Auditory Occlusion Based on the Human Body in the Direct Sound Path: Measured and Perceivable Effects

STEVEN MCSEVENEY, Edinburgh Napier University, UK
MONICA TAMARIZ, Edinburgh Napier University, UK and Herriot Watt University, UK
IAIN MCGREGOR, Edinburgh Napier University, UK
BABIS KONIARIS, Edinburgh Napier University, UK
KENNY MITCHELL, Edinburgh Napier University, UK

Audio plays a key role in the sense of immersion and presence in VR. Improving the sense of presence has been a long-standing goal in media, as it correlates to improved enjoyment of content. We share results of a perception study on the ability of listeners to recognise auditory occlusion due to the presence of a human being in the direct sound path. We ran two-alternative forced choice trials to test for effects of occluder body type and distance from sound source on recognition of auditory occlusion. Results show that audio cues allow listeners to significantly detect the presence or absence of an occluder, and that position of the occluder relative to the listener and sound source, as well as occluder body type modulate detection rates. Synthesised audio achieved, in selected conditions, better occlusion detection than recorded audio. The work provides details on what filtering occurs across 26 1/3 octave frequency bands when a person comes between a listener and a sound source. This research will inform the recreation of auditory effects in virtual shared spaces due to the presence of user avatars.

CCS Concepts: • **Applied computing → Sound and music computing**; • **Computing methodologies → Simulation evaluation**; **Simulation environments**.

Additional Key Words and Phrases: Sound, auditory occlusion, presence, mixed reality, simulation

ACM Reference Format:
Steven McSeveney, Monica Tamariz, Iain McGregor, Babis Koniaris, and Kenny Mitchell. 2024. Auditory Occlusion Based on the Human Body in the Direct Sound Path: Measured and Perceivable Effects. In . ACM, New York, NY, USA, 24 pages. https://doi.org/XXXXXXX.XXXXXXX

1 BACKGROUND

This paper examines the effects of the interposition of a human body in the direct sound path between sound source and listener. It includes a study on the perceivable differences in auditory occlusion as a result of position and body type. The presence of people affects the sound that a listener hears in a shared physical space, due to occlusion of sound sources. Reproducing these effects in shared online environments has the potential to improve the sense of presence and social connectedness. Within virtual reality applications, audio plays a key role in the sense of immersion and presence [58]. Social connectedness has been recognised as having a positive effect on physical as well as mental health [82]. For example, online dancing is an activity which has been shown to effectively enable a sense of connecting with others [50] and reducing stress [35]. Previous research also points to the importance of users feeling embodied in virtual

---

Manuscript submitted to ACM

environments, particularly regarding auditory feedback from their movements [49]. In VR this is important as users are capable of becoming more than an observer, but an active part of the world. Improving the sense of presence has been a long-standing goal in media as it correlates to improved enjoyment of content [87].

The COVID-19 pandemic has highlighted the importance of digital social interaction in the mental resilience of individuals in combating loneliness [66]. VR interventions have also been shown to reduce psychological distress and anxiety amongst users [39, 73].

In real-world crowds, each individual contributes to the absorption of sound within a shared space, such as a theatre or concert hall [51]. People's position and clothing also affect this absorption factor. In VR shared experiences, this appears to be unexplored, yet acoustic shadows are an important auditory phenomenon which humans rely on to identify objects in a space which we cannot see [25]. Auditory feedback of the environment with regards to the acoustic shadows of individuals could provide another method of enhancing the sense of presence and connectedness of users.

Our research questions are:

(1) Can participants detect presence / absence of an occluder from audio input?
(2) Can participants correctly detect body size and position from audio input?
(3) Does audio type (synthesised or recorded) help detection of presence, occluder body type and occluder position?
(4) Does participant's expertise with audio improve detection of presence, occluder body type and occluder position?

This paper is organised as follows. We start with an exploration of related work in the fields of acoustics and computing, examining previous research on how sound is treated in interactive media such as video games and mixed reality (MR) applications, as well as the technical challenges therein. Then we provide the details of our data gathering methods on the auditory occlusion of a human body. We then discuss the online perceptual study which was conducted to test the perception of occlusion and the results of the measured human auditory occlusion. A discussion of the results is provided, followed by the conclusion of our findings.

2 RELATED WORK

Here we examine various approaches to the improved realism of audio in virtual environments and the need for a sense of presence and connectedness in online applications. We first examine past research in realistic acoustic environments. Then we highlight the potential need for enhancing remote social interaction. Finally, we explore what technical challenges arise when developing audio systems which try to replicate real-world sound interactions.

2.1 Absorption and Occlusion

There is considerable literature on the absorption coefficients of materials [46], sound propagation [72] and reflections [78] with regards to virtual spaces [48]. This research is of particular interest to the games industry, with demand for realism and natural sounding environments to enhance the immersive player experience [34]. With the increased availability of consumer VR products such as the Meta Quest [55] and HTC Vive [38], recent research has focused on methods of improved sound localisation using generalised Head-Related Transfer Functions (HRTFs) [2] and spatial audio techniques [43, 99]. While this research considers how sound emitting objects should be heard, they do not examine the changes to the environment due to the presence of other people.

Less is available on how dynamic occlusion objects may affect the auditory environment. Some, such as work by Raghuvanshi [71], develop methods on auditory changes to the virtual environment through portals. Russell and Brown [75] investigate the salience in the perception of occlusion of sound due to occluding objects. Some studies have been carried out on how people change the sonic characteristics of spaces [76] and the absorption of the body [14, 42], however there is a paucity of research which fully explores or measures the effect of people as occluding objects across the audible spectrum.

Avatars in interactive media such as VR can typically be altered via number of physical attributes including height, weight, shape and gender. As such, when considering the audible changes a person could potentially make to a space, these traits are worth considering. We decided to use three basic body types (ectomorph, mesomorph and endomorph) as used in other research [21, 27, 32, 45], as a means of determining if any noticeable differences were measurable. Gender was also considered as a possible candidate for noticeable measurement differences. Some other aspects were considered such as body fat content and more specific body types, but were determined to be outwith the scope of the initial study. Should a difference in body type become apparent in the measurements, this would present a case for further work.

2.2 Presence and social connectedness online

As the worldwide lock-downs due to the COVID-19 pandemic demonstrated, social isolation exacerbates existing age-related health problems in elderly people, as well as detrimental effects on young people's mental well being [16]. Some social group activities such as dancing have been shown to improve the mental well being of participants [26]. By implementing an auditory feedback system that reacts intuitively as in real world scenarios, the user can be immersed more fully in the experience. This could aid in reducing user loneliness as they can feel more connected to the space and the users within it [66]. Around 15 percent of the world population are living with some form of disability [96]. Online discourse in recent years has led to the impetus for more accessibility in games, allowing those with a variety of disabilities to use and enjoy the medium [22]. A user focused approach to audio in virtual environments could result in the increased accessibility of MR applications in a broader sense.

Recreating humanoid avatars in MR is also a technical challenge, both physically and technologically. Recreating precise movements of participants is difficult, particularly with regards to smaller details such as finger movement tracking [37]. Authentically portraying such experiences in a digital setting requires precision in the visuals, but also in auditory feedback for participants. Effective recreation of subtle, personal sounds of people could be a key factor in building an immersive virtual experience. Combining visual, audio and haptic feedback has been shown to aid in collaborative tasks [57], therefore a combination of these systems are often considered for group social activities. A sense of presence relies on the user believing themselves to be within a virtual space, and yet there is very little research on the sense of agency and body ownership with regards to personal sound i.e. the sound that your avatar makes in the virtual world [31]. While methods of procedurally generating footsteps and materials has been examined [89, 93], the effect on the user's experience has not.

Methods of measuring the sense of immersion are plentiful, covering aspects from physical autonomic responses through to analysis of respondents feedback. In the case of the former, involuntary responses such as eye-movement can be used measure salience of auditory events [41]. Examining the latter, a common method is the use of questionnaires, however the efficacy of such practice in virtual reality scenarios has been questioned [79]. Questionnaires however, are still common practice in the measurement of subjective feelings [40, 52, 92].

2.3 Technical Challenges

Audio in Mixed Reality. MR, as the next step in interactive media, builds upon what has been learned from traditional platforms. With video games in particular, accurate physics based on the ground truth has been a goal for many development studios and platforms [5, 54, 71]. However, focus on how a listener perceives their environment in a virtual space has been largely unexplored [6]. The proliferation of internet enabled applications, especially those with voice-chat capability, compounds this problem by including multiple users within a virtual shared space. Sound designers now should consider how to balance challenges such as realism and immersion, as well as adjusting the audio mix at run-time within the virtual space against what a user requires for the most engaging experience. In addition, spatialisation of audio for a VR user is more important than in traditional media, for example when important elements of the experience occur outside of the users field of view [81].

With the increased expectation of consumers for immersive virtual environments however, comes the need for more resource intensive audio feedback, such as implementation of real-time audio occlusion and reflections, versus baked-in alternatives [71]. With the popularisation of new virtual, mixed and augmented reality technology such as Meta's Quest 3 [55], Valve's Vive Pro [91] and budget friendly options such as Google Cardboard, this appears to be more important than in previous generations of products. While advanced systems such as desktop computers are less susceptible to these resource bottlenecks, less powerful systems can have trouble smoothly running intensive applications[94]. It would be beneficial to discover best-practices for the allocation and focus of those resources in order to prioritise crucial elements of the experience to avoid problems such as voice starvation.

Real versus synthesised sound. It is possible to overlook a key difference in real-world versus virtual world auditory environments, that is using our ears rather than using some form of listening device such as headphones. Although advances have been made with HRTFs and acoustic modelling of virtual environments [9, 15, 64], technological constraints prevent real world interactions of sound with materials, other sounds, high order ambisonics and the listener themselves [44]. Auditory processing encompasses a wide range of acoustic and psychoacoustic phenomena including inter-aural time difference (ITD), inter-aural level difference (ILD), the Haas effect (also known as the precedence effect), masking, missing fundamental, just noticeable differences, equal loudness contours and even physical movement within the ear [83]. All contribute to understanding the surrounding environment. While some of these problems are simple to solve in a digital context, others require vast amounts of computation, and some are beyond the capabilities of today's technology. By understanding and utilising these phenomena, it may be possible to create systems for interactive media which could deliver an intended result while reducing the technical requirements.

A potential solution for these problems lies in the way the human brain handles the hundreds of sound sources, reflections and reverberations of daily life, known as the cocktail party problem [17, 56]. Similarly to how the eyes focus on a narrow band of our total visual field, humans are able to focus our hearing on a particular sound source, utilising top-down attention to group and segregate what we hear [29, 63]. This selective attention allows us to process new and important information quickly by deprioritising unnecessary sound - for example the background dirge of traffic - so as to focus on the footsteps of someone approaching, or a singular voice in a group of people. These concepts could be applied to VR formats in order to emulate the real-world auditory experience with virtual sound sources. It could be possible, for example, to use eye-tracking technology to identify the focus of the user and therefore mix the VR audio environment in real-time, reducing unnecessary sound in that moment. Taking account of the process of auditory scene analysis [10], it may be possible for current technology to be utilised to provide audio feedback in a way which can convey this same information more effectively [36].

