Aalborg University Copenhagen

Aalborg University Copenhagen
Frederikskaj 12,
DK-2450 Copenhagen SV
Semester Coordinator:
Secretary:

Semester: 10th

Title:
Social Skills Training for Autistic Children:
Bringing Music Therapy into Immersive Virtual Reality

Project Period:
1st Feb-28th June

Semester Theme:
Master thesis

Supervisor(s):

Ali Adjorlu

Member:

Linnea Bjerregaard Pedersen

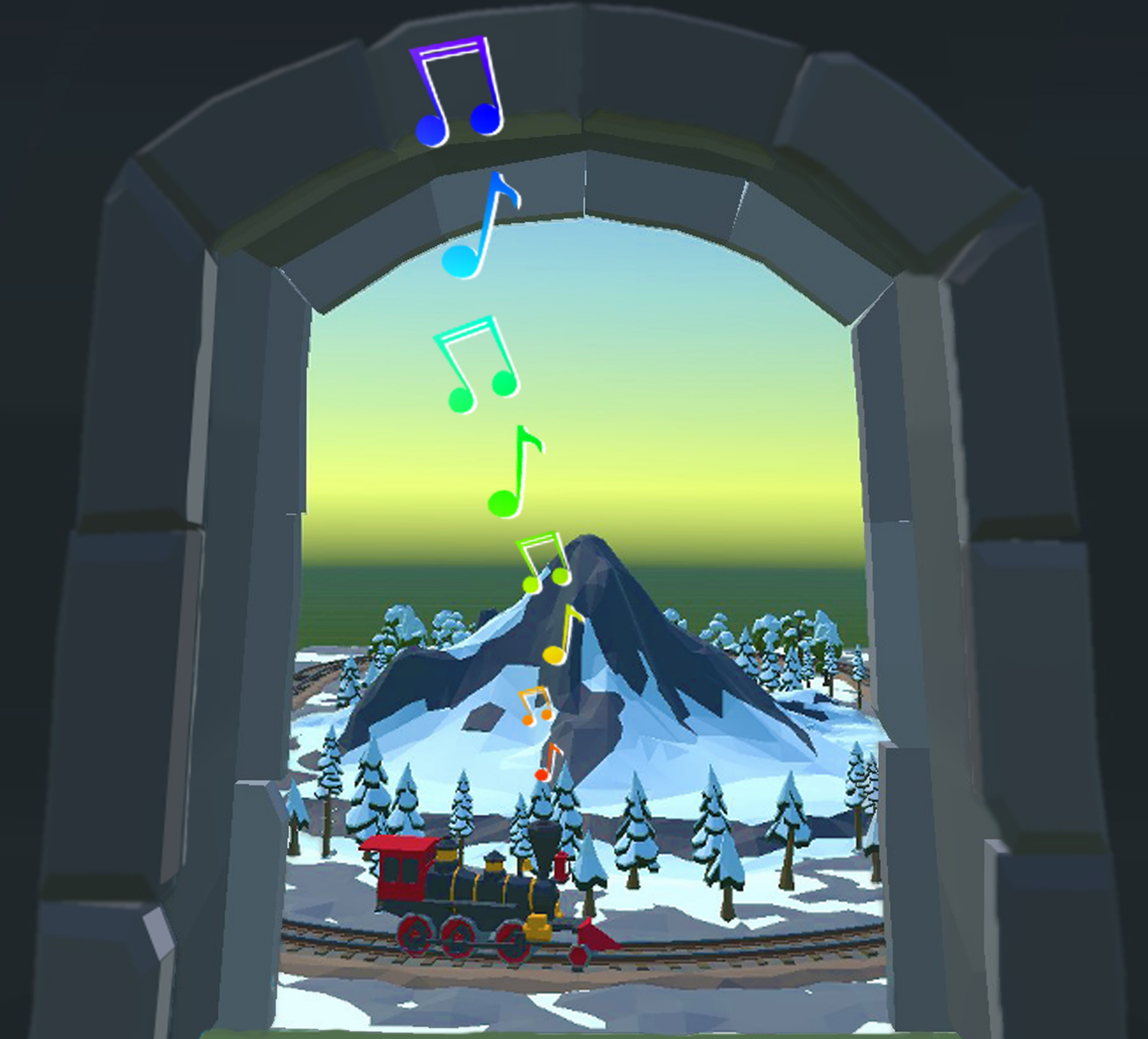
Copies: 1
Pages: 25
Finished: 25th May

Abstract:

Autism spectrum condition (ASC) comes with persistent challenges in social and communication skills, that lead to social withdrawal and anxiety, reducing the possibilities of the individual becoming a productive member of society.
To train social skills, an immersive virtual reality (IVR) therapeutic tool was designed in cooperation with autism experts. The IVR tool is based on the basic principles of improvisational music therapy than facilitates social and communication skills. The tool further exploits the capability of IVR technology of prompting valid reactions, which increases ecological validity.
Sessions with the IVR tool were conducted with autistic children to observe their social interaction abilities. To assess said abilities, a questionnaire was formulated from established music therapeutic assessment methods.
Comparing the IVR sessions to standard mediation sessions, the participants showcased significantly higher levels of social interaction in the IVR sessions. Hence, the IVR tool was deemed effective in facilitating social interaction.

MUSIC QUEST

Social Skills Training for Autistic Children: Bringing Music Therapy into Virtual Reality



