

This is a repository copy of *Measuring the Behavioral Response to Spatial Audio within a Multi-Modal Virtual Reality Environment in Children with Autism Spectrum Disorder*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/149430/>

Version: Published Version

Article:

Johnston, Daniel Ian, Egermann, Hauke Wolfgang orcid.org/0000-0001-7014-7989 and Kearney, Gavin Cyril orcid.org/0000-0002-0692-236X (2019) Measuring the Behavioral Response to Spatial Audio within a Multi-Modal Virtual Reality Environment in Children with Autism Spectrum Disorder. *Applied Sciences*. 3152. ISSN 2076-3417

<https://doi.org/10.3390/app9153152>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Article

Measuring the Behavioral Response to Spatial Audio within a Multi-Modal Virtual Reality Environment in Children with Autism Spectrum Disorder

Daniel Johnston ^{1,†} , Hauke Egermann ² and Gavin Kearney ^{1,*} 

¹ Communications & Signal Processing Research Group, Department of Electronic Engineering, University of York, York YO10 5DD, UK

² York Music Psychology Group, Music Science and Technology Research Cluster, Department of Music, University of York, York YO10 5DD, UK.

* Correspondence: dij502@york.ac.uk; Tel.: +44-(0)1904-324-229

† Current address: Communications & Signal Processing Research Group, Department of Electronic Engineering, University of York, Heslington, York YO10 5DD, UK

Received: 30 June 2019; Accepted: 31 July 2019; Published: 2 August 2019



Abstract: Virtual Reality (VR) has been an active area of research in the development of interactive interventions for individuals with autism spectrum disorder (ASD) for over two decades. These immersive environments create a safe platform in which therapy can address the core symptoms associated with this condition. Recent advancements in spatial audio rendering techniques for VR now allow for the creation of realistic audio environments that accurately match their visual counterparts. However, reported auditory processing impairments associated with autism may affect how an individual interacts with their virtual therapy application. This study aims to investigate if these difficulties in processing audio information would directly impact how individuals with autism interact with a presented virtual spatial audio environment. Two experiments were conducted with participants diagnosed with ASD ($n = 29$) that compared: (1) behavioral reaction between spatialized and non-spatialized audio; and (2) the effect of background noise on participant interaction. Participants listening to binaural-based spatial audio showed higher spatial attention towards target auditory events. In addition, the amount of competing background audio was reported to influence spatial attention and interaction. These findings suggest that despite associated sensory processing difficulties, those with ASD can correctly decode the auditory cues simulated in current spatial audio rendering techniques.

Keywords: autism spectrum disorders; virtual reality; auditory processing; assistive technology; tools for therapy; multi-sensory; spatial audio; ambisonics

1. Introduction

Autism spectrum disorder (ASD) is a lifelong neurodevelopmental disorder that affects approximately 1% of the worldwide population [1]. It is characterized through core symptoms that include impaired development in social interaction, communication, repetitive behaviors, restrictive interests, and sensory processing difficulties [2]. Autism is heterogeneous, and despite these symptoms being commonly identified with the condition, the scope of which these difficulties affect everyday life are often unique to the individual [3]. With no cure for ASD, there is an increasing need for intervention programs that target appropriate behavioral and educational complications with the aim to improve the quality of life for the individual and their family [4].

1.1. Virtual Reality and ASD

Virtual Reality (VR) has been an active area of research in the development of interventions for this population for over two decades. Empirical studies have recognized VR as an important therapy tool for addressing the core symptoms associated with ASD such as difficulties in social interaction [5] and communication [6]. In addition, virtual reality presents dynamic and controllable environments in which educational and training practices such as disaster awareness [7], driving [8] and independent functioning [9] training can take place.

Early work conducted by Strickland [10] in 1996 suggested key aspects of this technology which can be applied to effective and engaging therapeutic solutions for both children and adults living with autism, and despite extensive advancements in how VR is delivered these can still be applicative to contemporary research practices. Firstly, many individuals with autism respond to and enjoy using virtual reality technology [11]. This translates to users actively engaging with interventions, with the enjoyment providing motivation to progress with the therapy program [12,13]. Secondly, virtual environments provide an experience with consistent and predictable interactions and responses [14]. This can be in the form of limiting social and communication pressures that can often cause anxiety in real world situations [15]. Furthermore, those with autism can frequently experience complications in processing sensory information, with particular prevalence in the auditory and visual domains. These irregularities in sensory receptiveness can often be a catalyst for a deterioration in behavior observed as aggressive or autonomic fear responses [16]. Virtual reality can control all audio and visual stimuli that are encountered during an intervention session, simplifying complex sensory arrays and thus reducing the elicitation of anxiety [10,17]. The control of VE's and interactions within them can also construct important opportunities for individualized treatment [10]. Re-adjusting the sensory and learning parameters of the intervention to compensate for the complex needs of the individual [11]. Finally, virtual reality technology can replicate realistic representations of real-world situations in which participants can learn and practice new behaviors and real-life skills. The increased feeling of presence and ecological validity as a result of this realism can have a positive impact on the generalization of newly acquired skills from the virtual intervention to their real-world applications [11]. However, the visual experience of the user is often the focus of research when investigating virtual reality interventions for ASD. Although what the user sees can increase generalization via realistic visual similarities with the real world, the sense of realism is reliant on the extent and fidelity of the VR technology delivering the multi-sensory stimuli to the user, including audio.

1.2. The Importance of Spatial Audio

Larson et al. [18] suggest that the spatial properties of virtual auditory environments have been of significant importance since the introduction of stereo sound in the 1930s. Stating that spatial sound is used to simulate an auditory reality that gives the user the impression they are surrounded by a 3D virtual environment. In terms of precision, spatial acuity is somewhat inferior to vision and somatic sensory system when perceiving the environment [19]. However, it is not unimportant. Spatial hearing is a critical way in which the user perceives space, providing information about events and objects that are far beyond the field of view [20].

Two studies conducted by Hendrix and Barfield [21] investigated how three-dimensional sound influences the user experience within a stereoscopic virtual environment (VE). The first compared a silent VE to environments with spatial sound. The second was a comparison between two environments with spatialized and non-spatialized sound sources. The results indicated that spatialized audio significantly increased the feeling of presence, with interactive sounds to be originating from their visual cues perceived as more realistic.

To perceive the real world, the human brain must decode a constant stream of multi-modal information from the various sensory channels [22]. An example of which is most cases of face to face communication. The brain must interpret visual information such as moving lips and facial

expressions, while also listening to localized speech sound from the mouth [23]. This cross modality synchronization can also be echoed in VR by using 3D audio, by simulating the shared spatial qualities of both audio and visual stimuli that occupy the same time and space.

1.3. Spatial Audio Rendering

For healthy listeners, the human auditory system is adept at decoding how a sound wave diffracts and interacts with the head, torso, and pinnae, resulting in both temporal and amplitude differences changes as well as spectral cues [24].

In terms of localization interaural time and level differences (ITD and ILD) are the two major cues attributed to sound localization along the horizontal plane (see Figure 1). For low frequency sound sources (<1.5 kHz), the detection of ITD is the primary mechanism [25]. This frequency threshold is a result of the distance a sound wave must travel to reach the furthest ear, causing the phase angle to differ upon arrival. The corresponding angle being both the function of the wavelength and the source positioning. With frequencies below 1.5 kHz the wavelength is greater than then maximum difference in time, this phase difference will therefore provide accurate location cues [26]. For higher frequency targets, ILD detection takes over from temporal differences as the primary horizontal localization mechanism [25]. Shorter high frequency sound waves are more susceptible to the acoustic shadowing caused by the human head. It is this increased shadowing effect that produces a greater difference in sound levels between the two ears [27].

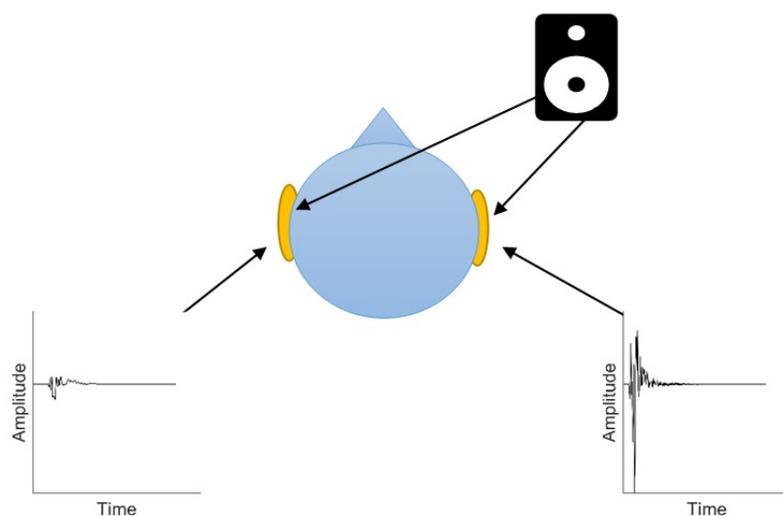


Figure 1. Binaural cues for horizontal localization. Plots display two time-domain representations of Head Related Transfer Functions (HRTF) recordings (45° azimuth & 0° elevation) extracted from the SADIE Database [28].