Shared Auditory Environments. In group social scenarios, particularly with a partner or crowds of people, it may be possible to create a more active listening experience for users in interactive MR applications, rather than the passive approach seen in today's media. For example, with music playing loudly in a club, it is often impossible to hear the speech from other individuals, even when in close proximity due to the overall volume within the space, often exceeding 100 dB [67] - hearing damage can occur after just 15 minutes of exposure at such loudness levels. In such cases, it is often necessary to shout loudly directly into the ears of someone in order to be heard; a situation which has been shown to lead to frustration [24]. Besides the risk of irreparable hearing damage, recreating such a scenario in virtual environments (VE's) could also cause users to feel overwhelmed . Instead, MR technologies have the potential of creating new and better than real-world experiences [97]. Cognitive overload has been shown to arise in noisy environments, due to the increased effort required to attend to an individuals focus [11]. Indeed, age-related hearing loss (see Fig. 1) has been shown to lead to decreased speech intelligibility [59, 95]. However, clinical hearing assessment typically discounts frequencies above 8 kHz, which is important to speech perception [60]. Accounting for such issues and moving away from accurate representation of acoustics in virtual spaces could provide a more enjoyable and immersive experience for users of MR applications [68]. It could be possible then, to enable a system of focused listening for users which allows them to remain in a shared, noisy virtual environment, while being able to naturally hold a conversation and with less cognitive overload due to reducing the required cognitive effort [47, 97].

3 EXPERIMENT 1: MEASURING HUMAN BODY SOUND OCCLUSION

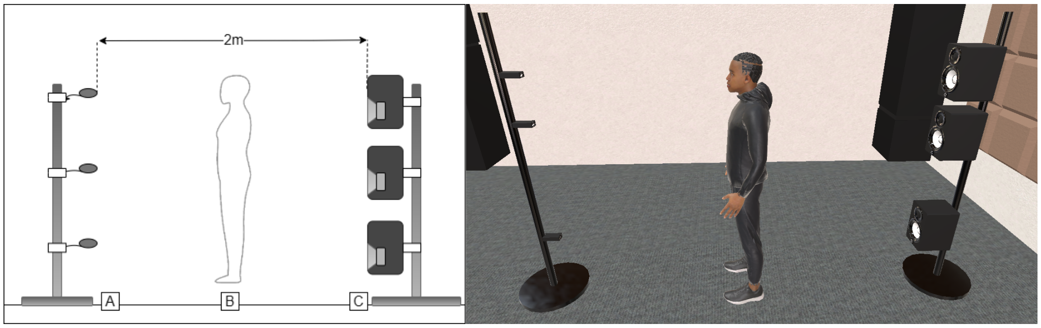
The follow section explains the method of gathering measurement data which would be used for the development of the follow-on perception study in the next chapter. We detail the setup and equipment used, as well as the process of gathering sound occlusion measurements for the variables of gender, body type and distance.

Participants. We invited 25 participants to take part in this measurement study (15M, 10F) from the Merchiston campus at Edinburgh Napier University. All participants were over 18 in accordance with University consent policy. Average age 31, min 18, max 55. Participants were asked to wear only base and mid layers of clothing (see Fig. 1) to avoid extraneous colouration of results due to heavy clothing materials or large garments which would affect their occluding silhouette.

Fig. 1. Participants during the measurement experiments were asked to wear online base and mid layers of clothing. [23]

Metrics including age, height, body type (ectomorph, mesomorph or endomorph [32, 45]) and gender were captured. The participants were fully informed of the purposes of the experiment, and gave consent to take part. Earplugs were available, but the experiment measurements were taken at less than 85dB(A) RMS, below the recommended level set by safety at work regulations in the UK [33]. This was to prevent any possible hearing damage to participants.

Design A suitable room, the Auralization Suite at Edinburgh Napier University [69], was prepared. This room was chosen as it is acoustically treated to reduce reverberation, as well as prevent excessive noise from adjacent rooms and corridors. A loudspeaker array - in this case three Genelec 8030A's [30] - was set up at one side of the room, and a microphone array consisting of 3 Audio-Technica AT831b's [3] at the other (see Fig. 2). Positions for participants to stand were marked on the floor at distances from the microphone of 0m, 0.5m, 1m, 1.5m and 2m. The distance between the speaker array and microphone array was 2m. It should be noted that all distances are given at distance from the microphone. Therefore 0m was directly in front of the capture microphone (without touching the capsule), and 2m from the loudspeaker.



(a)

(b)

Fig. 2. (a) Diagram of measurement equipment setup (not to scale). Microphone array on the left consisting of three AT831b's, loudspeaker array on the right consisting of three Genelec 8030A's. Array's were spaced 2m apart. Participant stood at points A (0m from microphone, referred to as the *close* position for experiment 2), B (1m from microphone, *mid* position) and C (2m from microphone, *far* position). Human outline [88]. (b) Mock-up of the configuration in Unity engine.

In order to eliminate height as a variable, the speakers were adjusted for each participant to be at head, chest and knee height for each individual. The microphones were positioned in the same manner. This was determined the best way to examine the occlusion of the direct path between speaker and microphone through the body and measure a complete picture of the individuals occlusion profile.

Participants self-recorded their body type from three basic somatotypes - ectomorph, mesomorph and endomorph (see Fig. 3)- as explained by Giovanni and Valentina [32] and Lam et al. [45]. These criteria were chosen for this initial study in order to determine if deeper investigation of these factors may be required in potential follow-on studies, but such studies were deemed beyond the scope of this work.

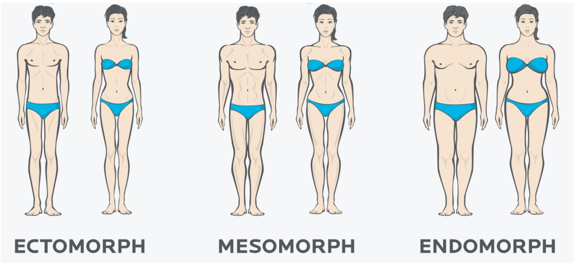


Fig. 3. Illustration shown to participants to help self-record body type for measurements. Image [20]

Materials. The tests were conducted by playing a white noise burst for 500ms through each loudspeaker at 80dB(A) RMS, measured at 1 meter from the loudspeaker, with one second of silence between each burst. Although pink noise is generally used when measuring acoustics, it presents a lower volume at high frequencies. With higher frequencies being shown to improve hearing challenges such as identifying speech in noise [60] and sound localization [1] we therefore chose to use white noise due to expectations that higher frequencies would be most affected in the tests. In this way, we could compare the absolute change in dB across each 1/3 octave frequency band. It also provided an unweighted measurement that could be used for validation testing in other applications.

Procedure. The white noise track was used to capture the sound within the empty room in the first instance, to establish a baseline measurement. Then the participant was asked to stand at the first marked position. The measurements were repeated at each position, to determine the changes at the listener position due to the changing distance of the occluding person. Measuring different distances provides information regarding the changes in occlusion and transmission due to proximity to the loudspeaker and microphone.

The recorded results were then quantified via spectral analysis to determine what changes took place with regards to the propagation of sound between the sound source and the listener when occluded by a human body. The analysis was carried out using MATLAB R2022b [53], where the audio data was then represented as spectrograms, 31-band graphic EQ bar charts, as well as extracting averaged dB levels for each 1/3 octave band as Microsoft Excel data for more precise comparison. By subtracting the values at each position from the empty room (unoccluded) signal values, it was then possible to illustrate the EQ profile of the occluded signals, and examine the precise values for each 1/3 octave band. It should be noted that due to the size of the room used to record the measurements, some low frequency phase issues occurred. As such, we present the EQ charts with 26 1/3 octave bands, starting at 63Hz.

4 EXPERIMENT 2: HUMAN AUDITORY OCCLUSION PERCEPTION STUDY

The following sections describe the details of the perception study, in which we tested conditions such as difference in distance, gender and occluder body type to examine which factors provide perceivable changes in auditory occlusion. The design of the tests and signals were based on the measurement data captured in the first experiment detailed in the previous section of the paper. This study involves a two-alternative forced choice task, and a rating task.

Participants. After gathering measurement data, we then proceeded with a perception study. This was conducted online in order to reach a greater number of participants. For the study, 64 participants (44 male, 16 female, 4 did not disclose) were recruited both online and from Edinburgh Napier University, Merchiston campus (average age = 42, min = 21, max = 70). All participants gave informed consent prior to taking part.

Design. We designed a virtual environment which placed the participant facing an occluding object (a virtual avatar) and a sound source. Fig. 4 illustrates the variables examined during the 2AFC trails.

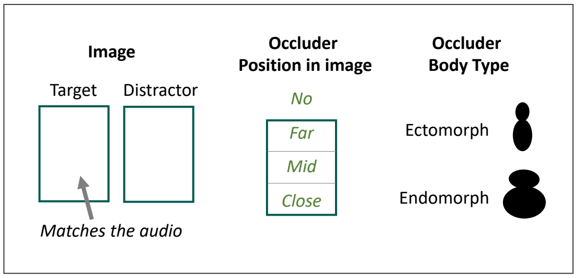


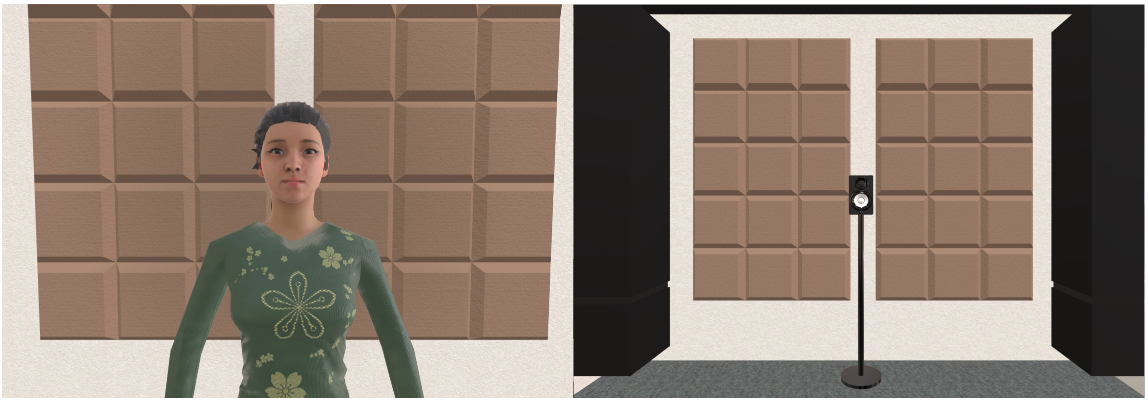
Fig. 4. Variables in the two-alternative forced choice task.