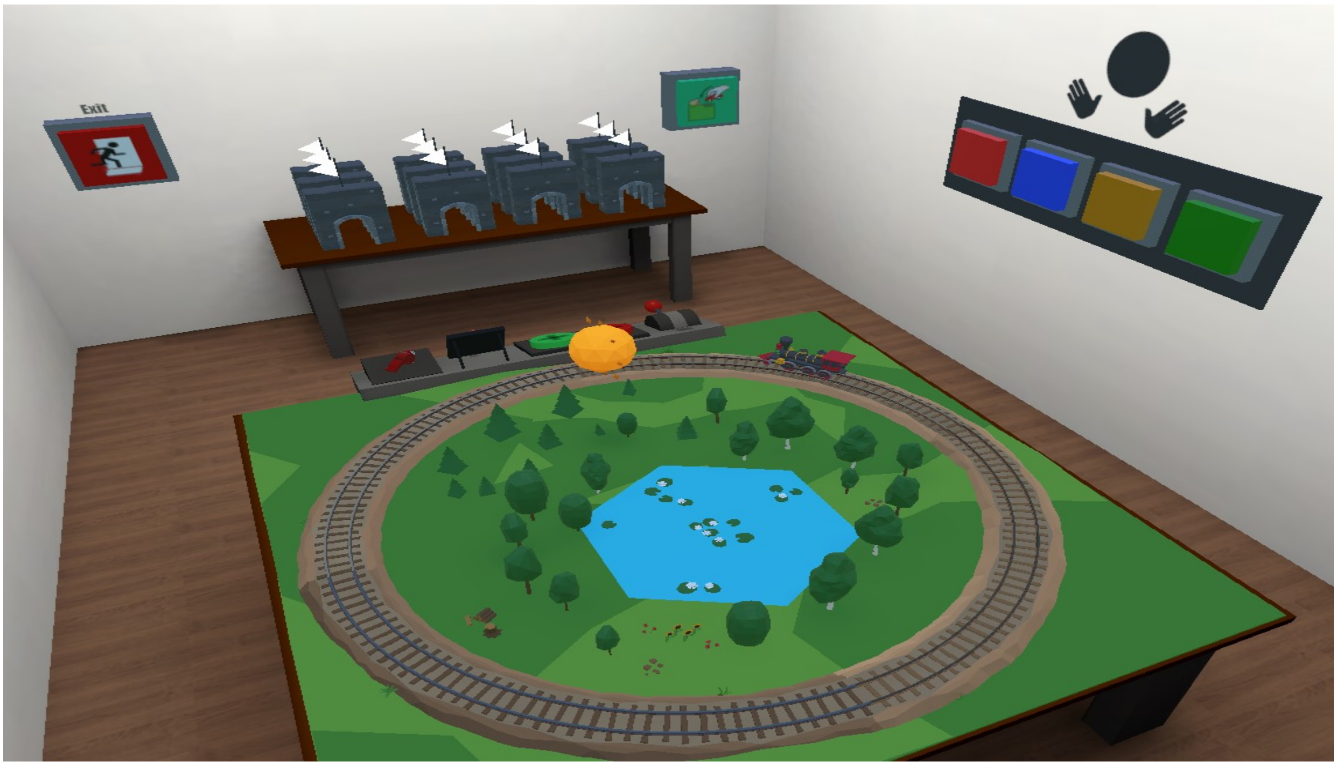
Linnea Bjerregaard Pedersen

Master Thesis, Summer 2022

Aalborg University in Copenhagen

Social Skills Training for Autistic Children: Bringing Music Therapy into Immersive Virtual Reality

Linnea Bjerregaard Pedersen
Aalborg University Copenhagen
lbpe17@student.aau.dk



The virtual environment used in the interventions with autistic children.

Abstract

Autism spectrum condition (ASC) comes with persistent challenges in social and communication skills, that lead to social withdrawal and anxiety, reducing the possibilities of the individual becoming a productive member of society. To train social skills, an immersive virtual reality (IVR) therapeutic tool was designed in cooperation with autism experts. The IVR tool is based on the basic principles of improvisational music therapy than facilitates social and communication skills. The tool further exploits the capability of IVR technology of prompting valid reactions, which increases ecological validity. Sessions with the IVR tool were conducted with autistic children to observe their social interaction abilities. To assess said abilities, a questionnaire was formulated from established music therapeutic assessment methods. Comparing the IVR sessions to standard mediation sessions, the participants showcased significantly higher levels of social interaction in the IVR sessions. Hence, the IVR tool was deemed effective in facilitating social interaction.

1 Introduction

Autism spectrum condition (ASC) is a neurodevelopmental condition that affects one in 100 individuals [3]. The diagnosis comes with challenges, especially in the social-emotional domain and difficulties with creating and maintaining meaningful relationships [8, 23]. A significantly higher rate of autistic adolescents reports social anxiety, depression, suicidal thought patterns, and other forms of psychopathology compared to their non-autistic peers [1, 8]. Because of the challenges autistic individuals experience, the majority remain dependent on their parents and social agencies for support into adulthood [1]. Therefore, to increase the likelihood of social involvement for improving the level of functioning and independence, and thus overall life quality, interventions for training social and communication skills are imperative [8, 23, 30, 36]. Several methods have been studied aiming to improve the recipient's social skills, including the use of social stories [27], social robots [2, 67], and video modelling [60]. Another method is music therapy (MT) - a method that can bypass linguistic impairments and facilitate pre-verbal and social communication skills [31, 33]. This is possible due to the innate 'communicative musicality', which originates far back in human history and thus the evolution of the brain. Communicative musicality is seen in early interactions between child and parents and is based on rhythm, pulse, timing, etc., to where the infant develops the ability to participate in non-verbal communication [31]. For supporting individuals who are challenged in developing this ability, MT can provide an intensified version of such interactions to prompt pre-verbal skills, i.e., imitation, taking initiative, and joint attention [31, 33]. Joint attention is the skill of sharing experiences with other people by making gestures and facial expressions pertinent to the situation [31, 57]. Developing joint attention is highly dependent on the individual's ability to imitate other people, which also is necessary for developing language skills [31]. Especially, early MT interventions are seen to facilitate social communication skills [30, 31].

A limitation within the teaching strategies for social skills training is achieving a generalisation and application of the learned skills in natural situations outside the therapeutic environment [35]. Here, immersive virtual reality (IVR) has shown great potential, as the technology can simulate convincing environments and enable activities that resemble the recipient's daily life tasks [23, 34, 48, 50]. Furthermore, IVR allows the therapist to assess the recipient in real-time and thus intervene appropriately to the valid reactions prompted by the multimodal stimuli provided by the IVR technology. Multimodal stimuli are important because social interactions involve the convergence of multiple sensory systems for perceiving social cues [50]. These include visual inputs such as gestures and poses, and auditory inputs, such as verbal utterances and prosody (increasing or decreasing intonation, weaker volume, slower tempo) [31, 50]. Therefore, providing multimodal stimuli is important for ecological validity, hence generalising the learned skills to other situations [34, 50].

This article continues my previous study: *Exploring the Possibilities of Music Therapy in Virtual Reality: Social Skills Training for Adolescents with Autism Spectrum Disorder* [52] (attached as Appendix A). The previous study aimed to develop an IVR tool with music therapists (MTs), which they could utilise to train social and communication skills in autistic children. The resulting IVR is evaluated in this article by two psychologists based on IVR interventions with autistic children. The primary focus is on the system's capability of facilitating social skills for autistic children and secondarily on the design's suitability for autistic perception.

2 Background

This section investigates ASC and associated theories about the characteristic of the condition. With this, the autism perception is examined, as it can provide vital information about the perceptual challenges that autistic individuals have and thus aid in preparing the IVR tool for autism interventions. From here, the section continues into a study about MT and its impact on autistic individuals. The final segment of the section is about IVR technology, its strengths, and its limitations.