Localization within the vertical plane is calculated from the spectral shaping of acoustic energy by torso and the unique and complex geometry of the pinna. Frequencies are distributed and consequently reflected causing minute but significant delays. It is these delays that will act as a comb filter effect on sound entering the inner part of the ear, representing a three-dimensional function of the elevation of the sound source [29,30].

The realistic reproduction of three-dimensional auditory environments through headphones, known as binaural-based spatial audio, is dependent upon the correct simulation of the interaural differences and spectral shaping cues caused by the torso, head and pinnae called Head Related Transfer Functions (HRTF) [31]. HRTF's are the functions of both frequency and position for both

azimuth and elevation, containing direction-dependent information on how a free field sound source reaches the ear [32,33]. Binaural rendering can then be achieved through the convolution of a monoaural anechoic sound source and the HRTF of both the desired direction and the corresponding ear [32,34].

Today most VR systems rely on head-tracked headphone-based systems (see Figure 2). Head tracking is definitive in creating a plausible and realistic virtual audio scene. Current virtual reality head devices can provide stable multi-modal interaction through accelerometers which can detect force, 360 orientation and measure the device's position within a 3-D space. Once the rendering system has received the X, Y, and Z directional information, it performs interpolation between the closest HTRF's resulting perceived localization of a dynamic virtual sound source [35,36]. Furthermore, by reproducing the audio over headphones it is possible to isolate the left and right audio channels and successfully reproduce the binaural localization cues. However, there are several auditory processing difficulties associated with autism spectrum disorders that may affect how a user may respond to auditory cues within a VR intervention environment.

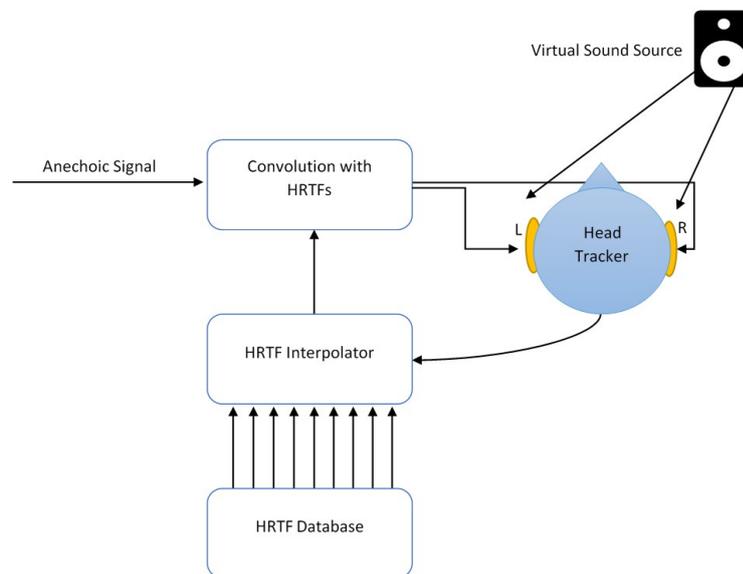


Figure 2. Block diagram illustrating audio signal chain for headphone-based spatial audio.

1.4. Atypical Auditory Processing in Autism

Commonly, children with autism experience complications in processing sensory information with up to 65% of individuals with ASD displaying complications in auditory domain [37–39]. There has been extensive research pertaining to auditory processing in ASD. Observations are diverse, ranging from enhanced abilities in musical pitch discrimination and categorization [40], to hypersensitivity of particular sound stimuli which evoke extreme negative behavioral reactions [41]. Of particular note, there have been several psychoacoustic studies which have investigated auditory scene analysis.

It has been reported that in the presence of competing background noise, those with autism often have difficulties in separating relevant auditory information (i.e., speech and target noises) into discrete auditory objects connected to the different sound sources within the environment [42]. These complications in processing sound information in noise have links to a deficit in auditory spatial attention, which is often exhibited as a failure to orient to speech and other social stimuli [43].

Soskey et al. [44] found that compared to typically developed (TD) controls, children with ASD showed a significant impairment in attending to sounds, in particular speech sounds.

During typical development, an infant of 6-weeks can display a significant sensitivity to social stimuli, with particular focus to the features of the human face and speech [45,46]. Autistic children on the other hand have been reported to display atypical cortical responses to synthetic speech-like stimuli, culminating in reduced involuntary attention to social auditory events [47]. These early limited reactions to social stimuli represent the core social impairments in the earliest and most basic form and have negative impacts on future social and communication development [48].

In addition, Lin et al. [42] observed a reduced response to speech in noise in individuals with autism, alongside reduced discrimination for ITD and ILD. These binaural cues are processed by an area of the auditory pathway known as the medial superior olive [49]. Interestingly, postmortem examinations of the brainstem olivary complex in those with ASDs observed neurological malformations resulting in a reduced size of the medial superior olive [50,51]. Despite this, people with autism spectrum disorders have been shown to perform equally well as healthy controls in noise localization experiments across the horizontal plane [52]. Therefore, an impaired sensitivity to ITD and ILD cues in autism have been linked to deficits in selective hearing, which can affect orientation to target sounds within noise [42].

2. Design and Hypothesis

This study aims to examine auditory spatial attention and sound localization ability of children and adolescents with ASD within a multi-modal virtual reality game environment. The therapeutic outcomes of VR interventions often rely on the immersion of the participant through place illusion [53]. This is particularly important when designing virtual interventions for individuals with autism [11], a population who have difficulties in generalizing newly skills to real-world applications. By conducting investigations within a VR environment similar to which participants may experience within an intervention setting, test parameters may be precisely controlled whilst the user interactions and behavioral models can be robustly recorded. Furthermore, the inclusion of dynamic visuals and game mechanics maintains participant immersion, therefore examining a pattern of behavior closer to that within a conventional virtual reality experience removed from experimental conditions [54]. Previous work by Wallace et al. [55] measured spatial presence within the 'blue room', an advanced collaborative virtual environment (CVE). The room makes use of visual images projected onto four screens located on three walls and the ceiling, with audio being delivered over a surround sound loudspeaker system. The paper reported that children with ASD showed equal amounts of attention to the virtual content than healthy controls and engaged with the virtual environment requirements. However, loudspeaker-based systems are often less effective at simulating the necessary ITD and ILD cues needed for three-dimensional spatial audio reproduction. Speaker crosstalk causes significant signal interference between each speaker and the opposite ear of the listener, resulting in impaired localization performance [56]. In contrast, head-track headphone-based systems are more successful at recreating dynamic soundfields which respond to the listeners position within the virtual world [36].

This investigation followed an in-between subject design. Participants were randomly allocated to either a 3D spatial audio, or control group. The 3D audio group would be exposed to spatial audio reproduced binaurally over headphones. The control group would listen to a monaural representation of the virtual auditory scene presented to both ears over headphones. Each participant group completed a two-phase experiment which first tested spatial audio attention, followed by a sound localization task. Both within a multi-modal virtual reality environment (see Figure 3).

It is hypothesized that subjects exposed to binaural-based spatial audio would show more spatial attention to auditory stimuli compared with the mono sound control group, indexed through more precise head orientation towards the presented audio targets. Furthermore, spatial audio stimuli would influence participant positional movement with the virtual environment. Additionally, based on previous studies of auditory scene analysis deficits in ASD, it is predicted that localization times will

be higher for audio cues with the most amount of competing environmental background noise. Finally, by examining spatial attention towards complex speech and non-speech sounds, it is possible to investigate if differences in spatial auditory attention are the consequence of a reduced orientation towards social stimuli.

