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

Mark A. Steadman[1a], Chungeun Kim[1], Jean-Hugues Lestang[2], Dan F. M. Goodman[2] and Lorenzo Picinali[1]

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

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

(Dated 2nd October 2017)

Running title: Sound localization in virtual reality

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

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

I. INTRODUCTION

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

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

It is generally thought that the differences between an individual's true HRTF and these generic HRTFs have a detrimental effect on the perceived realism of virtual sound sources. It has been noted, for example, that the listeners are able to localize a sound that is spatialized using their own HRTF with a similar accuracy to free field listening, albeit with poorer elevation judgments and increased front-back confusions (Morimoto and Aokata, 1984; Wightman and Kistler, 1989). These errors are typically exacerbated where non-individualized HRTFs are used (Wenzel *et al.*, 1993; Møller *et al.*, 1996). Furthermore, it has been suggested that the use of non-individualized HRTFs results in an auditory perception with reduced 'presence' (Väljamäe *et al.*, 2004).

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

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

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

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

II. METHODS

A. Experimental Design

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

B. Participants

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

C. Procedure

Participants were seated on a swivel chair in the center of a quiet room. A virtual environment was presented using a head-mounted display and auditory stimuli were presented over headphones. During both training and evaluation phases, participants initiated a trial by orienting towards a button in the virtual scene and activating it using a handheld controller. Doing so initiated playback of a randomly selected auditory stimulus spatialized at a random location. Source locations were uniformly distributed over the upper hemisphere by setting $\theta=2\pi u$ and $\phi=sin^{-1}v$, where θ and ϕ are the azimuth and elevation angles respectively and u and v are random variates uniformly distributed on the interval [0, 1].

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

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

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

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

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

Day 1	*Day 2*	*Day 3*
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		

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

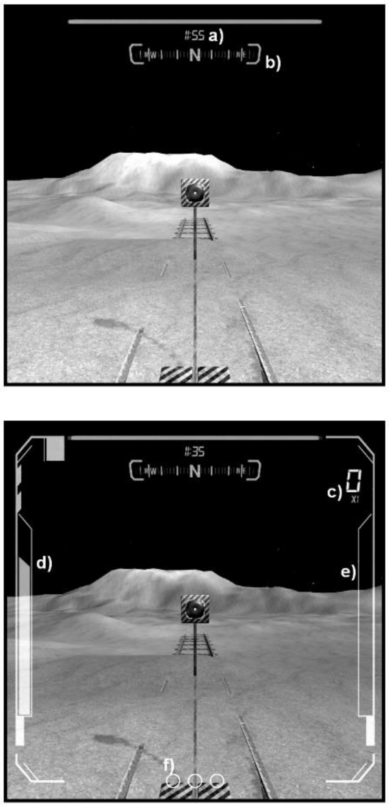

Figure 1: Screenshots of the application as seen by the participant undergoing non-gamified (upper) and gamified (lower) sound localization training. HUD elements are labelled in light text and correspond to a) time remaining, b) orientation (compass), c) player score, d) player health, e) stimulus playback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

D. Materials and stimuli

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

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

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

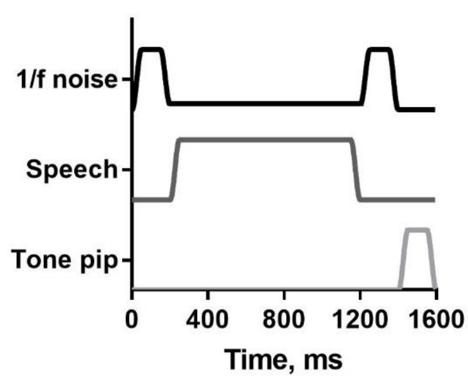

Figure 2: Schematic representation of the target stimulus comprising pink (1/f) noise, a segment of Italian speech and a 1 kHz tone.

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

III. RESULTS

A. Reduction of sound localization error

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

Further analyses were carried out to elucidate the factors underlying improvements in localization error. Confusions between the front and rear hemispheres are common in virtual audio systems, and are thought to be resolved most easily through the effect of relative motion of the source and listener on ITD, ILD and spectral contrasts in the HRTF (Wightman and Kistler, 1999). Figure 3 (right) shows the mean per-participant front-back confusion rate for each evaluation session. The rates of front-back confusions were analysed using a two-way repeated measures ANOVA in the same manner as described above. There was a significant main effect of the number of training sessions, F(7, 105)=5.98, p<0.001. There was also a significant effect of the HRTF, F(1,15)=9.00, p=0.009. Confusion rates were, surprisingly, lower in general for the control HRTF that the trained one after the initial training session, by 5.9% on average. There was also a significant interaction between the effect of the training and the HRTF, F(7, 105)=3.69, p=0.001. This can be observed in the more pronounced initial improvement after the first training session for the control HRTF, which is not apparent in the trained HRTF.