2.1 Autism Spectrum Condition

ASC is a heterogeneous neurodevelopmental condition that is described in the *Diagnostic and Statistical Manual of Mental Disorders (DMS-5)* [3] as "*persistent deficits in social communication and social interaction across multiple contexts*", including social-emotional reciprocity and nonverbal communicative behaviours, which are sufficient to hamper independence in social, occupational, and other important areas of functioning [3, 38]. Diagnosis of ASC is made based on current behaviour and history following interviews with caregivers, observations made by several clinicians, and self-assessment questionnaires, if possible. Usually, it is made when the child is 12 to 24 months but can be made earlier if the child has severe symptoms [3]. Diagnosis can also be made later in life as symptoms can change with development or hide due to co-occurring conditions. Comorbidities are, amongst others, hyperactivity disorder, depressive and bipolar disorder, and anxiety disorder.

The ratio of males to females who are diagnosed with ASC is quoted as 4:1. It is hypothesised that because females tend to have less salient symptoms, it is not as often discovered [3]. Yet, when researching the socio-communicative differences between males and females on the autism spectrum, no significant difference is found [59].

Manifestations of ASC are associated with stimming, hyper-/hyporeactivity to sensory stimuli, and an intense interest in the environment's colours, odour, textures, etc. [3]. Based on the severity of the symptoms, which are placed on a continuum ranging from mild to severe, the severity level of the diagnosis is specified and is usually the determinant for what level of support the individual needs [3].

2.1.1 Theories on Autism

The three most influential theories that seek to explain autism or aspects of autism are executive dysfunction, weak central coherence theory, and Theory of Mind deficit [42]. **Executive dysfunction** suggests that individuals on the autism spectrum have difficulties with controlling their attention and with actions that are associated with activity in the frontal lobes. This would explain the strong need for repetition, specific interests, and the difficulties autistic people experience with planning, as well as inhibiting cognitive interference (i.e. the Stroop effect) [19]. **Weak central coherence** proposes that autistic people have a limited understanding of context and instead focus on piecemeal and detailed information [19]. This manifests in autistic people having 'islets of ability' in for example mathematics, music, or remembering, i.e., railway timetable information [6]. The **lack of Theory of Mind** was formulated in the 1980s by Simon Baron-Cohen. It builds upon the dated idea that autistic individuals cannot impute mental states to themselves and others such as purpose, intention, liking, etc. [56].

The presented theories have been criticised for the lack of specificity for autism, the narrow perspective they put on the autistic community, and for setting non-autistic individuals (N-AIs) in a superior perception - a perception that leads to stigmatisation and a greater risk of causing autistic individuals to socially isolate [10, 14, 19, 38, 42]. For those reasons, Chown [14, 15] advocates for less acknowledged, yet modern theories such as Gallagher's interaction theory [24] and Klin and colleagues' hypothesis on the enactive mind [35]. The **interaction theory** is based on phenomenological and developmental studies and focuses on how people understand each other through embodied practice and interpersonal interaction. Thus, a greater emphasis is placed on the perception of the other person's bodily movements rather than reading the other person's mind, as in Theory of Mind. Because autistic individuals have a unique perception of the world, the meaning they deduce from the interactions is not the same as when deduced by N-AIs [29]. The theory of the **enactive mind** proposes a framework in which the assumption of a pre-given set of cognitive abilities to handle social interactions is discarded. It is not a matter of *knowing about* but *knowing how* to navigate social situations, which is acquired through interactions within social environments [35]. For N-AIs it seems natural to gravitate towards social engagement but for the autistic individual, the social stimuli seem less salient and less relevant. Thus the tools required for flexibly adapting to social activities are not available, which results in slow and truncated interactions [35].

It is worth mentioning that these two theories seem coherent with the philosophical view of 'embodiment', where the individual becomes skilful through bodily interactions with the environment [20]. Exactly this is relevant when designing for training skills in IVR, that allows for interactions that resemble everyday interaction.

To further address the 'neurotypical syndrome', where the non-autistic practices are prioritised and made superior, the 'double empathy problem' was formulated. It describes the miscommunication that occurs when two people of different conceptual understandings attempt to communicate a meaning [14, 42, 49]. More on this in the next segment.

2.1.2 Social Interactions

The double empathy problem explicates how it is not only the autistic perception that causes a social disjunction but that N-AIs are as estranged from expressions made by autistic individuals. In a study by Sheppard et al. [68], they suggest that the double empathy problem is caused by N-AIs perceiving autistic individuals as less expressive. To evaluate this hypothesis, non-autistic participants were asked to watch short video clips of other people's (targets') reactions and categorise them into given scenarios: being told a story, being told a joke, waiting, and being given a compliment. The group with targets consisted of an equal amount of autistic and N-AIs. The participants could not as accurately categorise the reactions made by the autistic targets as they could with the non-autistic targets. Yet, the autistic targets were not rated less expressive but more ambiguous. The exact cause of the misinterpretations is yet to be discovered. Nonetheless, it is proposed that autistic individuals teach themselves to react in certain ways to present a socially acceptable facade - hence, the reported ambiguity [68]. Keeping up such a facade is a common compensation strategy for autistic individuals, and can be stressful for the person [3, 43].

In another study, it was noticeable how different the gaze patterns of autistic and N-AIs were during a conversation. Where N-AIs usually gaze at the conversion partner while listening, and have a deviant gaze pattern while speaking, autistic individuals did the opposite [30].

These findings add to explaining the occurrence of the negative social interactions that autistic people can experience [68]. Typically, autistic individuals have heightened physiological responses to novel and social situations, which make the individual vulnerable to developing social anxiety. This impact is amplified by negative social interactions. Therefore, it is suggested that training in social skills should focus on arousal regulation and relaxation skills [8].

2.1.3 The Autistic Perception

Hyper- and hyposensitivities in multiple modalities are reported by more than 96% of individuals on the autism spectrum [39]. The distress caused by sensory stimuli can cause self-injurious and aggressive behaviour, especially in cases where the individual is unable to communicate their duress. The sensitivity can be explained by the differences between the sensory systems of autistic and N-AIs, including a measurable difference in the early auditory cortex, which causes complex stimuli to be overwhelming for the autistic individual. Hypersensitivity towards tactile information causes an aversion to light touches, as well as some thermal stimuli [38, 39]. Regarding the visual system, autistic individuals are seen to have more robust responses to high spatial frequency information (SFI) and less robust to low SFI. High SFI is related to the detailed perception of, i.e., sharp edges, whereas low SFI is about the general shape and proportions of objects. The latter is crucial for processing facial information, which is a common difficulty for autistic individuals [39, 76].