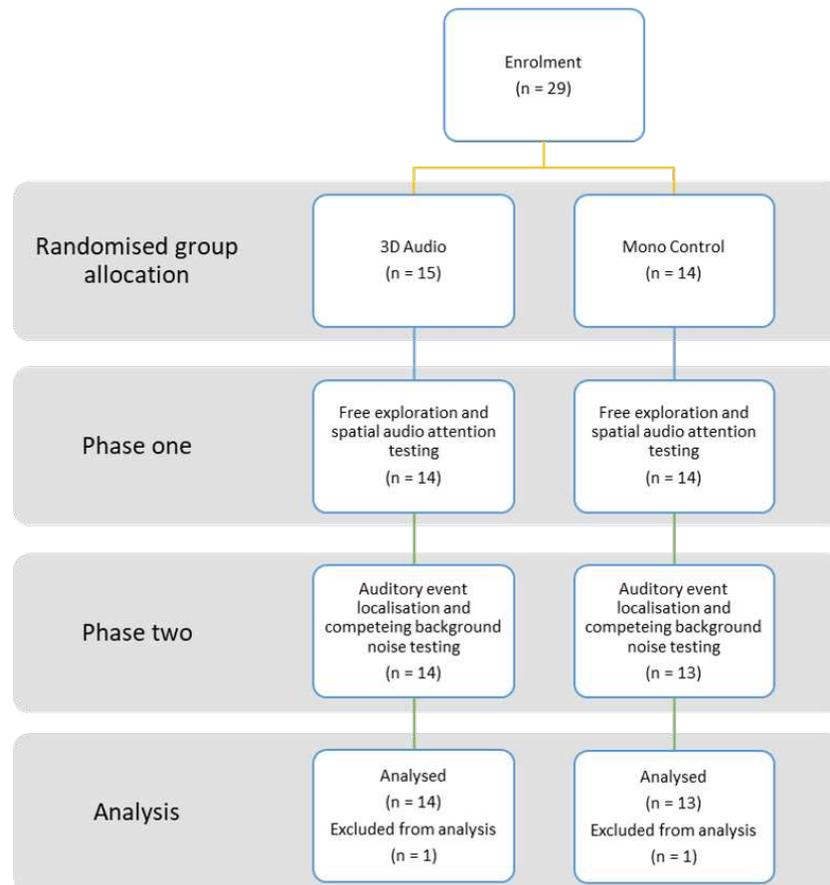


Figure 3. Flow chart displaying experimental procedure.

3. Methods

3.1. Participants

A study group consisted of 29 children and adolescents (27 male and two female, mean age = 14, SD = 2.42, range of 9–19 years). Participants were recruited from two special education schools and one charity. All participants had a formal diagnosis of autism spectrum disorder obtained from their local national health trust, displaying social, cognitive, and motor functioning associated with moderate to high functioning autism. Exclusion criteria were self-reported hearing problems; physical disabilities that would limit movement around experiment space; and an inability to finish the task. Two participants were excluded due to not being able to complete the full experiment. This study and methods were approved by the University of York board of ethical approval, and an information package was provided to the participants parent(s) or legal guardian(s). Participants were admitted into the study after informed consent and assent was obtained from their parent(s) or legal guardian(s).

3.2. Equipment

All audio and visual stimuli were rendered using the Oculus Rift CV1 head mounted display (HMD) with built-in on-ear headphones. Head tracking was also achieved using the HMD, with motion tracking of participant position calculated by the Oculus Rift sensors [57].

3.3. Stimuli

Visual Environment: The virtual environment used for this experiment represents an enchanted forest setting at night (see Figure 4). It has been noted that when designing visual aspects of virtual reality for ASD applications, the presence of visual clutter significantly affects user performance at completing tasks [58]. Therefore, the graphical environment was designed to be engaging without presenting visual elements that may have driven user attention away from the auditory stimuli. Furthermore, individuals with autism that also exhibit abnormalities in processing visual information can often attempt to avoid visual input such as bright lights [59]. The use of a darker environment reduces the possibility of anxiety related behaviors that are a result over sensory over-stimulation [60].



Figure 4. Visual Environment—360 degree capture.

Auditory Stimuli: Figure 5a,b are overhead views of the virtual environment displaying the auditory stimuli placement and movement area, with a summarization of each audio object in Tables 1 and 2. All stimuli were recorded or edited using Reaper digital audio workstation and Wwise [61] game audio engine. Experimental groups were based upon audio rendering techniques. The 3D audio group would listen to binaural-based spatial audio rendered via the Google Resonance spatial audio SDK [62]. Audio was presented at a fixed playback level of 68 dbA. This was matched to a -20 dBFS RMS value of a pink noise test signal playing at a distance of 1m from the listener.

Phase One: Eight spatial auditory events (see Table 1) were placed in the VE. All audio events were stationary with the exception of event 7 which moved around the entire environment over the course of 30 s. Previous research conducted by the authors has noted that target stimuli among competing audio could prompt negative effects on auditory spatial attention within VR environments for individuals with ASD [63]. This is further supported by research that reports within competing background noise, those with autism often display difficulties in separating relevant auditory information (i.e., speech and target noises) into discrete auditory objects connected to the different sound sources within the environment [42]. To minimize the effects of any complications in auditory scene analysis associated with ASD, environmental audio was reduced to -43 LUFS while each Phase One event was rendered.

Phase Two: Eight spatial audio events representing virtual characters were placed throughout the VE. In terms of type, audio stimuli were divided into simple and social types. Simple stimuli consisted of a bell sound effect, with the social stimuli representing a speech-like sound. All events were stationary except for movement towards the player once the character was found.

A spatialized composite background audio track was provided which was derived from several discrete auditory objects placed throughout the VE, matching the visual scene. This included sounds representing wind, water, frogs, birds, crickets and wind chimes.

The level of the background sounds was either ≥ -43 LUFS or ≥ -53 LUFS. These levels were taken from a past study that evaluated the effects of background noise on spatial attention in VE’s with participants with ASD [63]. Using an audio object-based hierarchical model based on a simplified technique designed by Ward et al. [64] to increase the accessibility of complex audio content for those with hearing impairments. Background audio levels would be reduced by removing the bird, cricket, and wind chime ambient sound objects; the aim of this process is to maintain participant immersion.

To further investigate auditory detection in noise, target stimuli were played at differing signal-to-noise ratio (SNR). For auditory events with +10 SNR, the stimulus level would be 10 LUFS higher than the background noise level. For auditory events with 0 SNR, the stimulus would be set at an equal level to that of the background noise.

Table 1. Phase One Auditory Stimuli descriptions.

Auditory Event	Stimulus	Visual Cue
1	MIDI-based music–Synth Pad	Present
2	Bell sound effects	None
3	Bubble sound effects	Present
4	Explosion & Speech-like stimuli	None
5	MIDI-based music–Synth Pad	Present
6	Crow sound effects	None
7	Explosion & Speech-like stimuli	Present
8	Explosion & Speech-like stimuli	None

Table 2. Phase Two Auditory Stimuli–SNR represents either the stimulus played at +10 LUFS above the background audio level, or set to a level equal to that of the background audio.

Auditory Event	Stimulus	Background Sound Level (LUFS)	SNR
1	Simple	≥ -53	+10
2	Social	≥ -43	+10
3	Social	≥ -43	0
4	Simple	≥ -43	0
5	Simple	≥ -53	0
6	Simple	≥ -43	+10
7	Social	≥ -53	0
8	Social	≥ -53	+10

3.4. Procedure

Participants were permitted to move freely around a pre-defined tracked experimental space of 1.6 m × 1.6 m while wearing the HMD. The combination of dynamic head and positional tracking allowed movement around the virtual environment with 6 degrees-of-freedom.

Throughout the experiment a support worker would be present to communicate with the participant and provide assistance if they became distressed. However, they were not permitted to deliver instructions. Prior to the start of the session, the experimenter would explain the use of the virtual reality system, the structure of the experiment and place the HMD on the participant in order for them to become familiarized with the device.



Figure 5. Top-down view of auditory stimuli positions within the virtual environment. (a) Phase one; (b) Phase two.

3.4.1. Phase One—Free Exploration and Spatial Audio Attention Testing

Prior to the main experimental task in Phase One, subjects were introduced to the virtual reality equipment and environment for a period of two minutes with minimal environmental audio and visuals. The experimenter explains to the participant that they are free to move around and explore the VE. This brief initial time was used to provision for any emotional excitement experienced by using virtual reality. The participant also receives support from the experimenter to become familiar with the virtual environment. Following this, participants were given a further five minutes free exploration during which they would be exposed to eight auditory events played once at pre-determined times, four of which had visual accompaniments and four had no visual accompaniments (see Table 1). This period would be used to record participant head rotation and horizontal tracked positional data. Once again the experimenter tells the participant that they are free to move around and explore the VE. After this verbal communication with the participant was minimized to control any influence that may lead to any guided exploration during the experiment.

