
377 sound source location. In this scenario, listeners learn to rely less on features specific to their own
378 HRTF and more on features that are common between theirs and the other HRTFs. A simple
379 example might be that listeners may begin to rely on interaural level differences more than time
380 differences if they are more reliably informative when the HRTF is altered. Such a mechanism
381 relies on redundancy in auditory-spatial cues, but would explain the observation of little to no

382 after-effects and has been put forward as a process underlying auditory perceptual learning in other
383 contexts (Kumpik *et al.*, 2010; Jones *et al.*, 2013).

384 **C. Effects of gamification**

385 An idea that has been receiving considerable attention is that the introduction of game-design
386 elements can have an amplifying effect on perceptual learning, which has been well studied in the
387 visual domain (see INTRODUCTION) but has been relatively little explored in audition (Honda
388 *et al.*, 2007; Whitton *et al.*, 2014; Zhang *et al.*, 2017). It has been proposed that gameplay initiates
389 the release of neural reward signals, which promote synaptic plasticity associated with learning
390 (e.g. Jay, 2003; Harley, 2004). A gamified interface for the sound localization task in this study
391 was implemented, which incorporated performance-related feedback by, for example, awarding
392 points for hits and decreasing player ‘health’ for misses. However, the introduction of these game-
393 design elements had no significant effect on the efficacy of the training. This may be because the
394 visual positional feedback given during training sessions in the non-gamified training provides a
395 level of performance related feedback, enough to activate stimulate similar reward mechanisms,
396 and the introduction of scoring mechanics is superfluous.

397 **D. Effects of active listening**

398 A review of many HRTF adaptation studies has suggested a possible augmenting role of
399 sensory-motor interaction in the process (Mendonça, 2014); paradigms that enable the listener to
400 actively move the source relative to the head tend to be more effective than those that do not. The
401 second experiment presented here was therefore designed to investigate the potential role of this
402 ‘active listening’ on the efficacy of the training process. However, analyses revealed no significant
403 effect of incorporating active listening in the training paradigm. Indeed, the training had similar

Sound localization in Virtual Reality

404 effects regardless of the paradigm used (Figure 5, lower panel). It could be that differences between
405 these training paradigms emerge over longer timescales, but since this study was restricted to short
406 training sessions, and the resulting effects are small, such differences were not apparent.

407 **E. Concluding remarks**

408 Virtual audio systems may be viewed as a useful tool to create realistic, ecologically relevant
409 environments whilst retaining a high degree of experimental control. One exciting possibility is
410 that they could even be used in future to assess hearing impairment in realistic virtual auditory
411 environments. In such an application, one would be interested in optimizing the system rapidly.
412 Whilst it seems that short-term localization training leads to the brain adapting to non-
413 individualized cues in a generic HRTF, the effects are small over short timescales (<1 hour). Future
414 work could investigate how important these effects are in the context of other factors such as the
415 use of appropriate reverberation, or in more ecologically relevant tasks, such as speech recognition
416 in ‘cocktail party’-type scenarios (Cherry, 1953).

417 Another application of virtual audio could be to address psychological and neurophysiological
418 questions about the mechanisms of perceptual learning, given the ease with which such systems
419 can be used to manipulate factors that would be difficult with conventional loudspeaker setups.
420 The question of whether HRTF ‘adaptation’ can be attributed to the development of parallel
421 internal auditory-spatial maps or cue re-weighting, for example, remains open.

422 **V. ACKNOWLEDGEMENTS**

423 This work was supported by the 3D Tune-In project (Eastgate *et al.*, 2016; Levtov *et al.*, 2016),
424 European Union's Horizon 2020 research and innovation programme under grant agreement No
425 644051. The authors would like to thank Brian FG Katz and his team at LIMSI-CNRS for the use
426 of the LIMSI Spatialization Engine.

VI. REFERENCES

Ahissar, M., and Hochstein, S. (1993). "Attentional control of early perceptual learning," *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5718-5722.

Carlile, S., and Blackman, T. (2014). "Relearning auditory spectral cues for locations inside and outside the visual field," *J. Assoc. Res. Otolaryngol.* **15**, 249-263.

Cherry, E. C. (1953). "Some experiments on the recognition of speech, with one and with two ears," *J. Acoust. Soc. Am.* **25**, 975-979.

Eastgate, R., Picinali, L., Patel, H., and D'Cruz, M. (2016). "3D Games for Tuning and Learning About Hearing Aids," *Hear. J.* **69**, 30-32.

Fuchs, E., and Flügge, G. (2014). "Adult neuroplasticity: more than 40 years of research," *Neural Plast.* **2014**.

Gardner, W. G., and Martin, K. D. (1995). "HRTF measurements of a KEMAR," *J. Acoust. Soc. Am.* **97**, 3907-3908.

Green, C. S., and Bavelier, D. (2007). "Action-video-game experience alters the spatial resolution of vision," *Psychol. Sci.* **18**, 88-94.

Harley, C. W. (2004). "Norepinephrine and dopamine as learning signals," *Neural Plast.* **11**, 191-204.

Hofman, P. M., Van Riswick, J. G., and Van Opstal, A. J. (1998). "Relearning sound localization with new ears," *Nat. Neurosci.* **1**, 417-421.

Honda, A., Shibata, H., Gyoba, J., Saitou, K., Iwaya, Y., and Suzuki, Y. (2007). "Transfer effects on sound localization performances from playing a virtual three-dimensional auditory game," *Appl. Acoust.* **68**, 885-896.