By examining the threshold for the 'flash-beep' illusion - an illusion where audio induces the perception of multiple blinks, albeit just one visual stimulus given, results showed that the autistic participants needed a greater threshold between the auditory and visual stimuli to uncouple the two, compared to the non-autistic participants [39]. This indicates inefficiency in filtering inputs accurately and processing multiple channels simultaneously. The same tendencies are seen with higher-order multisensory integration tasks, i.e., the McGurk effect, where audio is timed properly to a video of a person moving their lips in different ways, which changes the perception of the auditory input. At the first showcase of the illusion, there was no difference between the autistic and N-AIs. However, with practice, the non-autistic participants learned how to lip-read the correct sound. This was not the case for the autistic participants who were still reliant on auditory information. Marco et al. [39] propose that these perceptual differences are why autistic people often struggle with focusing their attention. In a complex environment, the many inputs cannot be processed by the autistic brain, thus resulting in "*the individual's inability to attend to their environment in a flexible, productive, and meaningful way*" [39]. The regions of the brain that processes multimodal sensory inputs have significant changes in terms of neuronal density. It is hypothesised to be the reason behind the challenges that autistic individuals have with sensory integration. These are the same regions that consist of the 'mirror neuron' network, which can explain the difficulty in imitating other people, which is fundamental for developing social skills [31, 38, 39].

Knowing about autistic perception, it is understandable why MT is used for autism interventions, as autistic individuals seem more reliant on auditory stimuli than visual. Furthermore, a connection can be made between the double empathy problem and especially improvisational MT, where the therapist and recipient make music together, as the therapist seeks to understand how the recipient communicates. Improvisational MT is studied in the following section.

2.2 Improvisational Music Therapy

Already in the 1950s-1960s, autistic children's affinity towards music was noticed [61]. Nordoff and Robbins pioneered creative MT, which later became improvisational MT [46, 61]. They proposed that music was perceived as a non-threatening medium by autistic children, whom Nordoff and Robbins had observed engaged in musical activities. Through several international cases with behaviourally challenged children, they established the theoretical foundation for improvisational MT for, amongst others, autistic children [61].

When MT is utilised in autism interventions, music can provide the needed level of predictability that is attractive for autistic individuals, yet still challenge this facet, and build musical and interactive expectations [30, 31]. These are made by contributing with variations in rhythm, harmony, creating sudden pauses, etc. Oftentimes the rhythm is created based on an activity that the child finds soothing (i.e., bouncing on a gym ball) [30]. This creates the possibility of the child recognising a connection between their actions and the music. When the connection has been established the therapist can begin to make slight changes to the allotted sound to challenge the child to gradually become flexible in the interaction as well as 'teasing' the child in a musical humorous way [31]. The MTs also imitate the child's rhythm to turn-organise their interaction. Turn-taking practice includes eye contact, nodding, facial expressions, and prosody. There are signs of turn-yielding and turn-taking cues, as well as turn-maintaining and turn-denying cues which the MTs can train with the help of using the rhythm created between themselves and the child. A sudden pause can be utilised to see how the child reacts and if the child perceives the turn-taking cue to initiate making a sound [31, 57]. Especially in the beginning of a music therapeutic process and the individual sessions, it is important to follow the child's initiatives and view them as intentional communication. Noticeable is that child-led interventions have indicated more frequent and longer events of joy and spontaneous participation from the child [25, 31]. Thus, when training communication and social skills, this should not be to 'appropriating' the autistic child's behaviour but allow for the child to express themselves and let the intervention participants learn how to communicate together [49]. In that way, the child's ability for emotional engagement and attunement can be disclosed, as well as new details about the challenges the child may experience, which otherwise can be hidden in cognitive tests and verbal interaction [25, 31]. Having attracted the child's attention - a prerequisite for interaction, it can be directed towards the social interplay which is entangled in the musical activities.

Supporting the child's desire for social participation is important, as it creates the foundation for later language development. With parental involvement, MT can be practised in the home to reinforce the learned skills in the child's daily life [31].

2.2.1 The Impact of Music Therapy

MT research primarily comprises case studies or, if quantitatively comparative, with small sample sizes. Therefore, is the validity of the research done within the field, as well as the therapeutical value of MT, questioned [40, 61].

In the systematic review by Reschke-Hernández [61], she deduced that throughout the about 80 years of using MT for children on the autism spectrum, MT assessment methods have not progressed much and the development of MT techniques has primarily relied on trial and error. Together with sparse coherent research available, the validity of MT research is limited. Reschke-Hernández hypothesises that the literature does not represent the music-therapeutic practice, as the MTs may not have engaged in clinical research or published their work. This is supported by the findings from an international survey of MTs, revealing an incongruence between the published literature and practice [4, 61]. Albeit, throughout the reviewed articles originating from the 1940s, MTs tend to include the same or similar techniques in their interventions, indicating converging and consistent methods within the discipline [61], thus adding to the combined validity of MT's impact. Yet, it is suggested to support the findings with quantitative measures, i.e., neuroimaging [11, 40].

In the literature review by Holck [31], she found that MT has a significant effective impact on autistic children compared to autistic controls that participated in play or storytelling activities. The children receiving MT, improved significantly their nonverbal communication (i.e., gestural imitation) and verbal communication, albeit less notable [31].

These findings are supported by the study of Bieleninik and colleagues [9], who utilised established methods for measuring autism severity, to compare improvisational MT and enhanced standard care (routine care available on-site). They followed 364 children on the autism spectrum over five months. The children receiving MT were either in a group of high intensity (three sessions per week) or low intensity (once per week). Measures were taken with the Autism Diagnostic Observation Schedule (ADOS) and the Social Responsiveness Scale (SRS). The first measures autism symptom severity and needs to be administered by a trained professional. The latter measures social competencies and is a parent-rated measure. According to the ADOS measure, there was no significant difference in symptom severity over the five months. Bieleninik et al. suggest that this was caused by missed sessions or that they should have extended the intervention process, as it is common to continue MT for years before it has a noticeable impact. However, a small but nominally significant difference was seen when looking at the social motivation, autistic mannerisms, and social awareness subscales in the SRS, where MT was prevailing over standard care. In addition, parents reported positive growth in their children's enjoyment and benefit of MT, and also experienced the effect of their involvement in the process. As delayed effects are not uncommon in interventions for autism, a post-hoc responder analysis was conducted five months after the first evaluation. Here, the participants who received high-intensity MT had significantly improved life quality, compared to the standard care group. Bieleninik et al. conclude that their findings do not support the use of improvisational MT for symptom reduction. However, they propose that symptom reduction is insignificant to autistic individuals, compared to well-being, engagement in learning, the ability to flexibly adapt to different situations and to build meaningful relationships, which improved during and after their study [9].

This section established techniques within MT, that prompt social and communication skills, as well as the impact that MT have on the lives of autistic individuals. For facilitating skill generalisation, IVR is suggested as an aiding tool for MT. The following section examines this technology.

2.3 Immersive Virtual Reality

When developing IVR solutions for autism interventions, it is essential to know what possibilities, strengths, and limitations the technology has. In IVR research, the terms 'immersion' and 'presence' are inevitable, as well as the often-discussed issue with IVR experiences, cybersickness. These three subjects are investigated in this section.