An accuracy metric (a) was used to evaluate participant performance during the spatial audio attention task. This was quantified by calculating the relative difference between the participant head rotation on the azimuth axis (y) and the target position of the reproduced stimulus (n) at 100 ms intervals.

$$a = 1 - \left| \frac{(n - y)}{180} \right| \tag{1}$$

Further data recorded the target azimuth location in degrees of each spatial auditory event with the respective start and end times, and virtual environment positions. Higher values of a during auditory stimulus playback time would represent greater spatial attention towards presented auditory targets.

To better understand if spatial audio influences subject behavior tracked horizontal plane positional data within the VE represented by x and z was also collected at 100 ms intervals. A distance score was then calculated by measuring the distance between the participant and each of the presented auditory stimuli during playback time. Smaller distance values during auditory stimulus playback time would indicate movement towards the presented stimuli, indicating an influence of participant movement.

3.4.2. Phase Two—Spatial Audio Localization and Background Noise Testing

Participants took part in a localization task which required them to locate a total of eight hidden characters within the virtual world based on localizing sounds that they emitted and limited visual representations. Each would play either social or simple stimuli (see Table 2).

The experimenter first explains the task clarifying that the participant must use their ears to find each character. To aid the participant in becoming familiarized with the target stimuli, they are presented with an example character in their field of view which plays both social and simple auditory stimuli. Following this, the experimenter would activate each hidden character at a time when the participant was comfortable to do so and repeat the explanation of the task.

Once again support worker communication was restricted so as to avoid an effect on the outcome of the investigation. The length of this phase would be dependent on the participants ability to progress. When each virtual character was found the participant would be given positive reinforcement in the form of verbal praise.

Localization times of auditory stimuli within the virtual environment during Phase Two were recorded for each participant for all eight virtual characters. Further to group comparisons, exploratory analyses of participant reaction time also examined the relationship between the amount of background ambient noises and spatial attention. To do so, localization times for each character were compared based on stimulus type, background audio level and SNR.

4. Results

4.1. Phase One: Spatial Audio Attention Testing

A two-way mixed ANOVA was conducted to investigate the impact of 3D audio and auditory stimuli on the accuracy metric during Phase One. A comparison between test groups yielded a significant difference ($F(1,26) = 29.43, p \leq 0.001, \eta_p^2 = 1.642$) across all auditory events, with the 3D audio test group ($M = 0.661$) scoring higher accuracy scores than the mono audio controls ($M = 0.487$). Figure 6 shows a comparison of group mean accuracy scores with 95% confidence intervals for each event.

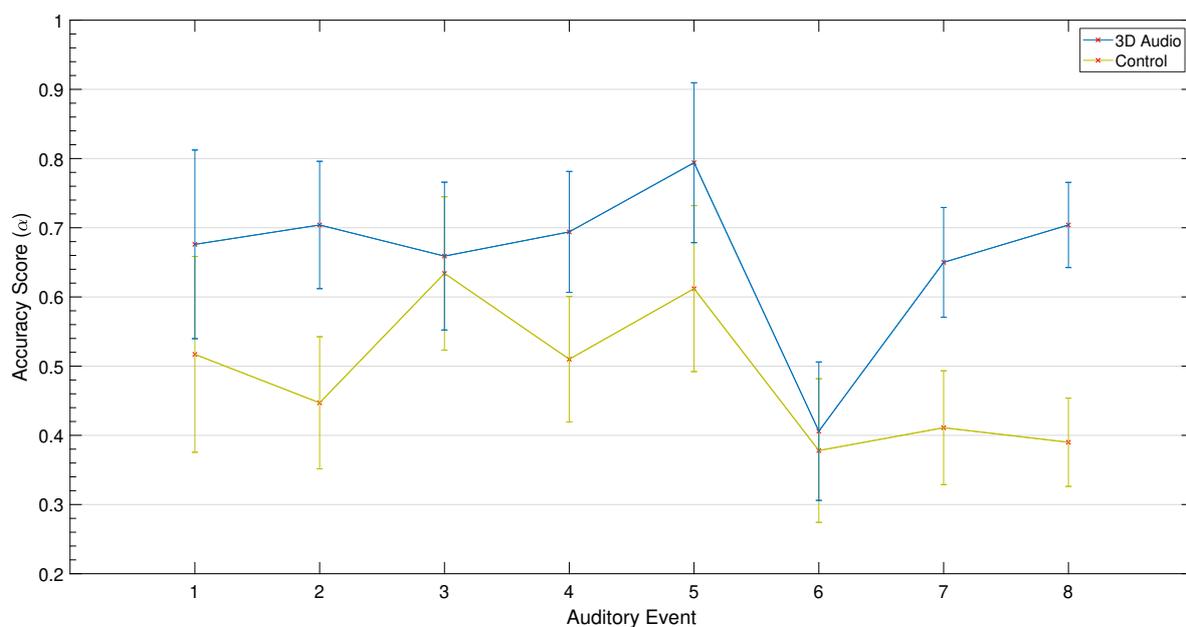


Figure 6. Mean accuracy (α) scores for each audio condition across all auditory events in Phase One. The whiskers denote 95% confidence intervals.

In addition, there is a significant effect of stimulus type on spatial attention of both groups ($F(7, 175) = 7.657, p \leq 0.001, \eta_p^2 = 0.227$), with a significant interaction between groups and stimulus type ($F(7, 175) = 2.426, p = 0.021, \eta_p^2 = 0.072$). Finally, the use of visual accompaniment had no significant effect on the accuracy of spatial attention in the spatial audio group yielded by visual accompaniment ($M = 0.627, SD = 0.187$) and no visual accompaniment ($M = 0.695, SD = 0.184$) paired-samples t-test; $t(55) = -1.903, p = 0.062$. However, there was a significant difference between visuals ($M = 0.543, SD = 0.235$) and no visuals ($M = 0.431, SD = 0.176$) within the mono control group; $t(51) = -2.827, p = 0.007$.

To better understand behavioral response to spatial audio within the virtual environment, tracked positional data across the two groups was also analyzed. The tracked positional coordinates of all participants on the horizontal axis collected throughout the entirety of Phase One and are visualized as 2-dimensional density plots shown in Figure 7. In addition, the figure also displays the directions and distance of all auditory events inside and outside of the tracked area.

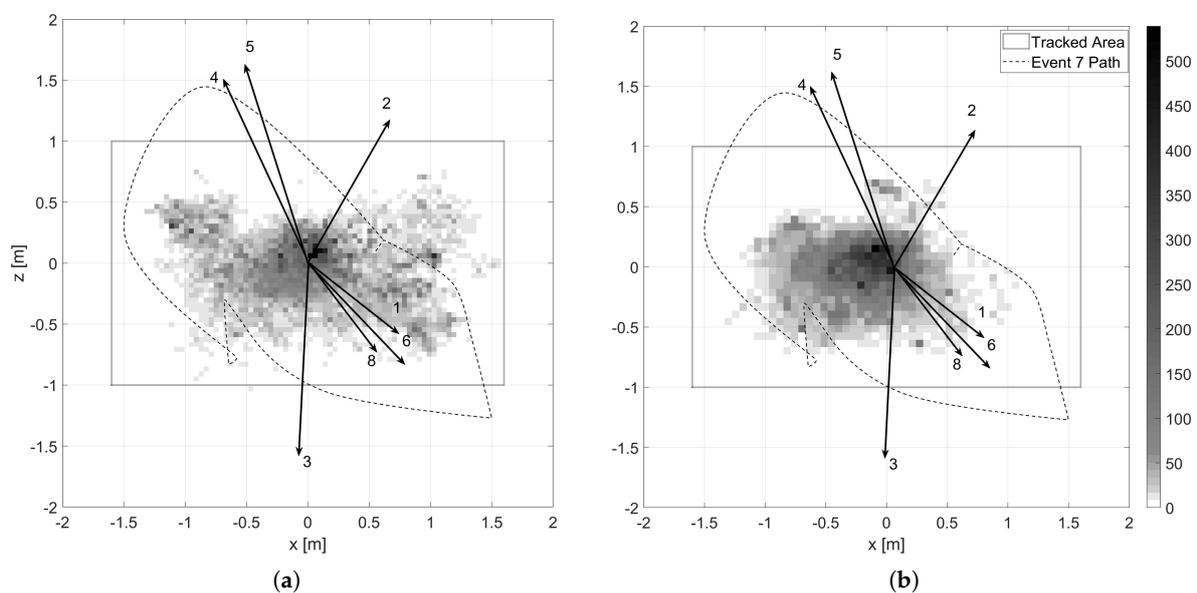


Figure 7. Two-dimensional density plot of positional tracking data throughout Phase One for all participants. Arrows point toward the target audio objects outside of the plot area. Distance from origin to audio objects—1 (6.8 m), 2 (4.57 m), 3 (5.98 m), 4 (12.43 m), 5 (5.96 m), 6 (3.86 m), 8 (5.74 m). (a) Spatial audio condition; (b) Mono condition.