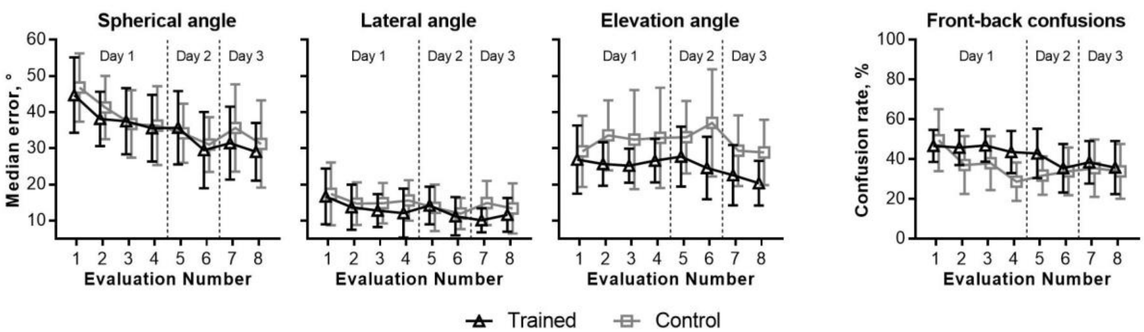

Figure 3: Summary of localization errors where the head was fixed during stimulus playback. Points correspond to the mean of the per-participant median error made in each evaluation for sounds spatialized using the trained (Δ) and control (□) HRTFs.

Reductions in overall angle error could indicate improved lateralization, presumably reflecting better ability to utilize interaural time difference (ITD) and interaural level difference (ILD) cues. They could also indicate improvements in elevation judgments, which rely more heavily on spectral cues. In order to investigate the relative contributions of these, target and response coordinates were converted to an interaural-polar coordinate system (Morimoto and Aokata, 1984). In this auditory-inspired coordinate system, a lateral coordinate represents the angle of the source from the median plane. An elevation coordinate represents a rotation around the interaural axis from the horizontal plane on what is known as a cone of confusion. Thus, errors in judgements based on ITD and ILD cues may be separated from those primarily relying on spectral cues.

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

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

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

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

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

B. Effects of gamification

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

C. Effects of active listening

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

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

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

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

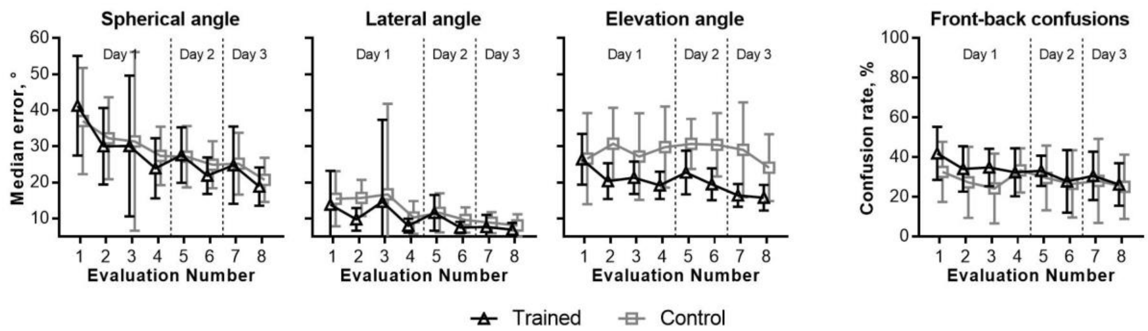

Figure 4: Summary of localization errors where the participant was free to move their head during repeated stimulus playback. Points correspond to the mean of the per-participant median error made in each evaluation for sounds spatialized using the trained (Δ) and control (□) HRTFs.

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

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

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

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

D. Timescale of HRTF adaptation

To assess at which point over the course of the experiment the changes in localization performance occurred, data for all participants were combined, regardless of the training paradigm used. These data are summarized in Figure 5 (upper panel). A one-way repeated-measures ANOVA, with evaluation number as the within-subjects factor and overall localization error as the dependent variable unsurprisingly revealed a significant effect of the training, F(7, 182), p<0.001. Bonferroni-corrected pairwise comparisons were made between each evaluation for localization errors where targets were spatialized with the trained HRTF, which are summarized in Table IV. This revealed a significant reduction in error after only a single training session (p=0.03). However, the improvement appears to be retained across the multiple days and improvements are ongoing; there was also a significant reduction in errors between the initial evaluation on day 2 and the final evaluation on day 3 (p<0.001).