The variables for the scenarios included the distance between the avatar and the participant, the audio signal occlusion method (real recorded occlusion or synthesized occlusion based on measurements in Experiment 1), and if the audio signal aligned with the avatar's distance from the participant. If the avatar was 1m from the participant, the signal would be a recorded or synthesized signal at 1m from the microphones if it was aligned, or a different distance value signal if mismatched. We refer to the "correct" response in each trial as the *target* and the incorrect response as the *distractor*. We tested three distances; 0 metres (directly in front of the participant) 1 metre and 2 metres from the participant (directly in front of the loudspeaker). We chose these particular distances to reduce the number of trials required, and as they showed the largest differences in occlusion. It should be noted that the 1m scenarios were the midpoint between the microphone and speaker, therefore we predicted a less perceivable occlusion effect than the 0m and 2m scenarios, and would be more likely to be confused with the no-occlusion scenarios. We also used three differing music tracks in order to check for consistency across sound sources. "Real" signals were recorded using a Neumann KU-100 binaural dummy head microphone [62], while synthesized signals were produced by applying EQ in relation to the combined average of the microphone array captures. Our decision to do this, and to test for stereo files versus mono synthesized files was with regards to how game engines such as Unity handle 3D audio objects [86]. In such cases, audio sources are treated as a single point source, and therefore it is useful to understand how this process could be utilised to enhance the virtual experience. For representation in binaural experiences, HRTFs would be applied at the final stage of processing, with differing approaches and datasets in use [98]. The next set of trials were conducted as hearing tests, with no images shown to participants. A subset of the audio from the 2AFC trials was used. This set of trials was to determine if participants would respond correctly without a visual prompt.

Materials. The online survey was built using Gorilla [13]. Participants could complete the survey on any PC or laptop device with headphone output. Instructions for calibration of their device was given before the trials commenced, in order to ensure a suitable and safe listening volume. Images for the trials were captured from a virtual representation of the measurement room recreated in the Unity game engine [90]. Copyright free music was acquired via websites FreePD.com [28] and Uppbeat.io [61], which were then used for the real occlusion recordings and the synthesised occlusion clips. To create the synthesised occlusion, the tracks were processed in Reaper [19] according to the measurement data taken in the first experiment.

Procedure. After giving informed consent to take part, participants were asked to complete a short demographic survey to allow examination of trends regarding trained versus untrained listeners. They were then presented with instructions for setting up their device appropriately, including setting volume level and wearing headphones. This aimed to prevent effects of vastly differing volume levels which may affect perception due to loudness contours [18, 85] or masking caused by noise across participant environments [65, 77], as well as to reduce the risk of hearing damage to participants [33].

For the first task, participants were presented with a two-alternative forced choice (2AFC) setup. All trials were randomised to prevent any ordering effect in the results. Each screen presented the participant with an audio sample and two images, representing the occluding person between them and the loudspeaker (see Fig. 5). The audio samples were prepared from the previous measurements acquired during experiment 1. Trials were designed to include comparisons of all combinations of body type (endomorph and ectomorph); gender (male and female), and position of the occluder away from the microophone (close: 0m, middle: 1m and far: 2m). Each combination was presented in a trial with a real recording and in another with a synthesised signal, recreated in a digital audio workstation (DAW) based on the measurements taken in the first experiment. In total, each participant completed 120 2AFC trials. For each trial, they listened to a 10 second audio sample, then chose which of the two images was most appropriate to the audio. From this we can examine which occlusion scenarios presented the most noticeable occlusion and the perception of occlusion as it relates to the visual representation.


(a)

(b)

Fig. 5. An example of one of the 2AFC trials. This trial included two scenarios: (a) a female ectomorph occluder in close position and (b) no occluder. Participants's task was to select the scenario they thought most appropriate to the audio clip being played.

In the second task, designed to establish whether audio input can accurately convey a sense of presence, we presented audio clips to participants with no visual elements. The participants were asked to rate how sure they were that someone was present in the room using a slider from -100 (completely sure there was no-one in the room) to +100 (completely sure there was someone in the room). As well as this continuous measure of presence, participants were also asked to rate "I feel a sense of presence when listening to this audio." in a 7-point Likert scale (1 = Disagree, 7 = Agree). The trials for this task included no occluder, or an occluder with all combinations the three positions, two body types and two genders mentioned above. Each participant completed 24 presence trials. Again, trials were randomised for each participant. Finally, participants were debriefed and thanked for their participation.

5 RESULTS

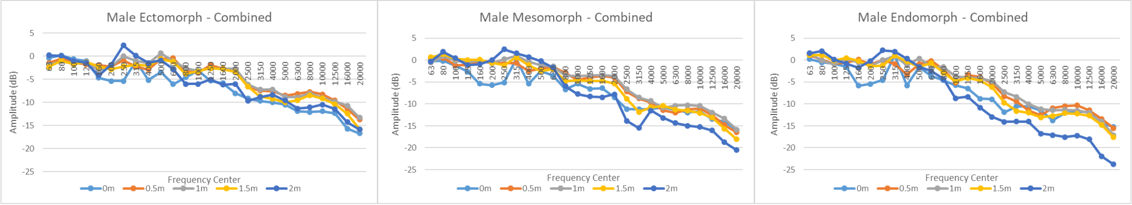
The results of both of our experiments is detailed as follows. Firstly, the measurement data is provided, illustrating the various equalisation profiles based on body type and distance. Next we provide the results from the perception study which we conducted online. The initial two-alternative forced choice trails are broken down into analysis of effects including occluder body type and position, no occluder trials, occluders with position as the control variable, occluders with body type as the control variable and effects of synthesised versus recorded audio trials. We provide additional results pertaining to effects of practice, *distractor* presentation and effect of occluder gender. Finally, we provide analysis of presence trials which were presented without visual elements as a pure listening test.

5.1 Measurement data on body type, position and gender

An examination of the audio data capture suggests that there are differences in the degree of occlusion not only between male and female genders, but also body types. There was an expected drop off in high frequencies in all measurements, particularly above 2.5kHz bands. Frequencies below 2kHz are less effected, but we do see some increased response in bands around 250-315Hz, particularly regarding the chest microphone measurements.

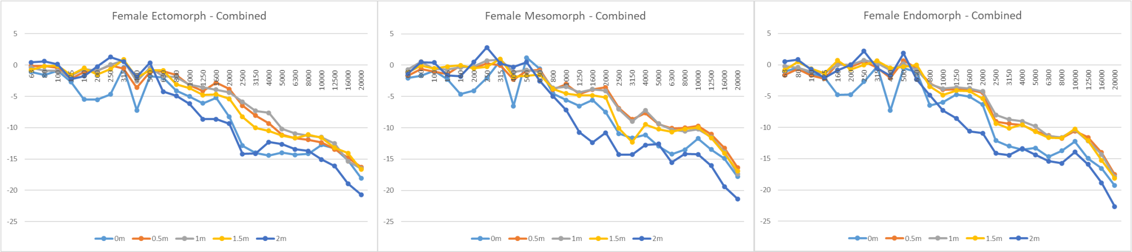
We also see from the measurements that the positions in closest proximity to the listener position and sound source create stronger attenuation at higher frequencies as well as intensification of responses in lower bands. The combined male versus combined female results show this more strongly for female participants. The combined results are derived from three of each body type for both male and female, so nine male participants for the combined male reading and nine for female. This was to ensure equal weighting of the results.

For body type we compare a combined 18 total participants, three male and three female per type to ensure equal weighting of each result. Of the three types, mesomorph shows the flattest results across each measurement. Endomorph and ectomorph both show intensification of response around the 200-315Hz at the head and chest measurements when not immediately at the listener position, as well as the 315-500Hz bands in the knee results. At the closest to listener position (0m), the attenuation was noticeably stronger across all body types in the 200-315Hz bands.



(a) (b) (c)

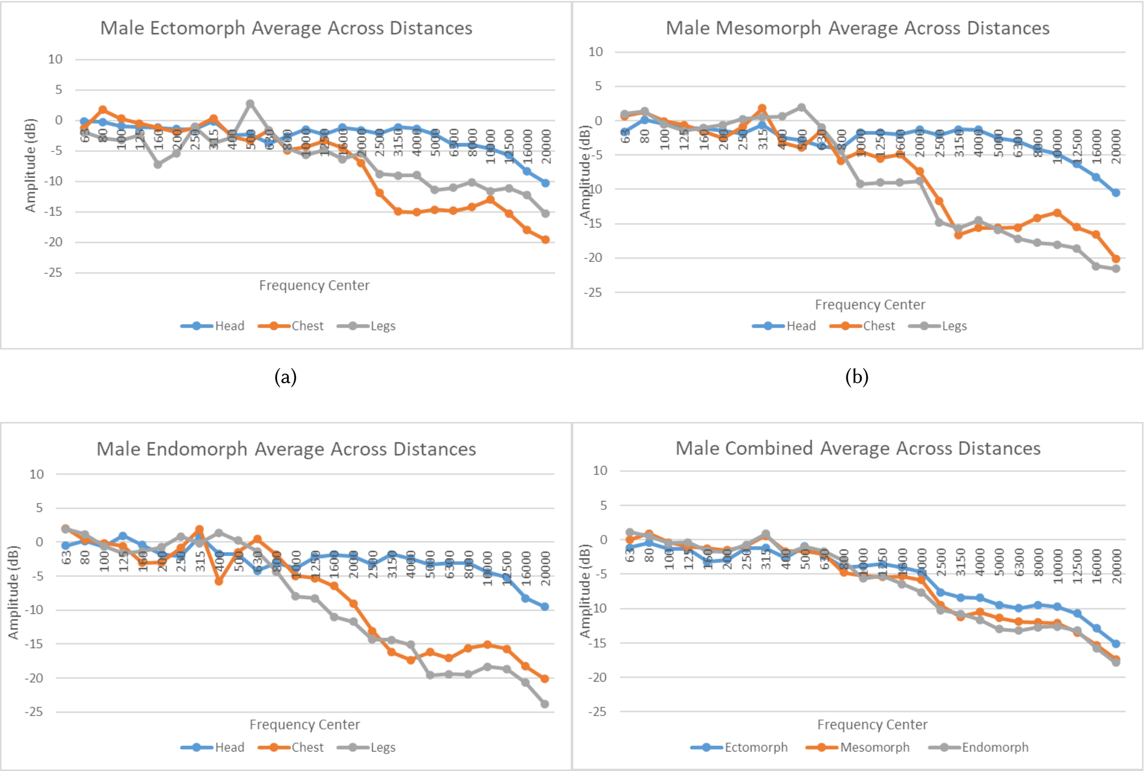
Fig. 6. Measurement data of male occluders. Three body types are shown (Ectomorph, Mesomorph, Endomorph) with measurements taken at distances from the microphone. Points are combined averages of head, chest and leg microphone measurements.



(a) (b) (c)

Fig. 7. Measurement data of female occluders. Three body types are shown (Ectomorph, Mesomorph, Endomorph) with measurements taken at distances from the microphone. Points are combined averages of head, chest and leg microphone measurements.

To investigate the nature of the changes across body type and distance, we also looked at averages across distance, body type and gender, such as with the male measurements in Fig. 6 and Fig. 8.