2.3.1 Immersion

The term 'immersion' originates from the physical feeling of being submerged in a medium [45]. An ongoing discussion in the IVR research community is whether immersion is prompted by technological properties or subjective experience. As a result, Nilsson et al. [45] propose a three-axis taxonomy of immersion including narrative immersion, challenge immersion, and system immersion. The first two involve the user's attentional surrender as a perceptual response to the mediated content, and the latter is viewed as an objectively measurable property of the system [45]. Following Slater's definition [70], immersion depends on what sensorimotor contingencies the system supports. This means that the more actions, which are utilised to perceive, that the system supports, the more immersive it is. Immersion is, therefore, prompted by high fidelity tracking and the quality of the displays that simulate the sensory modalities [45, 69].

2.3.2 Presence

With a system that supports sensorimotor contingencies approximated the physical reality, the user can experience the illusion of being located in the mediated world [70]. Slater termed this illusion the 'place illusion' or PI. PI covers the physics of the situation in the virtual environment (VE) and the quality of the actions that can be applied to perception. This means that a head-mounted display (HMD) that supports a 180° horizontal field of view supports higher quality actions than one supporting 10°. Another aspect of presence is the believability of the events occurring in the VE, which Slater calls the plausibility illusion or PsI. It relates to both the exteroceptive and interoceptive sensations, and to the often valid psychophysiological reactions that the VE can induce in the user. PsI can be elicited when, for example, non-player characters react and initiate contact with the player, without the player being the cause of the event [70].

PI and PsI merge in the representation of the body. Here, tracking enables a correlation between information from the proprioceptive and the exteroceptive visual system whenever the user moves within the VE. This can lead to a sense of body ownership of the virtual body. The theory of body ownership originates from the 'rubber hand illusion' paradigm, where a person is under the illusion of a rubber hand is theirs, while hiding their true hand under a table. When the rubber hand is touched, the person can 'feel' it. Body ownership is found to be elicited with the whole body in immersive VEs [70]. Presence and body ownership can lead to realistic responses in the VE [45, 70], which can further provide higher ecological validity [50].

It is important to acknowledge that both PI and Psi are perceptual illusions - not cognitive. Thus, while the user is experiencing Psi and PI, they also *know* that it is not real [70].

2.3.3 Cybersickness

One issue that remains with IVR is the risk of cybersickness. Here, the user can experience symptoms such as eye strain, headache, and nausea [18, 66]. The most prominent theory explaining the aetiology of cybersickness is the sensory conflict theory. It suggests that it is caused by a conflict between the visual system, the vestibular system, and nonvestibular proprioception [17]. The discrepancy occurs because the brain receives visual stimuli from the optical flow illusion perceived in the HMD but does not receive corresponding stimuli of movement from the vestibular system [37] Therefore, when implementing locomotion techniques for traversing the VE, it is recommended to use methods that instantly moves the user to the goal location, such as teleportation. If using continuous movement, developers should diminish the acceleration duration which is a known contributing factor to inducing cybersickness [54]. The risk of cybersickness can also be decreased by reducing lag and making sure that the HMD is calibrated suitably to the user [18, 66]. The ergonomics of the HMD are also a contributing factor, such as its size and weight [17].

The investigation of the limitations and possibilities for MT and IVR showed that the two components contribute differently to important aspects of social interaction; MT challenges the social flexibility and focuses on the interpersonal interactions, whereas IVR provides a space of focused multimodal stimuli, where the recipient can attend to relevant stimuli in the VE, without disturbances from the outside world.

3 Related Work

With the precautions that should be taken when developing IVR solutions (as mentioned in segment 2.3.3), it is also important to examine how to design for MTs, how to bring MT into IVR, as well as how autistic individuals can become engaged in IVR. These considerations are investigated by studying previous literature in this section.

3.1 Design Considerations for Technical Solutions in Music Therapy

In the study by Baltaxe-Admony et al. [5], six MTs from different countries were interviewed about their experiences and thoughts about incorporating technological devices into their practice. From the transcribed interviews, five themes emerged:

1. the device should be **versatile** as MTs often work with a diverse group of people,
2. the device should be easy to travel with, as many therapists conduct their sessions in the recipient's home, thus **form factor** was important,
3. the device should be easy to set up and **easy to use** for the therapist, which leads to,
4. the device should be **standalone**, as additional equipment is not always available (such as personal computers and monitors),
5. and the MTs should be able to **collect data** from the sessions.

The fifth theme was not agreed upon by all therapists, as one mentioned that data and evaluations could remove focus from providing a musical experience. Baltaxe-Admony et al. included it as it can be necessary for situations where the recipient's family needs the data for receiving financial support from their insurance companies.

Based on the design considerations, Baltaxe-Admony et al. developed a physical prototype. However, as the prototype was described as "*fairly bare-boned*" [5] (barely finished), it is not relevant for this project. Yet, the design considerations deduced from the interviews are of high importance for developing a suitable tool for MTs to use and accept.

3.2 Virtual Reality Musical Instruments

Serafin et al. [66] present a field overview for designing virtual reality musical instruments (VRMIs). They deduced nine design principles to follow. Some are consistent with the strengths and limitations that come with IVR technology, presented in section 2.3; a VRMI should elicit presence, represent the user's body suitably for the situation and instrument, and there should be a focus on avoiding cybersickness, reducing latency, and display ergonomics. In addition, Serafin et al. advocate for four more design principles:

- with IVR's ability to display multiple modalities, the design of **feedback and mapping** is especially relevant, as the motion and location of the visual virtual objects must match its audial equal. Also, appropriate tactile feedback is found to enhance the playability of VRMIs and can aid in learning to manage the instrument;
- **using existing skills** and extending the possibilities from existing instruments makes the VRMI more familiar and thus facilitates learning in the beginning;
- the VRMI should consider both **natural and 'magical' interactions**. Natural interactions comprise actions which are limited by physical constraints while magical interactions are not (e.g., the Go-Go interaction technique). Natural interaction has the advantage of familiarity and may increase usability,
- the last design principle is to make the experience **social**. Due to the HMD occluding visual communication from the physical world, the experience cannot be created with the usual gestures, i.e., pointing towards objects. Therefore, alternative methods should be used to facilitate social interaction. By providing a virtual social experience, social presence/co-presence can be elicited [66].

Some of the design principles tend to the novel user, which is relevant for this study. This includes exploiting the user's existing skills and providing tactile feedback. Further, the last principle of making the experience social is highly applicable for this study, where a multiplayer IVR tool is evaluated, exactly to prompt, i.e., joint attention gestures (pointing towards objects).