Jay, T. M. (2003). "Dopamine: A potential substrate for synaptic plasticity and memory mechanisms," *Prog. Neurobiol.* **69**, 375-390.

Jones, P. R., Moore, D. R., Amitay, S., and Shub, D. E. (2013). "Reduction of internal noise in auditory perceptual learning," *J. Acoust. Soc. Am.* **133**, 970-981.

Kahana, Y., Nelson, P. A., Petyt, M., and Choi, S. (1999). "Numerical modelling of the transfer functions of a dummy-head and of the external ear," in *AES 16th International Conference (AES)*.

Katz, B., Rio, E., and Piccinali, L. (2010). "LIMSI Spatialization Engine," Inter Deposit Digital Number: F.

Katz, B. F. (2001). "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *The Journal of the Acoustical Society of America* **110**, 2440-2448.

Katz, B. F., and Parseihian, G. (2012). "Perceptually based head-related transfer function database optimization," *J. Acoust. Soc. Am.* **131**, EL99-EL105.

Kumpik, D. P., Kacelnik, O., and King, A. J. (2010). "Adaptive reweighting of auditory localization cues in response to chronic unilateral earplugging in humans," *J. Neurosci.* **30**, 4883-4894.

Levtov, Y., Piccinali, L., D'Cruz, M., and Simeone, L. (2016). "3D Tune-In: The use of 3D sound and gamification to aid better adoption of hearing aid technologies," in *AES 140th International Convention (AES)*.

Li, R., Polat, U., Makous, W., and Bavelier, D. (2009). "Enhancing the contrast sensitivity function through action video game training," *Nat. Neurosci.* **12**, 549-551.

Lim, S. j., and Holt, L. L. (2011). "Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization," *Cogn. Sci.* **35**, 1390-1405.

Majdak, P., Goupell, M. J., and Laback, B. (2010). "3-D localization of virtual sound sources: effects of visual environment, pointing method, and training," *Atten. Percept. Psychophys.* **72**, 454-469.

Majdak, P., Walder, T., and Laback, B. (2013). "Effect of long-term training on sound localization performance with spectrally warped and band-limited head-related transfer functions," *J. Acoust. Soc. Am.* **134**, 2148-2159.

Mendonça, C. (2014). "A review on auditory space adaptations to altered head-related cues," *Front. Neurosci.* **8**, 219.

Mendonça, C., Campos, G., Dias, P., Vieira, J., Ferreira, J. P., and Santos, J. A. (2012). "On the improvement of localization accuracy with non-individualized HRTF-based sounds," *J. Audio Eng. Soc.* **60**, 821-830.

Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D. (1996). "Binaural technique: Do we need individual recordings?," *J. Audio Eng. Soc.* **44**, 451-469.

Morimoto, M., and Aokata, H. (1984). "Localization cues of sound sources in the upper hemisphere," *J. Acoust. Soc. Jpn.* **5**, 165-173.

Parseihian, G., and Katz, B. F. (2012). "Rapid head-related transfer function adaptation using a virtual auditory environment," *J. Acoust. Soc. Am.* **131**, 2948-2957.

Riesenhuber, M. (2004). "An action video game modifies visual processing," *Trends Neurosci.* **27**, 72-74.

Shelton, B., and Searle, C. (1980). "The influence of vision on the absolute identification of sound-source position," *Atten. Percept. Psychophys.* **28**, 589-596.

So, R. H., Ngan, B., Horner, A., Braasch, J., Blauert, J., and Leung, K. (2010). "Toward orthogonal non-individualised head-related transfer functions for forward and backward directional sound: cluster analysis and an experimental study," *Ergonomics* **53**, 767-781.

Trapeau, R., Aubrais, V., and Schönwiesner, M. (2016). "Fast and persistent adaptation to new spectral cues for sound localization suggests a many-to-one mapping mechanism," *J. Acoust. Soc. Am.* **140**, 879-890.

Väljamäe, E., Larsson, P., Västfjäll, D., and Kleiner, M. (2004). "Auditory presence, individualized head-related transfer functions, and illusory ego-motion in virtual environments," in *Proc. of 7th Annual International Workshop on Presence*.

Van Wanrooij, M. M., and Van Opstal, A. J. (2005). "Relearning sound localization with a new ear," *J. Neurosci.* **25**, 5413-5424.

Warufsel, O. (2002). "IRCAM Listen Database," <http://recherche.ircam.fr/equipes/salles/listen/> (accessed 15th August, 2017)

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (1993). "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.* **94**, 111-123.

Whitton, J. P., Hancock, K. E., and Polley, D. B. (2014). "Immersive audiomotor game play enhances neural and perceptual salience of weak signals in noise," *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2606-E2615.

Wightman, F. L., and Kistler, D. J. (1989). "Headphone simulation of free-field listening. II: Psychophysical validation," *J. Acoust. Soc. Am.* **85**, 868-878.

Wightman, F. L., and Kistler, D. J. (1999). "Resolution of front-back ambiguity in spatial hearing by listener and source movement," *J. Acoust. Soc. Am.* **105**, 2841-2853.

Zahorik, P., Bangayan, P., Sundareswaran, V., Wang, K., and Tam, C. (2006).

"Perceptual recalibration in human sound localization: Learning to remediate front-back reversals," J. Acoust. Soc. Am. **120**, 343-359.

Zhang, Y.-X., Tang, D.-L., Moore, D. R., and Amitay, S. (2017). "Supramodal enhancement of auditory perceptual and cognitive learning by video game playing," Front. Psych. **8**.