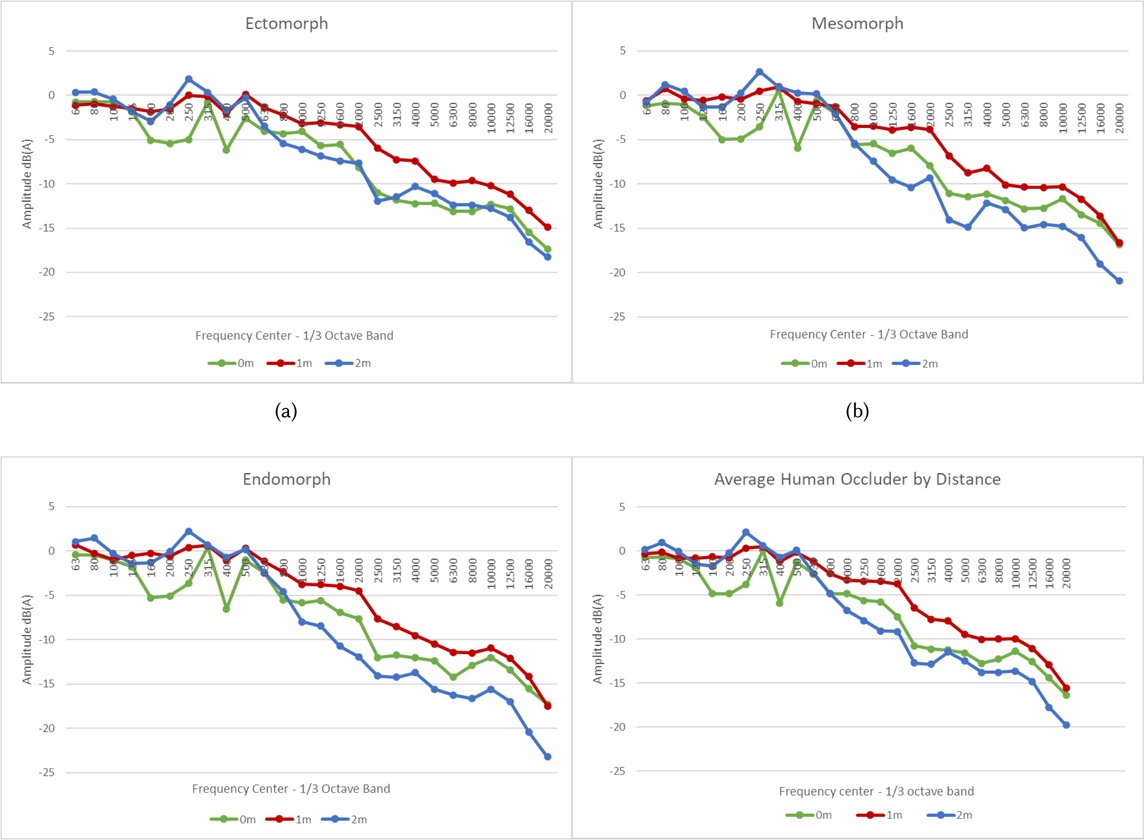


(a) (b)

(c) (d)

Fig. 8. Averaged values across all distance measurements for male body types at head, chest and legs. Images (a),(b) and (c) show average values across distance for each microphone in the array for the three male measured body types (Ectomorph, Mesomorph and Endomorph). Image (d) illustrates the combined average of the microphones for each body type.

We also see from the combined male and female data that there are differences in the profiles for body types (see Fig. 9). Particularly in frequencies above 500Hz we see divergence between the profiles for body types. Noticeably, the 1m readings are the most similar, but at 0m (closest to the microphones) and 2m (closest to the loudspeaker) we can see these results differ significantly. It should be noted that the perception study focused only on ectomorph and endomorph body types, as well as male and female genders in order to reduce the number of required trials. As such, we chose to focus on these variables at opposing ends of the spectra.



(a) (b)

(c) (d)

Fig. 9. Combined average of microphones for each body type (combined male and female). Images (a),(b) and (c) show average values across distance for the three measured body types (Ectomorph, Mesomorph and Endomorph). Image (d) illustrates the combined average for each distance measured.

5.2 Online Perception Study

The following section reports the analysis of data from 64 participants in our Two-Alternative Forced Choice trials. In what follows, *target* refers to the correct alternative in the trial, i.e., the image that corresponded to the audio being played, while *distractor* refers to the other, incorrect image. Unless otherwise indicated, binomial models and binomial generalised mixed-effect models with the outcome variable being whether participants correctly identified the *distractor* image corresponding to the audio and variable, and participant as random effect were conducted. The package lme4 [4] in the software R [70] was used to fit the models. The dashed line in figures represents the response rate predicted by chance.

In the non-image trails, presence was measured in two ways. First, to probe sense of presence, participants were asked to answer the question "I feel a sense of presence when listening to this audio" on a 7-point Likert scale from 1 = disagree to 7 = agree. (Fig. 10, left). A linear regression returned a significant effect of actual presence of an occluder correctly predicting Rated Presence (Likert) (β=1.69, p=0.000) and marginal R^2=0.16. Second, participants rated confidence of their perception of presence using a continuous slider with values (from -100 (completely sure there was no-one in the room) to +100 (completely sure there was someone in the room) (Fig. 10, right)). A binomial generalised mixed-effect model with Participant as the random effect and Rated Presence (Continuous) as the outcome variable returned a significant effect of presence on ratings (β=72.26, t=24.12, p=0.000), and marginal R^2=0.28. Both measures show that presence is clearly detected from audio.

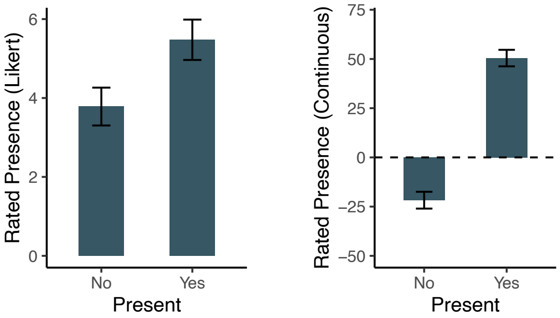


Fig. 10. Presence ratings. Left: Certainty of presence measured in a continuous scale from -100 to +100. Right: Sense of presence measured in a 7-point Likert scale.

Can participants correctly detect body type and position from audio input? Descriptive statistics are shown in Fig. 11. In the remainder of this section, unless otherwise indicated, binomial models and binomial generalised mixed-effect models with participants as random effects were conducted, with proportion of correct responses as the outcome variable.

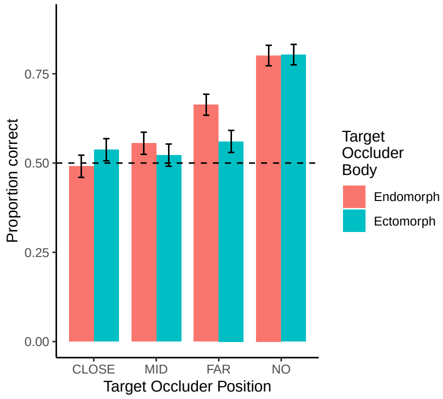
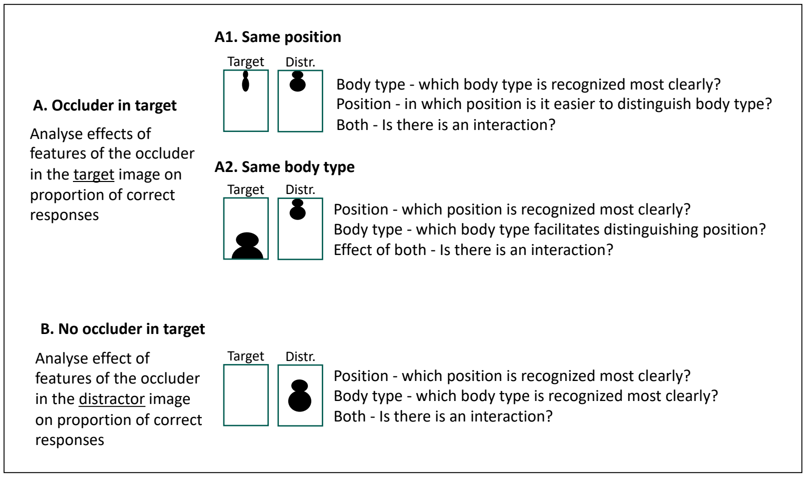


Fig. 11. Proportion of correct answers by position and body type of the occluder in the *distractor* image. Significant results were recorded for *endomorph* in *mid, far* and *no* occlusion trials. *Ectomorph* was significantly correct in *close, far* and *no* occluder trails.

From this data we can extract a third measure of detection of presence, this time on response Correctness in the forced-choice trials. A linear model revealed that when an occluder was present in any position (*close*, *mid*, or *far*), responses were significantly more correct than when there was no occluder present (β=-1.18, p=0.000). As seen in Fig. 11 presence is main influence on response correctness. To examine in detail to what extent position and body type are correctly inferred from the audio, we analysed the trials with an occluder in the audio (positions *close*, *mid* and *far*) and without one (position *no*) separately (Fig. 12).



A. Occluder in target

Analyse effects of features of the occluder in the target image on proportion of correct responses

A1. Same position

Body type - which body type is recognized most clearly?
Position - in which position is it easier to distinguish body type?
Both - Is there is an interaction?

A2. Same body type

Position - which position is recognized most clearly?
Body type - which body type facilitates distinguishing position?
Effect of both - Is there is an interaction?

B. No occluder in target

Analyse effect of features of the occluder in the distractor image on proportion of correct responses

Position - which position is recognized most clearly?
Body type - which body type is recognized most clearly?
Both - Is there is an interaction?

Fig. 12. Analyses performed to determine to what extent occluder position and body type are correctly inferred from the recorded audio.

5.3 Trials when an occluder was present in the target image

Effect of occluder body type (controlling for position). Considering only trials in which there was an occluder in the *distractor* image, to isolate the effect of occluder body type, we analysed only the trials where the occluder was in the same position in the *target* and *distractor* images. We found that *endomorph* occluders were detected more correctly than *ectomorph* (β=-0.31, p=0.003) (Fig. 13, left), but no effect of ordinal position on response correctness (β=0.04, p=0.53) (Fig. 13, middle). Regarding an interaction between position and size (Fig. 13, right), models with and without the interaction term were compared with an ANOVA test, which returned a strong interaction (X^2 = 36.13, p=0.000). This is driven by a trade-off between body type and position. *Endomorph* body type was correctly identified more often when they were further from the speaker, and *ectomorph* body types when they were closer to the speaker. In particular, *ectomorph* in the *far* position was often mistaken for *endomorph* body type (Fig. 13, right).

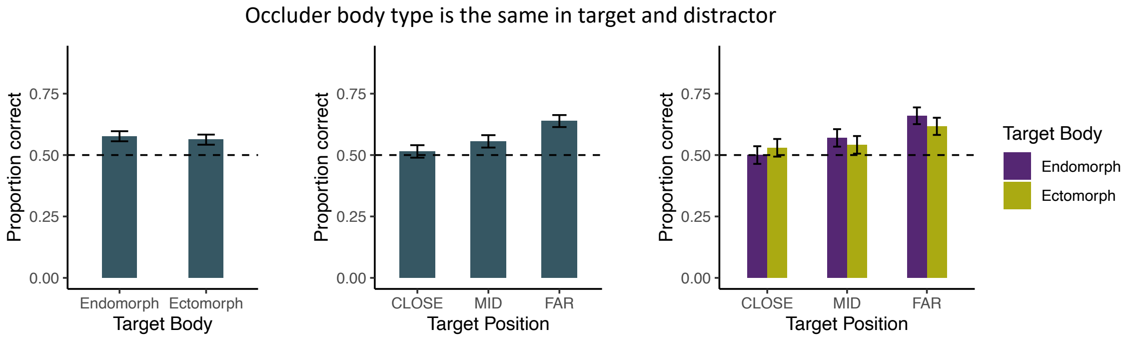


Fig. 13. Proportion of answers correct in trials with same occluder position in the *target* and *distractor* images, broken down by features of the occluder in the *target* image. Left: By *target* occluder body type, significantly correct for both. Centre: By *target* occluder position, significant in *mid* and *far* positions. Right: By *target* occluder body type and position. Only *mid* and *far* position show significant results for both body types.