The figure shows differences in the extent of exploration depending on the auditory condition. Despite both groups showing a condensation of exploration around the center of the tracked space, the 3D audio group show a distribution of exploration in the direction of the majority of auditory targets. Further analysis of the distance between each participant and the auditory target using a two-way mixed ANOVA also yielded a significant difference between the auditory conditions ($F(7, 175) = 23.111, p \leq 0.001, \eta_p^2 = 0.308$). With participants within the 3D audio group moving closer towards the virtual sound sources during their respective playback times.

4.2. Phase Two: Spatial Audio Localization and Background Noise Testing

A two-way mixed ANOVA showed that these differences in localization times were also significantly different ($F(7, 175) = 7.565, p \leq 0.001, \eta_p^2 = 0.232$). Figure 8 shows the mean localization times for both experiment groups across all auditory events in Phase Two, which shows shorter localization times in the 3D audio condition ($M = 8.4$ s) than in the mono control condition ($M = 17.8$ s).

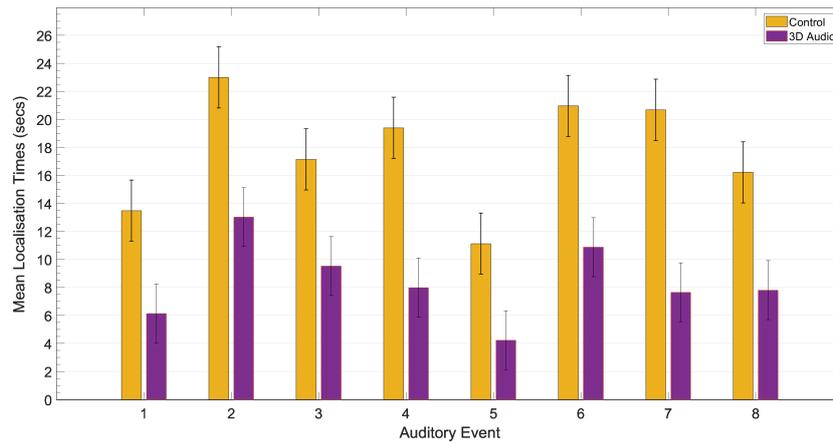


Figure 8. Mean localization times for each audio condition across all auditory events in Phase two. The whiskers denote 95% confidence intervals.

A hierarchical linear model was used to analyze the effects background audio, SNR, and type of stimulus have on localization times within both the 3D audio and mono control conditions, the results of which can be seen in Figure 9. Inferential statistics reported a significant overall effect of background audio on the time taken to locate each auditory event for both groups: 3D audio ($F(1, 54.713) = 15.243, p < 0.001$) and mono audio ($F(1, 84) = 10.476, p = 0.002$). Auditory stimuli within the higher background audio level of -53 LUFS (3D audio $M = 10.4$ s and mono audio $M = 20.9$ s) displayed higher localization times than those with -43 LUFS (3D audio $M = 6.5$ s, mono audio $M = 15.4$ s). Further to this, the effect of SNR between background sounds and auditory stimulus was only significant for the 3D audio group ($F(1, 71.029) = 5.206, p = 0.026$), with the time taken to correctly locate auditory stimuli $+10$ LUFS above the background audio level ($M = 7.4$ s) being lower than those with no SNR ($M = 9.5$ s). However, the background audio and SNR interaction was not significant for the 3D audio group ($F(1, 71.729) = 1.3586, p = 0.248$). Finally, the type of auditory stimulus yielded a significant effect on the results of localization times for both groups: 3D audio ($F(1, 68.778) = 5.545, p = 0.021$) and mono audio ($F(1, 84) = 4.930, p = 0.029$), showing participants taking longer to locate social stimuli (3D audio $M = 9.5$ s, mono controls $M = 19.33$ s) than simple ($M = 7.3$ s, mono controls $M = 16.17$ s).

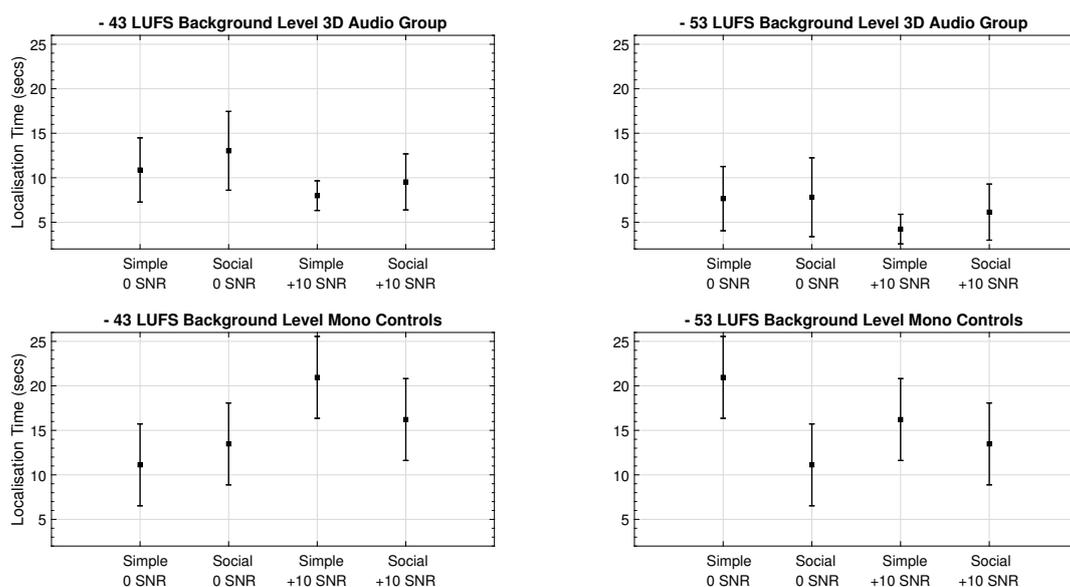


Figure 9. Estimated means data from hierarchical linear model, comparing the effect of background audio, signal-to-noise ratio, and stimulus type for binaural spatial audio and mono control groups. The whiskers denote 95% confidence intervals.

5. Discussion

5.1. Phase One: Spatial Audio Attention Testing

Throughout Phase One, subjects in both groups displayed overall attention towards the presented auditory targets; however, those within the 3D audio group displayed showed higher levels of head orientation in the direction of audio events. These results demonstrate that despite reported impairments in interpreting ITD and ILD binaural cues associated with autism [65], spatial audio was capable of accurately attracting participants to the areas of the virtual environment applicable to the auditory stimulus's location.

Individuals with autism have been recorded as having similar evoked brain responses to speech-like stimuli than their typically developed peers, but at the same exhibit less involuntary attention towards it [66,67]. With this in mind, it would be expected that under the test paradigms of free exploration that measures involuntary attention, accuracy scores for speech stimuli to be the lowest. However, the findings from Phase One of this experiment are contrary to this for both condition groups, with the auditory event with the lowest scores being the crow (event 6). This auditory stimulus type would appear to be similar to the ambient environmental audio track and despite the lack of competing audio present the event was not unique enough to warrant participant attention. In the case of attention towards speech-like sounds in Phase One, these events are a combination of explosion and non-human speech-like stimuli, designed to create a unique audio cue that attracted attention towards them. Furthermore, studies have observed hyperactivity in the brain with response to novel auditory stimuli in children on the autistic spectrum [68].

Research by Lokki and Grohn [69] revealed that despite audio cues being less accurate for sound source localization than accompanying visual cues, within a virtual environment audio-only cues are almost equally effective as visual only cues. In addition, when objects were placed outside of the participants field of view or subjected to occlusion of other visual objects, audio-only cues were more successful in the preliminary phase of object localization. This could serve as justification for the significant effect visuals accompaniments had on the overall accuracy scores for participants in the mono audio group. The visual representations would compensate for the spatial inaccuracies of the audio rendering. This would also account for the similar scores present in both groups for event three (e.g., bubble sound). This incorporated visuals that spanned a large area of the VE which would attract user attention once it arrived into the participant field of view.

Exploration of the virtual environment is also illustrated in the positional data density plot (see Figure 7). For both conditions tracked horizontal movement was concentrated close to the starting position of $(x, y = 0, 0)$. Nonetheless, participants listening to spatialized audio displayed clear movement in the direction of both the auditory only and audio-visual cues. Due to the grouping of both audio-visual and audio-only events it cannot be confirmed which events encouraged participant exploration in those areas. However, the reduced amount of tracked data in the direction of event three does suggest that the larger visual stimuli negate the need for participant exploration of the VE.