3.3 Engagement in Immersive Virtual Reality

Tarantino et al. [72] assessed the level of engagement that can be elicited in individuals on the autism spectrum by the extended reality (XR) technologies; IVR and augmented reality (AR). IVR was experienced with the Oculus Rift and AR with the Microsoft HoloLens. Because assessment methods for measuring engagement in IVR are based on research on N-AIs, Tarantino and colleagues resorted to observing the behaviour "*to prevent*

Contributor	**Year: 2021**	**Process focus**
Music Therapists	February	Preliminary interview: establishing the requirements for the VE.
	March-May	Development of the *first iteration.*
	May	Evaluation by one of the MTs.
Psychologists	September-December	Development of the *second iteration.*
	October-December	Four meetings providing feedback on the design.
Non-autistic children	December	A test focused on usability and initial attitude towards the application.
	Year: 2022	
Psychologists	February-April	Testing with the psychologists in IVR and *preparing for IVR sessions.*
	April-May	IVR sessions with the target group.
Autistic children	May-July	Continuation of the IVR sessions.

Table 1: *Timeline of the design process of the IVR tool. To the left is who was contributing to the project and to the right, is a recapitulation of where the focus was at the given time.*

potential errors and bias due to the difficulty that [autistic, ed.] people have in reflecting and reporting on their own behavior and emotions" [72]. The method for measuring engagement is interesting to investigate because of the different phenomenological perceptions autistic and N-AIs have.

The study was designed with the participation of three computer scientists, four psychologists, a medical doctor, and nine individuals on the autism spectrum (age 15-23). The study itself included six autistic males (age 21-23) who all were attending either high school or university courses, and with IQ scores ranging from 73-98. Each participant experienced four different scenarios, including photorealistic images and low-poly VEs, that at times invited interaction [72].

To measure the engagement of the participants, three observers were instructed to answer a Likert scale (1 to 5) that asked about different engagement factors. These were derived from studies about IVR and N-AIs and references from autism research. The factors were:

- **Emotional participation** in watching photorealistic and non-photorealistic scenarios. This is interesting as photorealistic environments are found to elicit a greater level of presence in N-AIs.
- **Suspension of belief**, where the virtual world is temporarily accepted as reality, which is desirable for IVR due to the place illusion, PI.
- **Body participation:** were there coherent body reactions to the mediated content? This is namely statistically significant correlating with the level of engagement.
- **Exploration** of the VE and ...
- ... exploration of **actions**, that both include voluntary participation which was not mandatorily required from the activity [72].

The observers were asked to describe the participants' reactions before, during, and after the experiences - specifically, their facial expressions, level of attention, emotional participation, and verbal reaction.

The responses from the Likert scale indicated a medium level of engagement, whereas the observational descriptions indicated high levels of engagement. Noticeably was the contrast to the usual findings of photorealistic scenes performing better in terms of engagement; here, the photorealistic scenario did not perform better than its low-poly alternative. Tarantino et al. propose that this complies with, that autistic individuals often get distracted by details of complex environments and thus have felt more relaxed in the simplified VEs. It should be mentioned that the cable on the Oculus Rift may have impacted the results regarding body participation and action [72].

Tarantino et al. thought that some of the VEs would be unpleasant to experience. However, when inquiring about them, the participants said that they could not be scared by a fictional world. That the VEs were perceived safe by the participants, is promising for utilising XR technologies in interventions and training. It was also found that familiarity influenced the level of engagement, hence personalisation of the IVR application should be incorporated [72]. Tarantino et al. further suggest that because IVR completely excludes visual inputs from the physical world, this could be used for beginners. With practice, the recipient may become desensitised to sensory inputs and outside disturbances and can move to AR technology that incorporates more of the physical world.

The findings from this section were highly important during the development of the IVR tool in the previous study [52], as well as when preparing the tool for the autistic children. The process of developing the tool is described in the following section.

4 The Interaction Design Process

As mentioned, this project is a continuation of a twelve-month design process. This section outlines the process and essential findings from the previous study [52]. Table 1 provides an overview of the process. The purpose of the previous study was to design a multiplayer IVR application for MT interventions for autism. The VE was initially designed in cooperation with two MTs and later with two psychologists. The cooperation changed due to the opportunity of evaluating the application's impact on the target group at the day hospital where the psychologists work.

The design process of the application followed a combination of the four approaches within interaction design: user-centred, activity-centred, systems, and genius design [62]. Summarily, in **user-centred design**, the designer focus on the user's needs and especially the user's goals, who are also often involved in the design process [55, 62]. **Activity-centred design** focus on the user's behaviour and the activities the users perform. **Systems design** is centred around the functionality of the system, such as the I/O-systems, controls, feedback, as well as the interaction with the device, other users, and the user self. Lastly, **genius design** is based on research and the designer's experience. Therefore, the genius design approach has little - if any - user testing [62]. At times, some techniques were more prominent than others during development, which is described in the following segments.

4.1 Designing a VE with Music Therapists

To establish the requirements of the MTs, a user-centred design approach was taken, commencing the design of the first iteration in February 2021. Based on the design considerations for developing digital solutions for MT (established in segment 3.1), it was decided to use the wireless HMD, Oculus Quest 2 [41]. This covered the considerations of form factor, being a standalone technology, and was relatively easy to use and set up.

The design of the first VRMI was based on the design principles established in segment 3.2, thus genius design approach was used. The mapping of audio and visual feedback was considered to create a suitable connection between the two and interaction was kept relatively simple and natural, meaning that the VRMI could be played by grabbing and placing objects on the VRMI, thus utilising existing skills. The social aspect of the VRMI was established through the online multiplayer feature. Because the tool needed an internet connection, the amount of data shared between the users was limited, for the sake of reducing latency.

With familiarity as an influencing factor for engagement (mentioned in segment 3.3), the VRMI was set in a familiar environment: a typical classroom, as seen in Figure 2. Because objects with reduced complexity are seen to induce relaxation without decreasing engagement [72], low-poly models were used to create the VE. Furthermore, with the tendency of looking at the piecemeal information instead of a context, it is proposed that the fidelity of the visual aspects in a VE is less important for autistic individuals, hence less important for generalising the skill in focus [48].

The avatar was also given a simple design. With one of the MTs expressing their worry about the avatar being intimidating if approaching a realistic design, and because autistic individuals are challenged with processing facial information (mentioned in 2.1.3), an iconic-looking avatar was designed. With the design of the avatar impacting the user's behaviour [66], a non-realistic avatar may invite less serious and more playful interaction, which is encouraged [7]. The avatar was later iterated, which is described in section 5.

After the evaluation by one of the MTs in May 2021, it was decided to re-design the VE and the VRMI, based on the provided feedback. In short, the VRMI was too complex for allowing versatility and did not show what contributions each user had made, which is important for prompting social interactions, i.e., joint attention [52]. This feature was later implemented, as will be presented in section 5. Also, since the majority of the objects in the classroom VE did not have a function other than being decor, it was mentioned that these could be distracting the recipient during the intervention and were removed. As it was no longer a classroom, the VE lacked familiarity. Therefore, to make the VRMI recognisable, it was re-designed as a model railway scenery, based on the knowledge about (autistic) children's affinity for trains [2, 64, 81], thus using the genius design approach. From here, the cooperation changed to working with the psychologists.