Effect of occluder position (controlling for body type). Considering only trials in which an occluder was present in the *target* image, to isolate the effect of occluder position we analysed only the trials where the occluders in the *target* and *distractor* images had the same body type. Body type did not influence correctness (β=-0.06, p=0.35) (Fig. 14, left), but ordinal position did have an effect, with occluders in the *far* position being detected most correctly (β=0.25, p=0.000) (Fig. 14, centre). There was no interaction between *position* and *body type* (Fig. 14, right) (X^2=4.70, p=0.09).

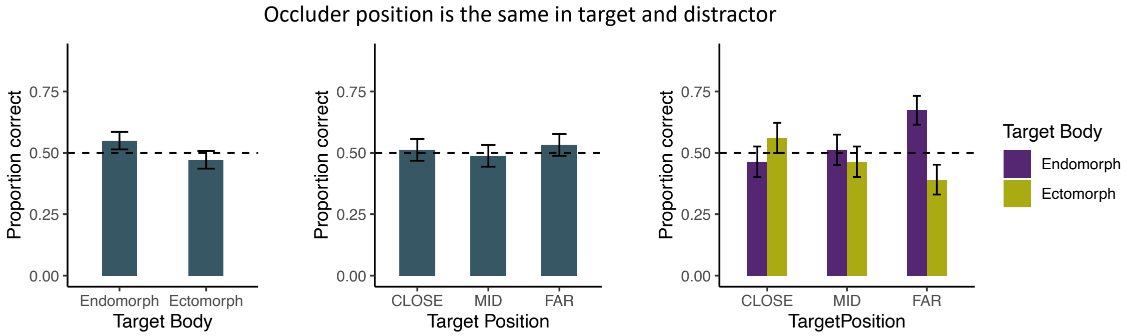


Fig. 14. Proportion of answers correct in trials with same occluder body type in the *target* and *distractor* images, broken down by features of the occluder in the *target* image. Left: By *target* occluder body type, *endomorph* is correct at siginificantly better than chance. Centre: By *target* occluder position, no significant results. Right: By *target* occluder position and body type.

In trials when there was no occluder in the *target* image (effects of *distractor* image). Considering only trials with no occluder in the *target*, we tested the effect of the body type in the *distractor* image on correctness of responses (Fig. 15, left). A mixed-effects model found no effect of body type (β=0.02, p=0.88). To test the effect of position of the occluder in the *distractor* (Fig. 15, middle), the categorical variable *distractor* position was made ordinal. A mixed-effects linear model found no differences between of positions on correct responses (β=0.10, p=0.23). No significant interactions were found either for the *endomorph* (β=0.07, p=0.40) or the *ectomorph* body type (β=0.13, p=0.16) (Fig. 15, right).

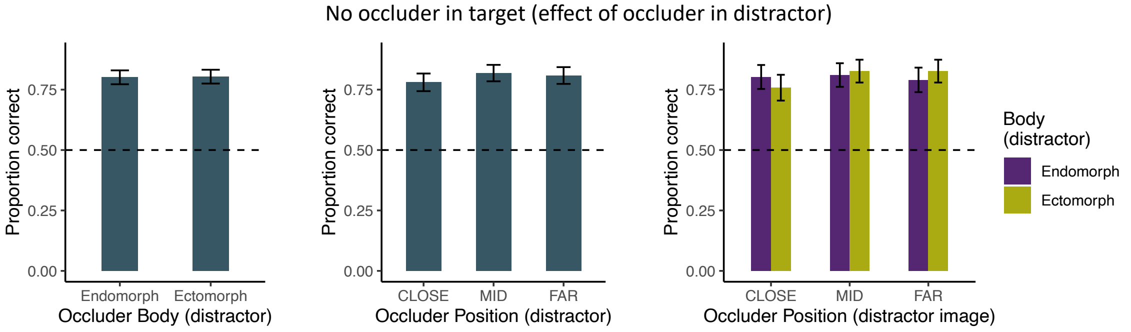


Fig. 15. Proportion of answers correct in trials where the *target* image did not include an occluder, broken down by features of the occluder in the *distractor* image: Left: By occluder body type. Middle: By occluder position. Right: By occluder body type and position.

5.4 Does audio type (synthesised or recorded) help detection of presence, occluder body type and occluder position?

To test the effect of audio type we considered trials from the presence study when there was an occluder present in the *target* only (in trials where there was no occluder, synthesised data was not tested).

Effects of Audio type on detection of occluder presence. A linear model with audio type as the predictor and presence (continuous) as the outcome variable returned significantly higher presence ratings for the synthesised audio (β=46.65, t=13.47, p=0.000), marginal R^2=0.16, indicating that synthesised audio led participants to rate presence more accurately than recorded audio.

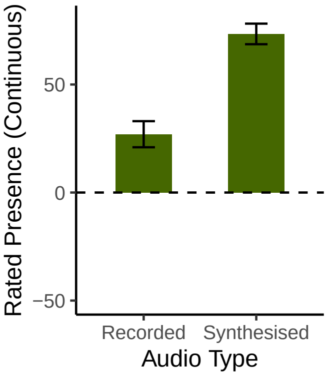


Fig. 16. Rated presence (continuous) in trials where an occluder was present, by audio type.

Effects of Audio type on detection of occluder body type and position. A binomial model with two interactions as predictors: audio type x occluder position (ordinal), and audio type x occluder body type, and with proportion of correct responses as the outcome variable returned R^2=0.02; a significant interaction between the audio type and occluder body type (β=-0.33, p=0.002); and a significant interaction between audio type and occluder position (ordinal) (β=-0.18, p=0.006). This indicates that with recorded audio, detection of position increases with distance of the occluder from the participant, and with synthesised audio it decreases, but only for the *ectomorph* body type (Fig. 17).

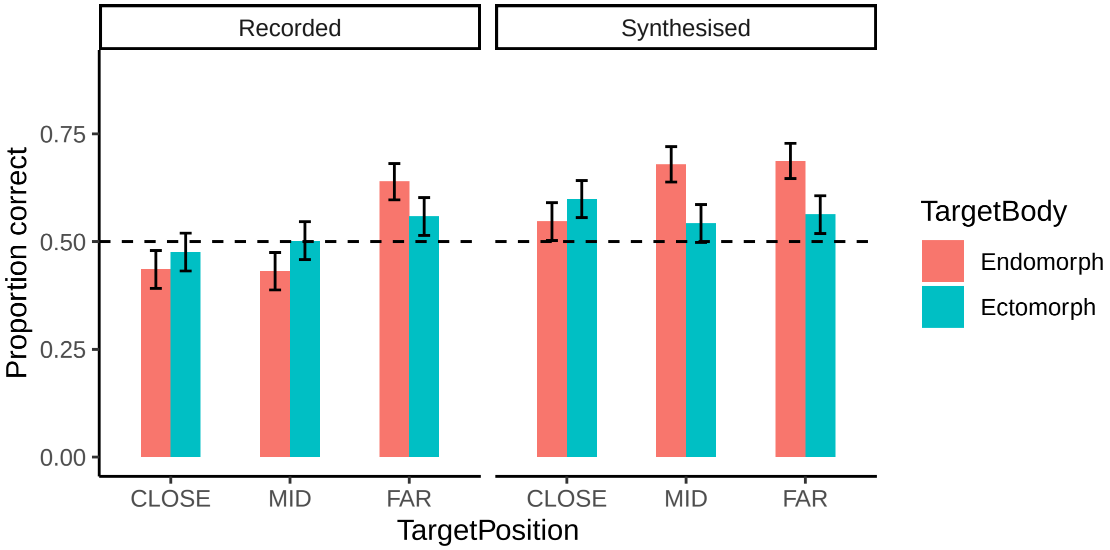


Fig. 17. Proportion of correct responses by audio type, occluder body and occluder position. Recorded audio resulted in incorrect responses in *close* and *mid* position for *endomorph* body types. Synthesised audio resulted in higher correct response for all conditions.

Overall, synthesised audio input obtained more accurate perceptions of occluder size and position than recorded input. This difference was most notable for *endomorph* occluders in *close* and *mid* positions. Indeed, *endomorph* in *close* and *mid* positions were predicted worse than expected by chance with the recorded input, but better than chance with synthesised input.

5.5 Additional results

Some additional exploratory tests were carried out to determine other impacts on responses.

Does participant's expertise with audio improve detection of presence, occluder body type and occluder position? The participant sample was very skewed in favour of high-expertise levels (1), which advises against further meaningful analysis. Further information on detection of presence and occluder size can be found in SM.x

Expertise	None	Beginner	Amateur	Semi-Pro	Professional	Not Known	Total
N	4	3	2	10	43	2	64

Table 1. Participant level of audio expertise.

Effect of practice. To test whether there was an effect of practice–whether there were more correct responses in later trials–, a generalised linear model was conducted with Correct (yes or no) as the outcome variable and trial number as the predictor. The results indicate an effect of practice (β=0.001, p=0.049).

Effect of target presentation side (left or right). To test whether there was an effect of *target* position (left or right), a generalised linear model was conducted with Correct (yes or no) as the outcome variable and *target* position (left or right) as the predictor. The results indicate no effect of image position (β=-0.02, p=0.83).

Effect of occluder gender. To test whether there was an effect of gender of the occluder (male or female) a generalised linear model was conducted with correct (yes or no) as the outcome variable and occlude gender as predictor. The results indicate no effect of gender (β=-0.09, p=0.07).

6 DISCUSSION

It would appear from the measurement readings that there is nuance to the attenuation of high frequencies when a person acts as an occluder (Fig. 6). The results show differences across bands due to gender and body type, as well as distance from the listener and sound source. Of particular interest are the changes in close proximity to the listener. In scenarios such as crowded gig venues or more intimate virtual environments with a partner or small group of users, this colouration of sound could have a noticeable impact on the listening experience, particularly for low frequencies [80]. We see that close proximity results in more severe attenuation at higher frequencies, but also typically the lower bands with frequency center's around 200Hz.

In Fig. 12 we see that the most noticeably correct response is for *endomorph* body type at the position closest to the loudspeaker. We also see that *ectomorph* body type is the most often mistaken for the *endomorph* body type at the *far* position, possibly indicating bias due to the visual elements of the trials. *Ectomorph* body types are most accurate in the *close* position. This might indicate that the effect of occlusion is less obvious in this position, but participants are still able to determine that occlusion is occurring in that trial. The *mid* position trials show a chance close to random for both body types, indicating that this position causes the most confusion amongst participants. This implies a less clear effect of auditory occlusion at this midpoint.