5.2. Phase Two: Spatial Audio Localization and Background Noise Testing

Alongside comparing localization times between auditory test conditions, Phase Two of this study was carried out to determine if reported impairments in decoding speech when background noise is present would have quantifiable effects on how they would respond to similar stimuli rendered via virtual spatial audio within a VR environment. Considering the heterogeneity of disorders in the autism spectrum present within the sample group, it is encouraging to see that all participants were capable of performing this task without any particular difficulties. However, the differences in performance between participants within the experimental groups may also provide explanation for the low effect size yielded by the two-way mixed ANOVA.

Firstly, results in Phase Two showed those in the 3D audio group performing significant better in the localization task. These results are consistent with similar research conducted with neurologically

healthy participants, reporting that three-dimensional auditory displays can reduce localization times and improve audio spatial attention in VR [70]. Furthermore, these results provide more evidence that those with ASD are capable of successfully interpreting binaural information to take advantage of the spatial acuity provided by spatial audio rendering techniques.

In regards to the effect of background levels on selective hearing, results showed that localization times were significantly longer when the level of background noise increased and/or the SNR between the auditory target and ambient noise decreased. Furthermore, the type of stimuli also had a noticeable impact on localization times, with participants taking longer to correctly locate speech-like stimuli. This data is comparable to evidence of speech processing deficits in children with ASD in non-spatialized audio test conditions [44,52,67]. Although the results from the mono control group would be sufficient in evaluating the effects of background noise on the response to target audio, mono audio is rarely used with virtual reality environments. Therefore, for this study it was important that competing background audio had a significant effect on both experiment conditions.

Individuals with ASD tend to demonstrate elevated speech perception thresholds, poor temporal resolution and poor frequency selectivity when compared to their typically developed peers [71]. The poor frequency selectivity alone could account for poor localization along the vertical plane [52], as this is primarily attributed to the analysis of the spectral composition of the sound arriving at the ear [49]. However, a combination of all three alongside a diminished ability to correctly translate ILD and ITD sound cues would account for poor selective hearing leading to difficulties with behavioral recognition of auditory targets in background noise [42,71].

6. Limitations and Future Work

One limitation of this study is the lack of a comparable typically developed control group. Comparison data may have highlighted any potential differences in performance between participants with autism and neurologically healthy counterparts. Nonetheless, participants in this study still showed higher spatial attention towards binaurally rendered spatial audio. Demonstrating quantifiable evidence of audio spatial interactions.

Another important consideration is the use of a non-individualized HRTF's database within this study. This database incorporates anechoic recordings using the Neumann KU-100 dummy head simulator which is based upon average dimensions of the human head and ears [28]. HRTF's differ greatly between individuals, due to the unique shape of the pinnae, head, and torso. Therefore, the use of a generalized HRTF can sometimes result in localization confusion and reduced externalization of target auditory events within a virtual environment [72,73]. In addition, recent research which compared individualized to non-individualized HRTF's did observe higher localization errors when using a generic database [74]. This may provide some explanation for the varying accuracy and localization times between participants within the 3D audio experiment group as well as the lower effect scores yielded in the statistical analysis. However, there are significant challenges involved in obtaining individualized recordings which warrants the use of generic dummy head recordings in the development of virtual reality soundscapes [75].

Future investigations could also build upon the work carried out by Wallace et al. [55] and measure self-reported presence within virtual environments that use spatial audio rendering techniques. Changes in behavior, spatial focus, and interactions when altering the audio reproduction could be compared. Furthermore, studies could be conducted to measure if removing aspects of the background audio has any negative effects on the feelings of presence felt by the participants.

7. Conclusions

Alongside realistic graphical rendering and natural approaches to computer interaction, spatialized audio can be used to significantly enhance presence and improve user interaction within a virtual reality environment [76]. In terms of clinical applications within autism research, presenting a more realistic experience would benefit interventions such as the treatment of phobias and vocational

training. By matching the visual and auditory sensory experiences of the real world, users will have a greater chance of generalizing newly acquired skills into their everyday lives, therefore increasing the possible positive outcomes of VR-based therapy [77].

To date, this study is the first to assess how those with autism respond to virtual spatial audio within a virtual reality game environment. The experiment involved 29 individuals diagnosed with autism being randomly allocated to 3D audio and mono control groups. All participants completed two experimental phases. The first compared spatial attention towards auditory stimuli by recording head rotation accuracy and horizontal movement towards presented audio. In Phase Two, participants were required to localize four social and four simple stimuli among varying levels of background audio. Results collated from head rotation and tracked positional data during the experiment do suggest that despite reported auditory processing difficulties, children with autism can make use of auditory cues rendered by spatial audio. In addition, the amount of background audio can have significant effects on the localization of auditory stimuli within a VE. These findings could provide valuable insight to those designing virtual reality interventions for the autistic population. Possible developers should make use of similar binaural-based spatial audio rendering approaches used in this study to increase the ecological validity of the virtual environment, deliver important information via auditory cues.

This study has shown that previously reported difficulties in auditory scene analysis associated with autism do extend into the realms of virtual reality and binaural-based spatial audio. The amount of competing background audio can have an effect on spatial attention towards virtual sound targets and so therefore this should be taken into consideration when designing three-dimensional virtual acoustic environments for ASD interventions.

Author Contributions: Conceptualization, D.J. and G.K.; Methodology, D.J.; Software, D.J.; Validation, D.J., H.E. and G.K.; Formal Analysis, D.J., H.E.; Investigation, D.J.; Resources, D.J.; Data Curation, D.J.; Writing—Original Draft Preparation, D.J.; Writing—Review & Editing, D.J., H.E. and G.K.; Visualization, D.J.; Supervision, H.E. and G.K.; Project Administration, D.J.; Funding Acquisition, G.K.

Funding: Funding was provided by a UK Engineering and Physical Sciences Research Council (EPSRC) Doctoral Training Award, via the Department of Electronic Engineering at the University of York, EPSRC Grant Number: EP/N509802/1.

Acknowledgments: The authors would also like to thank Accessible Arts & Media (York, UK), Springwater School (Harrogate, UK) and Riverside School (Goole, UK) for their participation and support of this research.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ASD	Autism Spectrum Disorders
CVE	Collaborative Virtual Environment
HMD	Head Mounted Display
SNR	Signal-To-Noise Ratio
TD	Typically Developed
VE	Virtual Environment
VR	Virtual Reality

References

1. Won, H.; Mah, W.; Kim, E. Autism spectrum disorder causes, mechanisms, and treatments: Focus on neuronal synapses. *Front. Mol. Neurosci.* **2013**, *6*, 19. [[CrossRef](#)] [[PubMed](#)]
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*; American Psychiatric Pub: Philadelphia, PA, USA, 2013.