TABLE IV: Summary of Bonferroni-corrected pairwise comparisons of average overall angle 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								-

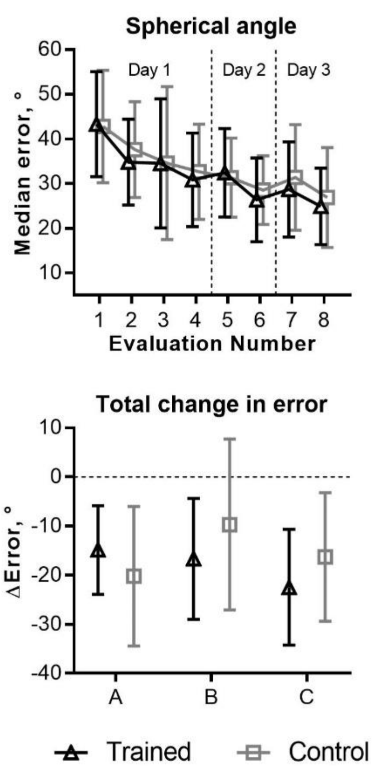

Figure 5: Summary of localization errors pooled across all training paradigms for sounds spatialized using the trained (Δ) and control (□) HRTFs (upper) and the total change in localization error from the initial to final evaluations for each training paradigm (lower). Groups A and B correspond to non-gamified 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.

IV. DISCUSSION

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

A. The effect of training on virtual sound localization accuracy

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

Despite the apparent rapid onset of the adaptation, it is not clear that the process has reached a plateau by the end of the experiments. Indeed, post hoc analysis indicated that mean errors are still decreasing between days two and three. It seems plausible, therefore, that further training sessions would lead to further improvements in localization accuracy at the expense of becoming impractical for consumer applications. Earlier experiments also seem to suggest that the timescale of adaptation is considerably longer (e.g. Kumpik *et al.*, 2010; Majdak *et al.*, 2013) and it may be that this timescale is imposed the rate of plastic changes in the brain, which would be difficult to circumvent.

B. Underlying mechanisms of HRTF adaptation

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

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

One reason for the finding that training generalized across both HRTFs could be that the two HRTFs used in this study were perceptually similar to each other. The method used to select them from the HRTF database would certainly not guarantee perceptual distinctiveness; the subset they were taken from was selected based on effectiveness for producing subjectively more realistic, spatialized percepts for the greatest number of listeners (Katz and Parseihian, 2012). It could be argued that this method would tend to produce a subset that epitomizes stereotypical features and minimizes idiosyncratic variation. To discriminate between changes due to HRTF-specific adaptation and other factors, it would be useful to select HRTFs that are perceptually distinct. This could be done using perceptually-based distance metrics, such as those proposed by So *et al.* (2010).

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

An alternative possibility, which could account for our finding that the adaptation appears to generalize to more than one HRTF, is that the process involves a re-weighting of acoustic cues for sound source location. In this scenario, listeners learn to rely less on features specific to their own HRTF and more on features that are common between theirs and the other HRTFs. A simple example might be that listeners may begin to rely on interaural level differences more than time differences if they are more reliably informative when the HRTF is altered. Such a mechanism relies on redundancy in auditory-spatial cues, but would explain the observation of little to no after-effects and has been put forward as a process underlying auditory perceptual learning in other contexts (Kumpik *et al.*, 2010; Jones *et al.*, 2013).

C. Effects of gamification

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

D. Effects of active listening

A review of many HRTF adaptation studies has suggested a possible augmenting role of sensory-motor interaction in the process (Mendonça, 2014); paradigms that enable the listener to actively move the source relative to the head tend to be more effective than those that do not. The second experiment presented here was therefore designed to investigate the potential role of this 'active listening' on the efficacy of the training process. However, analyses revealed no significant effect of incorporating active listening in the training paradigm. Indeed, the training had similar effects regardless of the paradigm used (Figure 5, lower panel). It could be that differences between these training paradigms emerge over longer timescales, but since this study was restricted to short training sessions, and the resulting effects are small, such differences were not apparent.

E. Concluding remarks

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

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

V. ACKNOWLEDGEMENTS

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