Synthesised input provided more accurate perceptions of occluder body type and position than recorded input (see Fig. 17). This could be due to some loss of complexity in the audio processing of the synthesised media, but this is unclear and would require further study. This difference was most notable for *endomorph* occluders in *close* and *mid* positions. Indeed, *endomorph* occluders in *close* and *mid* positions were predicted worse than expected by chance with the *recorded* input, but better than chance with *synthesized* input. We actually see that for *endomorph* in the *close* and *mid* positions, participants incorrectly identified the body type as *ectomorph* at worse rates than expected by chance (50% dotted line) when presented by recorded input. This means that participants were incorrectly identifying *endomorph* body types as *ectomorph* in the *close* and *mid* positions. This implies that the recorded audio caused the impression of weaker occlusion, but this is being partially corrected with the synthesised audio.

Certain variables of this study indicate no significant perceivable results, such as in the case of gender. However, distance to sound source and listener show significant results (β=0.25, p=0.000) (Fig. 14), as well as in the case of having no occluder (Fig. 15). Scenarios with the occluder closest to the sound source had the most significant results, whilst the mid-point position had the lowest accuracy. This indicates that proximity to listener or sound source has significant impact on the listeners ability to discern occlusion.

When asked to indicate perception of presence, participants indicated strong responses when someone was occluding the sound source (Fig. 16). There was less certainty in participant responses when no occluder was present. This indicates that participants could hear when someone is occluding a sound source and are more sure that there is occlusion taking place.

When no image was presented, participants in these trials were able to audibly perceive the presence of another person in the space between them and the loudspeaker. The results show a strong response to trials in which a person was present (see Fig. 10). The results for distance show higher accuracy of results at the *close* and *far* positions versus *mid*. This indicates that the occlusion effect was stronger at close proximity's to listener or sound source, with less certainty in the *mid* position, potentially due to a higher mix of reflections from the room. The strongest indicator to a person causing occlusion was in the *far* position - closest to the sound source - indicating that this is where the strongest occlusion effect takes place.

Gender had no effect on the results, and was therefore shown to not impact the perception of audio occlusion in this scenario. Participants were better able to recognise occlusion over time as they completed trials, indicating that participants are capable of learning what an occluded track sounds like, but the specific nature of what they are listening for in such cases was unclear and would require further study. The randomised nature of the *target* and *distractor* image positions was effective and showed no significant impact on the results. In Fig. 15 we can see that despite varying the *distractor* images between position and body type, there was no impact on correct responses for scenarios where no audio occlusion took place.

In Fig. 13 we see that *endomorph* body type closest to the loudspeaker had the highest correct response rate, indicating a stronger audio occlusion effect than in other positions. Interestingly, at the *close* position (closest to the listener position), the *ectomorph* body type had more accurate responses. This also suggests that the audio occlusion was more noticeable at the *far* position (closest to the loudspeaker) which was possibly being conflated with the *endomorph* occluding body type, as indicated by the level of incorrect response for *ectomorph* in the *far* position. Fig. 14 indicates that this position was most accurately judged when controlling for body type. Therefore the body type of the occluding person was not a significant variable for audio occlusion in these trials.

6.1 Further Work

As the measured data shows that higher frequencies are the most heavily attenuated, it may be reasonable to assess the perception of auditory occlusion against the background of presbycusis [95] where subtle changes in upper frequency content may become more difficult to discern. Our sample data is heavily skewed in favour of those with a high level of audio expertise. It may be the case that audio professionals have better listening training in versus the general population which therefore allows them to notice smaller shifts in frequency content [12]. However, broader studies would need to be carried out in order to determine if this is the case.

It could be argued that the stereo recorded audio caused confusion in the results as a result of the KU-100's HRTF profile due to listener mismatch [64]. In this case, it may be the case that the measurement data gave an exaggerated effect of occlusion and this presents a more noticeable effect for listeners versus a more acoustically accurate recording. Mono source capture may be advisable for future testing purposes. Further work may also include examining participant preferences in such cases.

The current work only explores auditory occlusion in the direct path of sound due to the human body. It would be beneficial to carry out similar work to examine changes in general absorption and frequency content for the human body, considering the full frequency spectrum. We plan to carry out such work in the next stage of this research. Such work is often difficult to conduct in occupied spaces, and previous work has focused on simulations to determine such estimates [7], particularly with regards to passive obstructions [84].

7 CONCLUSION

We conducted a survey examining the auditory occlusion caused by a human on the perception of someone being present in a shared space. Results indicate that listeners are able to correctly identify the auditory occlusion caused by someone positioned between them and a sound source with above chance accuracy. We show that gender of the occluding person had no significant impact on this perception in these trails, but the occluders' distance from the listener or sound source does have a significant impact on accuracy, as does the occluder body type in some instances. We also show that in listening-only tests, participants were highly accurate in judging the presence of another person. In scenarios where the occluder was in close proximity to the listener or the sound source, the ability to judge the presence of another person in the room was more accurate than at the midpoint between listener and speaker. Our results also provide evidence that synthesised audio occlusion effects, as measured and processed according to our mono capture source, are able to provide a degree of correction for confusion caused by binaural recordings of real occlusion with regards to occluder body type.

We conclude that listeners are able to determine the presence of an individual in a shared space due to their occlusion of a sound source. This ability to recognise when someone is in a shared space indicates a need for such systems in MR applications where users share a virtual space - such as concerts, social gatherings or multiplayer co-operative or competitive games. However, use of audio in MR for social experiences is under-researched [8]. We can determine from our results that while certain aspects of an avatar - such as gender - should not make material changes to the sound occlusion, other aspects such as proximity and overall body type should be considered.

We argue that the sound occlusion caused by people is an important aspect of real world listening environments that should be replicated in MR applications for player and non-player avatars to improve the plausibility of shared virtual environments. The ability for listeners to recognise the presence of another individual in a virtual space through the effect of sound occlusion makes the case that such a system could aid in the improvement of immersion in MR. The specific effects of visual cues on auditory processing are unknown, with some work showing confusion due to mismatch of visual environments in VR [74]. Although precise localisation was not the purpose of this study, we do show that the auditory occlusion caused by a person was strongly recognised in both visual and non-visual tests. In reality, people are acutely attuned to the presence of other people through various sound cues. This work makes a strong case that auditory occlusion caused by humans is one such cue that should be considered and implemented in the development of MR applications in order to increase the plausibility of the space and the presence of others, be they player or non-player controlled avatars.

ACKNOWLEDGMENTS

CAROUSEL+ is research and innovation project number 101017779, funded under the European Horizon 2020 FET Proactive program.

REFERENCES

[1] M. Yu. Agaeva and E. A. Petropavlovskaia. 2023. Localization of Correlated and Uncorrelated Audio Signals in the Horizontal Plane under Masking Conditions. *Human Physiology* 49, 1 (Feb. 2023), 44–54. https://doi.org/10.1134/S0362119722100012

[2] Areti Andreopoulou and Brian F. G. Katz. 2022. Perceptual Impact on Localization Quality Evaluations of Common Pre-Processing for Non-Individual Head-Related Transfer Functions. *Journal of the Audio Engineering Society* 70, 5 (May 2022), 340–354. https://doi.org/10.17743/jaes.2022.0008

[3] Audio-Technica. 2023. AT831. https://www.audio-technica.com/en-gb/at831

[4] Douglas Bates, Martin Maechler, Ben Bolker [aut, cre, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai, Fabian Scheipl, Gabor Grothendieck, Peter Green, John Fox, Alexander Bauer, Pavel N. Krivitsky (shared copyright on simulate.formula), Emi Tanaka, and Mikael

Jagan. 2024. lme4: Linear Mixed-Effects Models using 'Eigen' and S4. https://cran.r-project.org/web/packages/lme4/index.html

[5] Mirza Beig, Bill Kapralos, Karen Collins, and Pejman Mirza-Babaei. 2019. G-SpAR: GPU-Based Voxel Graph Pathfinding for Spatial Audio Rendering in Games and VR. In *2019 IEEE Conference on Games (CoG)*. 1–8. https://doi.org/10.1109/CIG.2019.8847959 ISSN: 2325-4289.

[6] Mirza Beig, Bill Kapralos, Karen Collins, and Pejman Mirza-Babaei. 2019. An Introduction to Spatial Sound Rendering in Virtual Environments and Games. *The Computer Games Journal* 8, 3 (Dec. 2019), 199–214. https://doi.org/10.1007/s40869-019-00086-0

[7] M.L. Benferhat, S. Debache Benzagouta, A. Bouttout, and F. Martellotta. 2024. On the simulation of occupied acoustic conditions of Djedid mosque in Algiers. In *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*. European Acoustics Association, Turin, Italy, 2553–2557. https://doi.org/10.61782/fa.2023.0963

[8] Isak De Villiers Bosman, Oğuz 'Oz' Buruk, Kristine Jørgensen, and Juho Hamari. 2023. The effect of audio on the experience in virtual reality: a scoping review. *Behaviour & Information Technology* (Jan. 2023), 1–35. https://doi.org/10.1080/0144929X.2022.2158371

[9] Hark Simon Braren and Janina Fels. 2021. Towards Child-Appropriate Virtual Acoustic Environments: A Database of High-Resolution HRTF Measurements and 3D-Scans of Children. *International Journal of Environmental Research and Public Health* 19, 1 (Dec. 2021), 324. https://doi.org/10.3390/ijerph19010324

[10] Albert S. Bregman. 1990. *Auditory scene analysis : the perceptual organization of sound*. MIT Press.

[11] Douglas S. Brungart, Mary E. Barrett, Julie I. Cohen, Calli Fodor, Calli M. Yancey, and Sandra Gordon-Salant. 2020. Objective Assessment of Speech Intelligibility in Crowded Public Spaces. *Ear and Hearing* 41, Suppl 1 (Oct. 2020), 68S–78S. https://doi.org/10.1097/AUD.0000000000000943

[12] Francesco Caprini, Sijia Zhao, Maria Chait, Trevor Agus, Ulrich Pomper, Adam Tierney, and Fred Dick. 2024. Generalization of auditory expertise in audio engineers and instrumental musicians. *Cognition* 244 (March 2024), 105696. https://doi.org/10.1016/j.cognition.2023.105696

[13] Cauldron Science. 2024. Gorilla Experiment Builder - Easily Create Online Behavioural Experiments. https://gorilla.sc/

[14] N.M.D. Chagok, L.D. Domtau, E. Agoyi, and E.E. Akpan. 2013. Acoustic Absorption Coefficient of Human Body in Octave Bands. https://www.auralexchange.com/knowledgebase/acoustic-absorption-coefficient-of-human-body-in-octave-bands/

[15] Panagiotis Charalampous and Despina Michael. 2014. Sound propagation in 3D spaces using computer graphics techniques. In *2014 International Conference on Virtual Systems & Multimedia (VSMM)*. 43–49. https://doi.org/10.1109/VSMM.2014.7136674

[16] Prasun Chatterjee and Santosh K. Yatnatti. 2020. Intergenerational Digital Engagement: A Way to Prevent Social Isolation During the COVID-19 Crisis. *Geriatric Research Education and Clinical Center* 68, 7 (2020). https://doi.org/10.1111/jgs.16563

[17] E. Colin Cherry. 1953. Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America* 25, 5 (Sept. 1953), 975. https://doi.org/10.1121/1.1907229 Publisher: Acoustical Society of AmericaASA.