4.2 Iterating the Virtual Environment with Psychologists

Initiating the cooperation with the psychologists in September 2021 and developing the second iteration (see Figure 1), a mixture of user-centred design and systems design was used. During this cooperation, especially the interaction between the user and the VE and related feedback were prioritised, considering the perceptual ability and mobility skills of the target group, consisting of autistic children. Concerning autistic perception, the implemented sound effects (SFX) were chosen based on studies [7, 16] that found a preference for low-pitched flat SFX together with melody sounds. Natural SFX (i.e., wind, water, birds) induced negative emotions in autistic children [16].

When non-autistic children tested the IVR application in December 2021, it was possible to observe their actions and reactions within the VE. Thus action-centred design approach was employed and led to the addition of a 'tidy-up button' to reset the interactive objects' positions, as well as resolving technical issues such as the player being able to move outside the VE and climb the walls.

The process up to this point has been thoroughly described in [52]. From here, this study continues the process.

4.3 Preparing the VE for the Autistic Perception

In preparation for the evaluation of the IVR tool with autistic children, a combination of user-centred and

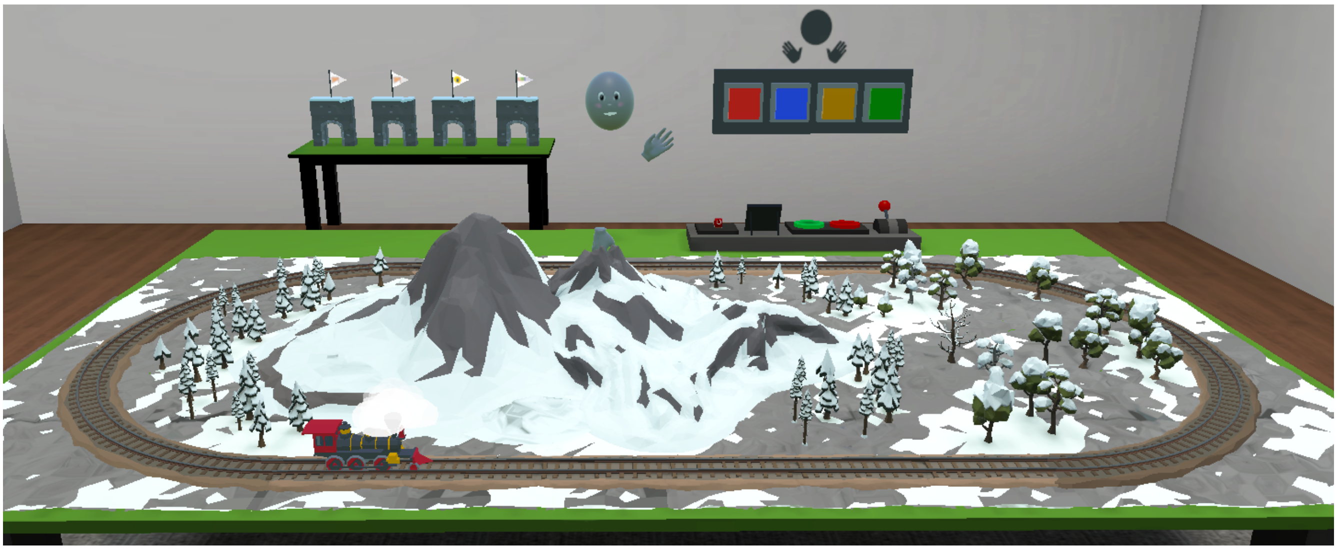


Fig. 1: *The second iteration of the IVR application. In the front: the VRMI as a model railway scenery. In the background: to the left, the gates that could be placed on the railway and emit audiovisual feedback; in the middle, the avatar; to the right, colour buttons, where the user could change the colour of their avatar.*

Fig. 2: *The first iteration of the IVR application. In the middle, is the VRMI in a VE designed as a typical classroom. Salient colours were given to the instrument to attract attention. To the right, is a xylophone which was omitted in the process.*

Fig. 3: *The version of the VE used for the evaluation. Different to the second iteration are the two buttons on the wall in the back: to the left, an exit button, and in the top middle, a tidy-up button. The railway scenery was also re-designed from a winter to a summer version.*

action-centred design approaches was used. Two informal meetings were conducted online within the developed VE with one of the psychologists and a colleague from the previous study (and the first author of [7], a study about the acceptability of XR technologies in autism interventions). The psychologist and colleague gave feedback while interacting in the VE. Based on the feedback, the VE was reduced to a relatively small size (about 3x4 meters). This was done to 1) fit the VE to the physical room wherein the interventions would take place, and 2) reduce the amount of movement the children needed to do, considering that some children have mobility issues. The IVR system allowed for real-walking (physical walking) and as an alternative, continuous locomotion, which was initiated with the right thumbstick, and snap rotations (45°) with the left. Here, the psychologist wanted teleportation instead of continuous locomotion, as this could be performed with just one controller, which is what come children can handle. He further suggested that teleportation would be easier for the children to use, as well as reduce the risk of cybersickness, stating that the children would most likely not voice if nauseated, and could get sick without notice. The psychologists further asked for implementing haptic feedback as the children have a liking for vibrations. Also, haptic feedback was advised for aiding the user to learn how to play the VRMI (mentioned in segment 3.2).

With the final revisions, the psychologists could begin the IVR interventions. The design and implementation of the VE used for the evaluation are presented in

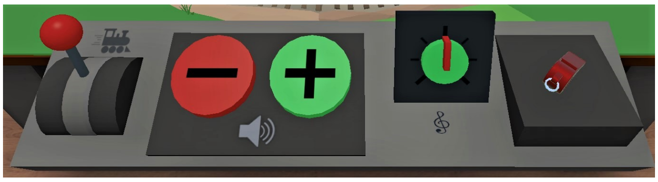
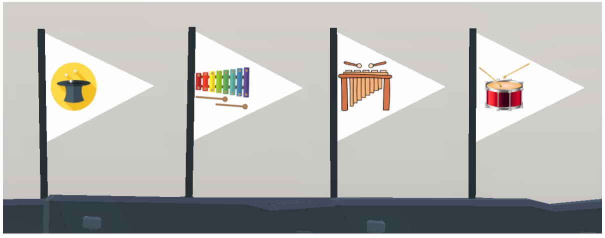
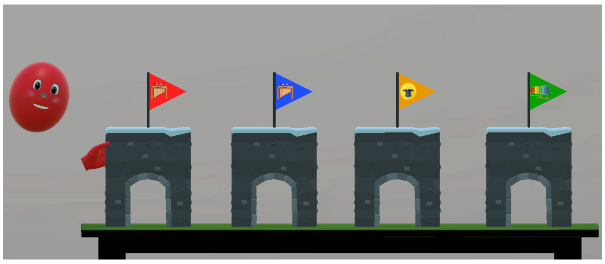