3. Lord, C.; Cook, E.H.; Leventhal, B.L.; Amaral, D.G. *Autism Spectrum Disorders*; Elsevier: Amsterdam, The Netherlands, 2000; Volume 28, pp. 355–363.
4. Weitlauf, A.S.; McPheeters, M.L.; Peters, B.; Sathe, N.; Travis, R.; Aiello, R.; Williamson, E.; Veenstra-VanderWeele, J.; Krishnaswami, S.; Jerome, R.; et al. *Therapies for Children with Autism Spectrum Disorder*; Agency for Healthcare Research and Quality (US): Rockville, MD, USA, 2014.
5. Kandalaf, M.R.; Didehban, N.; Krawczyk, D.C.; Allen, T.T.; Chapman, S.B. *Virtual Reality Social Cognition Training for Young Adults with High-Functioning Autism*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 43, pp. 34–44.
6. Max, M.L.; Burke, J.C. Virtual reality for autism communication and education, with lessons for medical training simulators. *Stud. Health Technol. Inform.* **1997**, *39*, 46–53. [[PubMed](#)]
7. Fino, R.; Lin, M.J.; Caballero, A.; Balahadia, F.F. Disaster Awareness Simulation for Children with Autism Spectrum Disorder Using Android Virtual Reality. *J. Telecommun. Electron. Comput. Eng. (JTEC)* **2017**, *9*, 59–62.
8. Cox, D.J.; Brown, T.; Ross, V.; Moncrief, M.; Schmitt, R.; Gaffney, G.; Reeve, R. *Can Youth with Autism Spectrum Disorder Use Virtual Reality Driving Simulation Training to Evaluate and Improve Driving Performance? An Exploratory Study*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 47, pp. 2544–2555.
9. Lamash, L.; Klinger, E.; Josman, N. Using a virtual supermarket to promote independent functioning among adolescents with Autism Spectrum Disorder. In Proceedings of the 2017 International Conference on Virtual Rehabilitation (ICVR), Montreal, QC, Canada, 19–22 June 2017; pp. 1–7.
10. Strickland, D.; Marcus, L.M.; Mesibov, G.B.; Hogan, K. *Brief Report: Two Case Studies Using Virtual Reality as a Learning Tool for Autistic Children*; Springer: Berlin/Heidelberg, Germany, 1996; Volume 26, pp. 651–659.
11. Newbutt, N.; Sung, C.; Kuo, H.J.; Leahy, M.J.; Lin, C.C.; Tong, B. *Brief Report: A Pilot Study of the Use of a Virtual Reality Headset in Autism Populations*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 46, pp. 3166–3176.
12. Johnston, D.; Egermann, H.; Kearney, G. Innovative computer technology in music-based interventions for individuals with autism moving beyond traditional interactive music therapy techniques. *Cogent Psychol.* **2018**, *5*, 1554773. [[CrossRef](#)]
13. Leonard, A.; Mitchell, P.; Parsons, S. Finding a place to sit: A preliminary investigation into the effectiveness of virtual environments for social skills training for people with autistic spectrum disorders. In Proceedings of the 4th International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT 2002), Veszprem, Hungary, 18–21 September 2002.
14. Silver, M.; Oakes, P. *Evaluation of a New Computer Intervention to Teach People With Autism or Asperger Syndrome to Recognize and Predict Emotions in Others*; Sage Publications: Thousand Oaks, CA, USA, 2001; Volume 5, pp. 299–316.
15. Iovannone, R.; Dunlap, G.; Huber, H.; Kincaid, D. *Effective Educational Practices for Students With Autism Spectrum Disorders*; Sage Publications: Los Angeles, CA, USA, 2003; Volume 18, pp. 150–165.
16. Stiegler, L.N.; Davis, R. *Understanding Sound Sensitivity in Individuals with Autism Spectrum Disorders*; Sage Publications: Los Angeles, CA, USA, 2010; Volume 25, pp. 67–75.
17. Keay-Bright, W. *Can Computers Create Relaxation? Designing ReacTickles© Software with Children on the Autistic Spectrum*; Taylor & Francis: Abingdon, UK, 2007; Volume 3, pp. 97–110.
18. Larsson, P.; Våljamäe, A.; Västfjäll, D.; Tajadura-Jiménez, A.; Kleiner, M. Auditory-Induced Presence in Mixed Reality Environments and Related Technology. *Eng. Mixed Real. Syst. Hum.-Comput. Interact. Ser.* **2010**, *10*, 143–163.
19. Stanney, K.M.; Hale, K.S. *Handbook of Virtual Environments: Design, Implementation, and Applications*; CRC Press: Boca Raton, FL, USA, 2014.
20. Hermann, T.; Ritter, H. Sound and meaning in auditory data display. *Proc. IEEE* **2004**, *92*, 730–741. [[CrossRef](#)]
21. Hendrix, C.; Barfield, W. The sense of presence within auditory virtual environments. *Presence Teleoperators Virtual Environ.* **1996**, *5*, 290–301. [[CrossRef](#)]
22. Kohlrausch, A.; van de Par, S. Auditory-visual interaction: From fundamental research in cognitive psychology to (possible) applications. In Proceedings of the Human Vision and Electronic Imaging IV, International Society for Optics and Photonics, San Jose, CA, USA, 23–29 January 1999; Volume 3644, pp. 34–45.

23. Larsson, P.; Vastfjäll, D.; Kleiner, M. Better presence and performance in virtual environments by improved binaural sound rendering. In Proceedings of the Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio, Espoo, Finland, 15–17 June 2002; Audio Engineering Society: New York, NY, USA, 2002.
24. Baumgarte, F.; Faller, C. Binaural cue coding—Part I: Psychoacoustic fundamentals and design principles. *IEEE Trans. Speech Audio Process.* **2003**, *11*, 509–519. [[CrossRef](#)]
25. Yin, T.C. Neural mechanisms of encoding binaural localization cues in the auditory brainstem. In *Integrative Functions in the Mammalian Auditory Pathway*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 99–159.
26. Zwislocki, J.; Feldman, R. Just noticeable differences in dichotic phase. *ASA* **1956**, *28*, 860–864. [[CrossRef](#)]
27. Handel, S. *Listening: An Introduction to the Perception of Auditory Events*; MIT Press: Cambridge, MA, USA, 1993; pp. 99–112.
28. Kearney, G.; Doyle, T. An HRTF database for virtual loudspeaker rendering. In Proceedings of the Audio Engineering Society Convention 139, New York, NY, USA, 29 October–1 November 2015; Audio Engineering Society: New York, NY, USA, 2015.
29. Musicant, A.D.; Butler, R.A. The influence of pinnae-based spectral cues on sound localization. *J. Acoust. Soc. Am.* **1984**, *75*, 1195–1200. [[CrossRef](#)]
30. Howard, D.M.; Angus, J.A.S.; Wells, J.J. *Acoustics and Psychoacoustics*; Focal Press: Waltham, MA, USA, 2013; pp. 107–119.
31. Larsson, P.; Våljamäe, A.; Västfjäll, D.; Tajadura-Jiménez, A.; Kleiner, M. Auditory-induced presence in mixed reality environments and related technology. In *The Engineering of Mixed Reality Systems*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 143–163.
32. Potisk, T. Head-related transfer function. In *Seminar Ia, Faculty of Mathematics and Physics*; University of Ljubljana: Ljubljana, Slovenia, 2015.
33. Zhong, X.L.; Xie, B.S. Head-Related Transfer Functions and Virtual Auditory Display. In *Soundscape Semiotics—Localisation and Categorisation*; IntechOpen: London, UK, 2014. [[CrossRef](#)]
34. Cheng, C.I.; Wakefield, G.H. Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space. In Proceedings of the Audio Engineering Society Convention 107, New York, NY, USA, 24–27 September 1999; Audio Engineering Society: New York, NY, USA, 1999.
35. Hayes, S.T.; Adams, J.A. Device Motion via Head Tracking for Mobile Interaction. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2016**, *60*, 774–778. [[CrossRef](#)]
36. Zotkin, D.N.; Duraiswami, R.; Davis, L.S. Rendering localized spatial audio in a virtual auditory space. *IEEE Trans. Multimed.* **2004**, *6*, 553–564. [[CrossRef](#)]
37. Chang, M.C.; Parham, L.D.; Blanche, E.I.; Schell, A.; Chou, C.P.; Dawson, M.; Clark, F. *Autonomic and Behavioral Responses of Children with Autism to Auditory Stimuli*; American Occupational Therapy Association: Bethesda, MD, USA, 2012; Volume 66, pp. 567–576.
38. Bishop, S.L.; Hus, V.; Duncan, A.; Huerta, M.; Gotham, K.; Pickles, A.; Kreiger, A.; Buja, A.; Lund, S.; Lord, C. *Subcategories of Restricted and Repetitive Behaviors in Children with Autism Spectrum Disorders*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 43, pp. 1287–1297.
39. Linke, A.C.; Keehn, R.J.J.; Pueschel, E.B.; Fishman, I.; Müller, R.A. *Children with ASD Show Links between Aberrant Sound Processing, Social Symptoms, and Atypical Auditory Interhemispheric and Thalamocortical Functional Connectivity*; Elsevier: Amsterdam, The Netherlands, 2018; Volume 29, pp. 117–126.
40. Heaton, P.; Hudry, K.; Ludlow, A.; Hill, E. *Superior Discrimination of Speech Pitch and Its Relationship to Verbal Ability in Autism Spectrum Disorders*; Taylor & Francis: Abingdon, UK, 2008; Volume 25, pp. 771–782.
41. Lucker, J.R. *Auditory Hypersensitivity in Children with Autism Spectrum Disorders*; SAGE Publications: Los Angeles, CA, USA, 2013; Volume 28, pp. 184–191.
42. Lin, I.F.; Shirama, A.; Kato, N.; Kashino, M. The singular nature of auditory and visual scene analysis in autism. *Phil. Trans. R. Soc. B* **2017**, *372*, 20160115. [[CrossRef](#)] [[PubMed](#)]
43. Dawson, G.; Toth, K.; Abbott, R.; Osterling, J.; Munson, J.; Estes, A.; Liaw, J. Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Am. Psychol. Assoc.* **2004**, *40*, 271–283. [[CrossRef](#)] [[PubMed](#)]
44. Soskey, L.N.; Allen, P.D.; Bennetto, L. *Auditory Spatial Attention to Speech and Complex Non-Speech Sounds in Children with Autism Spectrum Disorder*; Wiley Online Library: Hoboken, NJ, USA, 2017; Volume 10, pp. 1405–1416.