[18] Eike Claaßen, Stephan Töpken, and Steven van de Par. 2024. The influence of the reference level on loudness and preference judgements for spectrally manipulated fan sounds. *The Journal of the Acoustical Society of America* 155, 3 (March 2024), 1735–1746. https://doi.org/10.1121/10.0025161

[19] Cockos, inc. 2024. REAPER | Audio Production Without Limits. https://www.reaper.fm/

[20] CSCS, Helen Kollias, and Ryan Andrews. 2009. Body type dieting for ectomorphs, endomorphs, and mesomorphs: Is it right for you? https://www.precisionnutrition.com/all-about-body-type-eating

[21] Anton J. M. Dijker. 2008. Why Barbie feels heavier than Ken: The influence of size-based expectancies and social cues on the illusory perception of weight. *Cognition* 106, 3 (March 2008), 1109–1125. https://doi.org/10.1016/j.cognition.2007.05.009

[22] DOIT. 2020. A Growth of Accessibility in Video Games | DO-IT. https://www.washington.edu/doit/growth-accessibility-video-games

[23] Kevin Donahue. 2021. Layering Ultralight Clothing for Pilgrimage Backpacking | One Step Then Another. https://www.onestepthenanother.com/caminodesantiago/layering-ultralight-clothing-for-pilgrimage-backpacking/

[24] Jacob Donley, Vladimir Tourbabin, Jung-Suk Lee, Mark Broyles, Hao Jiang, Jie Shen, Maja Pantic, Vamsi Krishna Ithapu, and Ravish Mehra. 2021. EasyCom: An Augmented Reality Dataset to Support Algorithms for Easy Communication in Noisy Environments. http://arxiv.org/abs/2107.04174 arXiv:2107.04174 [cs, eess].

[25] Frank Dufour. 2011. Acoustic Shadows: An Auditory Exploration of the Sense of Space. *SoundEffects - An Interdisciplinary Journal of Sound and Sound Experience* 1, 1 (Dec. 2011), 82–97. https://doi.org/10.7146/se.v1i1.4074 Number: 1.

[26] Rachel Elliott. 1998. The Use of Dance in Child Psychiatry. *Clinical Child Psychology and Psychiatry* 3, 2 (1998), 251–265. https://doi.org/10.1177/1359104598032008

[27] Débora Franco, Isabel Fragoso, Mário Andrea, Júlia Teles, and Fernando Martins. 2017. Somatotype and Body Composition of Normal and Dysphonic Adult Speakers. *Journal of Voice* 31, 1 (Jan. 2017), 132.e9–132.e21. https://doi.org/10.1016/j.jvoice.2015.11.020

[28] FreePD. 2024. FreePD.com - Free Public Domain Music Creative Commons 0 Completely Royalty Free. https://freepd.com/

[29] Di Fu, Cornelius Weber, Guochun Yang, Matthias Kerzel, Weizhi Nan, Pablo Barros, Haiyan Wu, Xun Liu, and Stefan Wermter. 2020. What Can Computational Models Learn From Human Selective Attention? A Review From an Audiovisual Unimodal and Crossmodal Perspective. *Frontiers in Integrative Neuroscience* 14 (2020). https://www.frontiersin.org/articles/10.3389/fnint.2020.00010

[30] Genelec. 2013. 8030A - Genelec.com. https://www.genelec.com/previous-models/8030a

[31] Michele Geronazzo and Stefania Serafin (Eds.). 2023. *Sonic Interactions in Virtual Environments*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-031-04021-4

[32] Debby Fernando Giovanni and Anny Valentina. 2023. Creative Strategy: An Introduction to Recognize Body Type for Early Adolescents. *International Journal of Application on Social Science and Humanities* 1, 1 (Feb. 2023), 1106–1116. https://doi.org/10.24912/ijassh.v1i1.25721

[33] U. K. Government. 2020. Health and Safety at Work etc. Act 1974. (Feb. 2020). https://www.hse.gov.uk/legislation/hswa.htm Publisher: Statute Law Database.

[34] Georgina Guillen, Henrietta Jylhä, and Lobna Hassan. 2021. The Role Sound Plays in Games:: A Thematic Literature Study on Immersion, Inclusivity and Accessibility in Game Sound Research. In *Academic Mindtrek 2021*. ACM, Tampere/Virtual Finland, 12–20. https://doi.org/10.1145/3464327.3464365

[35] Charanya Gurusathya. 2019. Dance as a Catalyst for Stress Busting. *Central European Journal of Sport Sciences and Medicine* 26, 2 (2019), 15–29. https://bibliotekanauki.pl/articles/1030676

[36] Dillon A. Hambrook, Marko Ilievski, Mohamad Mosadeghzad, and Matthew Tata. 2017. A Bayesian computational basis for auditory selective attention using head rotation and the interaural time-difference cue. (2017). https://doi.org/10.1371/journal.pone.0186104 ISBN: 1111111111.

[37] Shangchen Han, Facebook Reality Labs BEIBEI LIU, Facebook Reality Labs ROBERT WANG, Facebook YE Reality Labs YUTING, Facebook D. Reality Labs CHRISTOPHER TWIGG, Beibei Liu, Robert Wang, Yuting Ye, and Christopher D. Twigg. 2018. Online Optical Marker-based Hand Tracking with Deep Labels. *ACM Trans. Graph* 37, 1 (2018). https://doi.org/10.1145/3197517.3201399

[38] HTC Corporation. 2023. VIVE United Kingdom | Discover Virtual Reality Beyond Imagination. https://www.vive.com/uk/

[39] Ashlee Humphries, Rachel Rugh, Morgan Patrick, and Julia C. Basso. 2022. *Enhancing Mental Health and Social Connection Through an Acute Online Dance Intervention.* preprint. In Review. https://doi.org/10.21203/rs.3.rs-1149930/v1

[40] Lars E F Johannessen, Eivind Engebretsen, Trisha Greenhalgh, Gemma Hughes, Julia Köhler-Olsen, Erik Børve Rasmussen, and Marit Haldar. 2021. Protocol for 'virtual presence': a qualitative study of the cultural dialectic between loneliness and technology. *BMJ Open* 11, 9 (Sept. 2021), e047157. https://doi.org/10.1136/bmjopen-2020-047157

[41] Oren Kadosh and Yoram S. Bonneh. 2022. Involuntary oculomotor inhibition markers of saliency and deviance in response to auditory sequences. *Journal of Vision* 22, 5 (April 2022), 8. https://doi.org/10.1167/jov.22.5.8

[42] B. F. Katz. 2000. Acoustic absorption measurement of human hair and skin within the audible frequency range. *The Journal of the Acoustical Society of America* 108, 5 Pt 1 (Nov. 2000), 2238–2242. https://doi.org/10.1121/1.1314319

[43] Jonathan W. Kelly, Taylor A. Doty, Morgan Ambourn, and Lucia A. Cherep. 2022. Distance Perception in the Oculus Quest and Oculus Quest 2. *Frontiers in Virtual Reality* 3 (2022). https://www.frontiersin.org/articles/10.3389/frvir.2022.850471

[44] Benjamin Kenwright. 2020. There's More to Sound Than Meets the Ear: Sound in Interactive Environments. *IEEE Computer Graphics and Applications* 40, 4 (July 2020), 62–70. https://doi.org/10.1109/MCG.2020.2996371 Conference Name: IEEE Computer Graphics and Applications.

[45] Luchcha Lam, Minsoo Choi, Magzhan Mukanova, Klay Hauser, Fangzheng Zhao, Richard Mayer, Christos Mousas, and Nicoletta Adamo-Villani. 2023. Effects of Body Type and Voice Pitch on Perceived Audio-Visual Correspondence and Believability of Virtual Characters. In *ACM Symposium on Applied Perception 2023*. ACM, Los Angeles CA USA, 1–11. https://doi.org/10.1145/3605495.3605791

[46] Kyulin Lee, Eugene Lee, Haeli Park, and Gilsoo Cho. 2013. Development of a Fabric Friction Sound Generator Simulating Body Movement and Evaluation of the Generated Sound. *Textile Science and Engineering* 50, 4 (2013), 241–246. https://doi.org/10.12772/TSE.2013.50.241 Publisher: The Korean Fiber Society.

[47] Orel Levy, Adi Korisky, Yair Zvilichovsky, and Elana Zion Golumbic. 2024. An ecological investigation of the effect of background noise on speech processing in a Virtual Classroom. https://doi.org/10.1101/2024.01.05.574300

[48] Ji Liu, Qiaoyi Chen, and Jingxia Dang. 2021. Examining risk factors related to digital learning and social isolation: Youth visual acuity in COVID-19 pandemic. *Journal of global health* 11 (2021), 05020–05020. https://doi.org/10.7189/jogh.11.05020 Publisher: INT SOC GLOBAL HEALTH.

[49] Danielle Lottridge, Rebecca Weber, Eva-Rae McLean, Hazel Williams, Joanna Cook, and Huidong Bai. 2022. Exploring the Design Space for Immersive Embodiment in Dance. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Christchurch, New Zealand, 93–102. https://doi.org/10.1109/VR51125.2022.00027

[50] Divine Maloney and Guo Freeman. 2020. Falling Asleep Together: What Makes Activities in Social Virtual Reality Meaningful to Users. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Virtual Event Canada, 510–521. https://doi.org/10.1145/3410404.3414266

[51] Francesco Martellotta, Michele D'alba, and Sabrina Della Crociata. 2011. Laboratory measurement of sound absorption of occupied pews and standing audiences. *Applied Acoustics* 72, 6 (May 2011), 341–349. https://doi.org/10.1016/j.apacoust.2010.12.008

[52] Anabela Marto and Alexandrino Gonçalves. 2022. Augmented Reality Games and Presence: A Systematic Review. *Journal of Imaging* 8, 4 (March 2022), 91. https://doi.org/10.3390/jimaging8040091

[53] MathWorks. 2023. MATLAB & Simulink. https://matlab.mathworks.com/

[54] Ravish Mehra, Atul Rungta, Abhinav Golas, Ming Lin, and Dinesh Manocha. 2015. WAVE: Interactive Wave-based Sound Propagation for Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* 21, 4 (April 2015), 434–442. https://doi.org/10.1109/TVCG.2015.2391858 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[55] Meta. 2023. Meta Quest 3: New mixed reality VR headset – Shop now | Meta Store. https://www.meta.com/gb/quest/quest-3/

[56] John C. Middlebrooks, Jonathan Z. Simon, Arthur N. Popper, and Richard R. Fay. 2017. *The Auditory System at the Cocktail Party*. Vol. 60. Springer International Publishing. https://doi.org/10.1007/978-3-319-51662-2

[57] Jonas Moll, Eva-Lotta Sallnäs, Kerstin Severinson Eklundh, and Sten-Olof Hellström. 2013. The Effects of Audio and Haptic Feedback on Collaborative Scanning and Placing. (2013). https://doi.org/10.1093/iwc/iwt031

[58] Julian Moreira, Laetitia Gros, Rozenn Nicol, Isabelle Viaud-Delmon, Cécile Le Prado, and Stéphane Natkin. 2019. Binaural sound rendering improves immersion in a daily usage of a smartphone video game. In *EAA Spatial Audio Signal Processing Symposium*. Paris, France, 79–84.

https://doi.org/10.25836/sasp.2019.23

[59] Christopher H. Morrell, Sandra Gordon-Salant, Jay D. Pearson, Larry J. Brant, and James L. Fozard. 1996. Age- and gender-specific reference ranges for hearing level and longitudinal changes in hearing level. *The Journal of the Acoustical Society of America* 100, 4 (Oct. 1996), 1949–1967. https://doi.org/10.1121/1.417906

[60] Lina Motlagh Zadeh, Noah H. Silbert, Katherine Sternasty, De Wet Swanepoel, Lisa L. Hunter, and David R. Moore. 2019. Extended high-frequency hearing enhances speech perception in noise. *Proceedings of the National Academy of Sciences* 116, 47 (Nov. 2019), 23753–23759. https://doi.org/10.1073/pnas.1903315116 Publisher: Proceedings of the National Academy of Sciences.

[61] Music Vine Limited. 2024. Free Music For YouTube Videos & Creators • Uppbeat. https://uppbeat.io/

[62] Neumann. 2018. Neumann KU-100 Specs. https://www.neumann.com/en-en/products/microphones/ku-100/

[63] Abigail L. Noyce, Jasmine A. C. Kwasa, and Barbara G. Shinn-Cunningham. 2021. Defining attention from an auditory perspective. *WIREs Cognitive Science* n/a, n/a (July 2021), e1610. https://doi.org/10.1002/wcs.1610 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcs.1610.

[64] Josefa Oberem, Jan-Gerrit Richter, Dorothea Setzer, Julia Seibold, Iring Koch, and Janina Fels. 2020. Experiments on localization accuracy with non-individual and individual HRTFs comparing static and dynamic reproduction methods. (March 2020), 11.

[65] M. Torben Pastore and William A. Yost. 2022. Spatial Release from Masking for Tones and Noises in a Soundfield under Conditions Where Targets and Maskers Are Stationary or Moving. *Audiology Research* 12, 2 (April 2022), 99–112. https://doi.org/10.3390/audiolres12020013 Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[66] Siyun Peng and Adam R. Roth. 2021. Social Isolation and Loneliness Before and During the COVID-19 Pandemic: A Longitudinal Study of U.S. Adults Older Than 50. *J Gerontol B Psychol Sci Soc Sci* XX (2021), 1–6. https://doi.org/10.1093/geronb/gbab068

[67] Martin Pienkowski. 2021. Loud Music and Leisure Noise Is a Common Cause of Chronic Hearing Loss, Tinnitus and Hyperacusis. *International Journal of Environmental Research and Public Health* 18, 8 (April 2021), 4236. https://doi.org/10.3390/ijerph18084236

[68] Constantin Popp and Damian T. Murphy. 2024. Speech Intelligibility versus Congruency: User Preferences of the Acoustics of Virtual Reality Game Spaces. *Virtual Worlds* 3, 1 (March 2024), 40–61. https://doi.org/10.3390/virtualworlds3010003 Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[69] Elena Prokofieva, C. Luciani, and I. McGregor. 2014. Design and Development of Auralization Room at Edinburgh Napier University. Audio Engineering Society. https://www.aes.org/e-lib/browse.cfm?elib=17151

[70] R Core Team. 2024. R: The R Project for Statistical Computing. https://www.r-project.org/

[71] Nikunj Raghuvanshi. 2021. Dynamic Portal Occlusion for Precomputed Interactive Sound Propagation. http://arxiv.org/abs/2107.11548 arXiv:2107.11548 [cs, eess].

[72] Nikunj Raghuvanshi and John Snyder. 2018. Parametric directional coding for precomputed sound propagation. *ACM Transactions on Graphics* 37, 4 (Aug. 2018), 1–14. https://doi.org/10.1145/3197517.3201339

[73] Giuseppe Riva, Luca Bernardelli, Gianluca Castelnuovo, Daniele Di Lernia, Cosimo Tuena, Alex Clementi, Elisa Pedroli, Clelia Malighetti, Francesca Sforza, Brenda K. Wiederhold, and Silvia Serino. 2021. A Virtual Reality-Based Self-Help Intervention for Dealing with the Psychological Distress Associated with the COVID-19 Lockdown: An Effectiveness Study with a Two-Week Follow-Up. *International Journal of Environmental Research and Public Health* 18, 15 (Aug. 2021), 8188. https://doi.org/10.3390/ijerph18158188

[74] Sarah Roßkopf, Leon O. H. Kroczek, Felix Stärz, Matthias Blau, Steven van der Par, and Andreas Mühlberger. 2023. The Effect of Audio-Visual Room Divergence on the Localization of Real Sound Sources in Virtual Reality. (2023).

[75] Michael K. Russell and Stephanie Brown. 2019. Using Sound to Create and Detect Occlusion of an Unseen Sound Source. *Auditory Perception & Cognition* 2, 4 (Oct. 2019), 207–229. https://doi.org/10.1080/25742442.2020.1773731

[76] Mostafa Sabbagh and Ahmed Elkhateeb. 2021. Effect of body posture on sound absorption by human subjects. *Applied Acoustics* 183 (Dec. 2021), 108317. https://doi.org/10.1016/j.apacoust.2021.108317

[77] B. Scharf. 1971. Fundamentals of Auditory Masking. *Audiology* 10, 1 (Jan. 1971), 30–40. https://doi.org/10.3109/00206097109072538 Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.3109/00206097109072538.

[78] Carl Schissler, Gregor Mückl, and Paul Calamia. 2021. Fast diffraction pathfinding for dynamic sound propagation. *ACM Transactions on Graphics* 40, 4 (Aug. 2021), 1–13. https://doi.org/10.1145/3450626.3459751

[79] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. Using Presence Questionnaires in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–12. https://doi.org/10.1145/3290605.3300590

[80] Elena Shabalina. 2014. *The Propagation of Low Frequency Sound Through an Audience*. Logos Verlag Berlin GmbH. Google-Books-ID: E64cAwAAQBAJ.

[81] Charles Spence, Jae Lee, and Nathan Van der Stoep. 2020. Responding to sounds from unseen locations: crossmodal attentional orienting in response to sounds presented from the rear. *European Journal of Neuroscience* 51, 5 (March 2020), 1137–1150. https://doi.org/10.1111/ejn.13733

[82] Ekaterina R. Stepanova, John Desnoyers-Stewart, Kristina Höök, and Bernhard E. Riecke. 2022. Strategies for Fostering a Genuine Feeling of Connection in Technologically Mediated Systems. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–26. https://doi.org/10.1145/3491102.3517580

[83] Daniel J Strauss, Farah I Corona-Strauss, Andreas Schroeer, Philipp Flotho, Ronny Hannemann, and Steven A Hackley. 2020. Vestigial auriculomotor activity indicates the direction of auditory attention in humans. *eLife* 9 (July 2020), e54536. https://doi.org/10.7554/eLife.54536 Publisher: eLife Sciences Publications, Ltd.

[84] Mattia Surricchio, Andrea Damiani, and Marco D. Santambrogio. 2022. Obstruction simulation in real-time 3D audio on edge systems. In *2022 IEEE 20th International Conference on Embedded and Ubiquitous Computing (EUC)*. 17–22. https://doi.org/10.1109/EUC57774.2022.00012

[85] Yôiti Suzuki, Hisashi Takeshima, and Kenji Kurakata. 2024. Revision of ISO 226 "Normal Equal-Loudness-Level Contours" from 2003 to 2023 edition: The background and results. *Acoustical Science and Technology* 45, 1 (Jan. 2024), 1–8. https://doi.org/10.1250/ast.e23.66

[86] Unity Technologies. 2024. Unity - Manual: Audio Source. https://docs.unity3d.com/Manual/class-AudioSource.html

[87] Ching-I Teng. 2010. Customization, immersion satisfaction, and online gamer loyalty. *Computers in Human Behavior* (2010), 8.

[88] TopPNG. 2019. Image Result For Body Outline Body Outline Outlines - Figure Drawi PNG Transparent With Clear Background ID 168232 png - Free PNG Images. https://toppng.com/free-image/image-result-for-body-outline-body-outline-outlines-figure-drawi-PNG-free-PNG-Images_168232

[89] Luca Turchet. 2015. Footstep sounds synthesis: Design, implementation, and evaluation of foot–floor interactions, surface materials, shoe types, and walkers' features. *Applied Acoustics* (July 2015). https://doi.org/10.1016/j.apacoust.2015.05.013

[90] Unity Technologies. 2023. Unity Real-Time Development Platform | 3D, 2D, VR & AR Engine. https://unity.com

[91] Valve. 2022. Steam Audio. https://valvesoftware.github.io/steam-audio/

[92] Jan B.F. Van Erp, Camille Sallaberry, Christiaan Brekelmans, Douwe Dresscher, Frank Ter Haar, Gwenn Englebienne, Jeanine Van Bruggen, Joachim De Greeff, Leonor Fermoselle Silva Pereira, Alexander Toet, Nirul Hoeba, Robin Lieftink, Sara Falcone, and Tycho Brug. 2022. What Comes After Telepresence? Embodiment, Social Presence and Transporting One's Functional and Social Self. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2067–2072. https://doi.org/10.1109/SMC53654.2022.9945544 ISSN: 2577-1655.

[93] Mark VanDam. 2014. Acoustic characteristics of the clothes used for a wearable recording device. *The Journal of the Acoustical Society of America* 136, 4 (Oct. 2014), EL263–EL267. https://doi.org/10.1121/1.4895015 Publisher: Acoustical Society of America.

[94] Yonghao Wang. 2017. Low Latency Audio Processing. (2017).

[95] Koichiro Wasano, Kimitaka Kaga, and Kaoru Ogawa. 2021. Patterns of hearing changes in women and men from denarians to nonagenarians. *The Lancet Regional Health - Western Pacific* 9 (April 2021), 100131. https://doi.org/10.1016/j.lanwpc.2021.100131

[96] World Health Organization. 2022. *WHO global standard for safe listening venues and events*. World Health Organization, Geneva. https://apps.who.int/iris/handle/10665/352277 Section: viii, 107 p..

[97] Fang Xu, Tianyu Zhou, Tri Nguyen, Haohui Bao, Christine Lin, and Jing Du. 2024. Synergizing Augmented Reality and Llms for Advanced Cognitive Support in Emergency Audio Communications. https://doi.org/10.2139/ssrn.4805664

[98] You Zhang, Yuxiang Wang, and Zhiyao Duan. 2023. HRTF Field: Unifying Measured HRTF Magnitude Representation with Neural Fields. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Rhodes Island, Greece, 1–5. https://doi.org/10.1109/ICASSP49357.2023.10095801

[99] Jiahong Zhao, Xiguang Zheng, Christian Ritz, and Daeyoung Jang. 2022. Interpolating the Directional Room Impulse Response for Dynamic Spatial Audio Reproduction. *Applied Sciences* 12, 4 (Jan. 2022), 2061. https://doi.org/10.3390/app12042061 Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.