45. Rochat, P.; Striano, T. Social-cognitive development in the first year. In *Early Social Cognition: Understanding Others in the First Months of Life*; Lawrence Erlbaum Associates: London, UK, 1999; pp. 3–34.
46. Morton, J.; Johnson, M.H. CONSPEC and CONLERN: A two-process theory of infant face recognition. *Am. Psychol. Assoc.* **1991**, *98*, 164–181. [[CrossRef](#)]
47. Boddaert, N.; Belin, P.; Chabane, N.; Poline, J.B.; Barthélémy, C.; Mouren-Simeoni, M.C.; Brunelle, F.; Samson, Y.; Zilbovicius, M. Perception of complex sounds: Abnormal pattern of cortical activation in autism. *Am. Psychiatr. Assoc.* **2003**, *160*, 2057–2060. [[CrossRef](#)] [[PubMed](#)]
48. Dawson, G.; Meltzoff, A.N.; Osterling, J.; Rinaldi, J.; Brown, E. *Children with Autism Fail to Orient to Naturally Occurring Social Stimuli*; Springer: Berlin/Heidelberg, Germany, 1998; Volume 28, pp. 479–485.
49. McAlpine, D. *Creating a Sense of Auditory Space*; Wiley Online Library: Hoboken, NJ, USA, 2005; Volume 566, pp. 21–28.
50. Kulesza, R.J.; Mangunay, K. *Morphological Features of the Medial Superior Olive in Autism*; Elsevier: Amsterdam, The Netherlands, 2008; Volume 1200, pp. 132–137.
51. Kulesza, R.J., Jr.; Lukose, R.; Stevens, L.V. *Malformation of the Human Superior Olive in Autistic Spectrum Disorders*; Elsevier: Amsterdam, The Netherlands, 2011; Volume 1367, pp. 360–371.
52. Visser, E.; Zwiers, M.P.; Kan, C.C.; Hoekstra, L.; van Opstal, A.J.; Buitelaar, J.K. *Atypical Vertical Sound Localization and Sound-Onset Sensitivity in People with Autism Spectrum Disorders*; Canadian Medical Association: Ottawa, ON, Canada, 2013; Volume 38, p. 398.
53. Slater, M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. R. Soc. Biol. Sci.* **2009**, *364*, 3549–3557. [[CrossRef](#)] [[PubMed](#)]
54. Rummukainen, O.; Robotham, T.; Schlecht, S.J.; Plinge, A.; Herre, J.; Habels, E.A. Audio Quality Evaluation in Virtual Reality: Multiple Stimulus Ranking with Behavior Tracking. In Proceedings of the Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality, Redmond, WA, USA, 20–22 August 2018; Audio Engineering Society: New York, NY, USA, 2018.
55. Wallace, S.; Parsons, S.; Bailey, A. Self-reported sense of presence and responses to social stimuli by adolescents with autism spectrum disorder in a collaborative virtual reality environment. *J. Intellect. Dev. Disabil.* **2017**, *42*, 131–141. [[CrossRef](#)]
56. Gardner, W.G. Head tracked 3-d audio using loudspeakers. In Proceedings of the 1997 Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 19–22 October 1997; p. 4.
57. Oculus Rift Head Mounted Display. Available online: <https://www.oculus.com/rift/> (accessed on 2 January 2019).
58. Bozgeyikli, L. Virtual Reality Serious Games for Individuals with Autism Spectrum Disorder: Design Considerations. Ph.D. Thesis, University of South Florida, Tampa, FL, USA, 2016; pp. 180–183.
59. Leekam, S.R.; Nieto, C.; Libby, S.J.; Wing, L.; Gould, J. *Describing the Sensory Abnormalities of Children and Adults With Autism*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 37, pp. 894–910.
60. Liss, M.; Mailloux, J.; Erchull, M.J. *The Relationships between Sensory Processing Sensitivity, Alexithymia, Autism, Depression, and Anxiety*; Elsevier: Amsterdam, The Netherlands, 2008; Volume 45, pp. 255–259.
61. AudioKinetic Wwise. Available online: <https://www.audiokinetic.com/products/wwise> (accessed on 10 September 2018).
62. Google Resonance—Spatial Audio SDK. Available online: <https://developers.google.com/resonance-audio/> (accessed on 30 June 2018).
63. Johnston, D.; Egermann, H.; Kearney, G. An Interactive Spatial Audio Experience for Children with Autism Spectrum Disorder. In Proceedings of the Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio, York, UK, 27–29 March 2019; Audio Engineering Society: New York, NY, USA, 2019.
64. Ward, L.; Shirley, B.; Francombe, J. Accessible object-based audio using hierarchical narrative importance metadata. In Proceedings of the Audio Engineering Society Convention 145, New York, NY, USA, 17–20 October 2018; Audio Engineering Society: New York, NY, USA, 2018.
65. Kashino, M.; Furukawa, S.; Nakano, T.; Washizawa, S.; Yamagishi, S.; Ochi, A.; Nagaike, A.; Kitazawa, S.; Kato, N. Specific deficits of basic auditory processing in high-functioning pervasive developmental disorders. In Proceedings of the 36th ARO (Association for Research in Otolaryngology) MidWinter Meeting, Baltimore, MD, USA, 16–20 February 2013.

66. Čeponienė, R.; Lepistö, T.; Shestakova, A.; Vanhala, R.; Alku, P.; Näätänen, R.; Yaguchi, K. Speech–sound-selective auditory impairment in children with autism: they can perceive but do not attend. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5567–5572. [[CrossRef](#)]
67. Whitehouse, A.J.; Bishop, D.V. *Do Children with Autism ‘Switch Off’ to Speech Sounds? An Investigation Using Event-Related Potentials*; Wiley Online Library: Hoboken, NJ, USA, 2008; Volume 11, pp. 516–524.
68. Gomot, M.; Belmonte, M.K.; Bullmore, E.T.; Bernard, F.A.; Baron-Cohen, S. *Brain Hyper-Reactivity to Auditory Novel Targets in Children with High-Functioning Autism*; Oxford University Press: Oxford, UK, 2008; Volume 131, pp. 2479–2488.
69. Lokki, T.; Grohn, M. Navigation with auditory cues in a virtual environment. *IEEE MultiMedia* **2005**, *12*, 80–86. [[CrossRef](#)]
70. Hoeg, E.R.; Gerry, L.J.; Thomsen, L.; Nilsson, N.C.; Serafin, S. Binaural sound reduces reaction time in a virtual reality search task. In Proceedings of the 2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE), Los Angeles, CA, USA, 19 March 2017; pp. 1–4.
71. Alcántara, J.I.; Weisblatt, E.J.; Moore, B.C.; Bolton, P.F. *Speech-in-Noise Perception in High-Functioning Individuals With Autism or Asperger’s Syndrome*; Wiley Online Library: Hoboken, NJ, USA, 2004; Volume 45, pp. 1107–1114.
72. Seeber, B.U.; Fastl, H. *Subjective Selection of Non-Individual Head-Related Transfer Functions*; Georgia Institute of Technology: Atlanta, GA, USA, 2003.
73. Wenzel, E.M.; Arruda, M.; Kistler, D.J.; Wightman, F.L. Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.* **1993**, *94*, 111–123. [[CrossRef](#)] [[PubMed](#)]
74. Rudzki, T.; Gomez-Lanzaco, I.; Stubbs, J.; Skoglund, J.; Murphy, D.T.; Kearney, G. Auditory Localization in Low-Bitrate Compressed Ambisonic Scenes. *Appl. Sci.* **2019**, *9*, 2618. [[CrossRef](#)]
75. Hur, Y.M.; Park, Y.C.; Lee, S.P.; Youn, D.H. Efficient Individualization Method of HRTFs Using Critical-band Based Spectral Cue Control. *Acoust. Soc. Korea* **2011**, *30*, 167–180. [[CrossRef](#)]
76. Naef, M.; Staadt, O.; Gross, M. Spatialized audio rendering for immersive virtual environments. In Proceedings of the ACM Symposium on Virtual Reality Software and Technology, Hong Kong, China, 11–13 November 2002; pp. 65–72.
77. Dautenhahn, K. Design issues on interactive environments for children with autism. In Proceedings of the ICDVRAT 2000, the 3rd Int Conf on Disability, Virtual Reality and Associated Technologies, Alghero, Sardinia, 23–25 September 2000; University of Reading: Reading, UK, 2000.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).