Fig. 4: *The control panel where the user controls the sound. From left: the tempo lever, the volume buttons, the pitch rotator, and the whistle.*



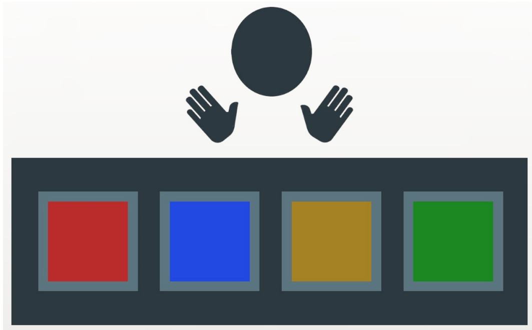
(a) The icons representing the SFX of the gates.



(b) The flags change colour to the interacting user's avatar colour.

Fig. 5: *The gates that are used for creating music with the VRMI.*

(a) To the left is the red exit button and to the right is the green tidy-up button.



(b) The colour buttons used to change the colour of the user's avatar.

Fig. 6: *The buttons on the walls of the VE.*

the following section.

5 The Virtual Environment

Figure 3 shows the version of the VE which was used for the evaluation. The design and implementation of the VE are described in the following segments.

5.1 Design

As mentioned in segment 4.1, the VRMI is designed as a model railway environment, placed on a table in the middle of the VE. The VRMI comprises a locomotive circulating the environment, a lake, trees, flowers, and a sun slowly rotating in the air. On another table, twelve gate models are placed, which are used to create the musical sound. This is done by placing a gate across the railway, and when the locomotive drives through it, the gate emits audiovisual feedback.

The user controls the properties of the gates' SFX by interacting with a control panel placed on the side of the railway environment (see Figure 4). This includes 1) a lever to control the tempo of the locomotive, 2) two buttons to increase and decrease the volume, and 3) a rotator knob that increased and decreased the pitch. Four SFX are included; a marimba (G5 note), a xylophone (C6 note), a bass drum, and a 'magic plings' effect, that are visualised with an icon on the flags of the gates, as shown in Figure 5a.

Next to the control panel, the user finds a whistle, which when taken near the user's head (for mimicking blowing in a real whistle), the whistle makes whistle SFX. As a response, the locomotive emits a horn sound and changes the colour of the smoke from its chimney (see Figure 8). The smoke is further used to visualise the current state of the settings of the control panel. This is outlined in Table 2.

On the walls of the VE, a button row with different coloured buttons (see Figure 6b). These can be pressed to change the colour of the user's avatar and when the user grabs a gate, the colour of the gate's flag and visual feedback will change to the current colour of the avatar, as seen in Figure 5b. This is done to visualise the users' contributions. The gate's visual feedback is a particle system that bursts silhouettes of musical notes, which is activated when the locomotive drives through it.

On another wall in the VE, is a red exit button to quit the application, and a green tidy-up button to reset the location of the gates and the whistle for all users (see Figure 6a).

Interactions with objects are direct (natural) and are initiated by the user pressing one of the grip buttons on the controllers. Haptic feedback (vibrations) is provided by the controllers when the user interacts with the control panel, and when pressing any button in the VE (buttons in Figure 6). The user will also feel vibrations when hovering their virtual hands over objects such as the locomotive, the sun above the railway environment, and other users' avatar components.

The avatar comprises two hands, a torso, and a spherical face with a 2D drawing as a face (see Figure 7). The mouth changes height depending on the microphone

Controller	Purpose	Other feedback
Lever	Locomotive drives faster/slower.	–
+/− buttons	Volume of the gate sounds is decreased/increased.	Transparency of the smoke is decreased/increased.
Rotator	Pitch of the gate sounds is decreased/increased.	Size of the smoke is increased/decreased.
Whistle	Whistles and the locomotive makes a horn sound.	Smoke colour changes to one of the avatar colours.

Table 2: *The controllers on the panel, their functions and related feedback [52].*

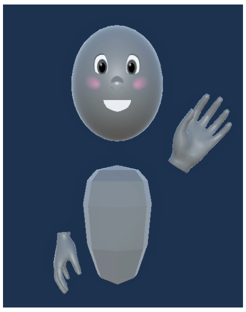


Fig. 7: *The default avatar. When changing avatar colour, both hands, torso, and the spherical head change colour. The mouth changes height depending on microphone input.*

Fig. 8: *The particle system used for the smoke from the locomotive's chimney. The colours of the smoke can change between the avatar colours (more about this in the text).*

input volume from the user.

Besides real-walk locomotion, the user can teleport in the VE. This is initiated by pressing the 'X' or 'A' buttons, or by tilting either of the two thumbsticks on the controllers. The different possibilities were provided for the sake of versatility.

5.2 Implementation

The application is built for the Oculus Quest 2 [41]. For implementation, Unity version 2020.3.18f1 [74] is utilised with *XR Interaction Toolkit* v1.0.0-pre.8 for enabling XR interactions [75], and *Normcore* v2.1.10 as the networking framework [47].

6 Methodology

The aim of this study is to evaluate if the developed IVR application, designed with MTs and psychologists, can facilitate social skills training and subsequently be deemed appropriate for autistic perception and useful for autism interventions. The evaluation is based on IVR sessions with autistic children, conducted at the day hospital where the two participating psychologists work. In the evaluation, the IVR session is compared to other non-IVR sessions, that are offered on-site. A questionnaire was formulated to inquire about the psychologists' opinions about the child's behaviour. This was done, as self-assessment questionnaires are not recommended (mentioned in segment 3.3) and further were not possible due to many of the children being nonverbal. The formulation of the questionnaire is outlined in segment 6.1.

A mixed-methods approach was utilised to avoid the bias of subjectivity, as recommended in [11]. Here, data about what object the participant was gazing at was print to a file. It should be mentioned, that the utilised HMD does not support eye-tracking, thus the forward direction of the gaze was used as a representation. The gaze direction can indicate if the child directed their attention towards the other person (the other avatar), which may signify that the child can identify relevant aspects of the social situation [35, 53], which otherwise is difficult for autistic individuals, following the enactive mind theory (presented in segment 2.1.1). This information may prove beneficial when correlated with the aforementioned questionnaire.

After the non-IVR and IVR sessions were done, a semi-structured interview was conducted with the psychologists to follow up on the sessions and inquire into responses in the questionnaire that may be difficult to interpret. In addition, this interview was also to ask about their general opinion about the IVR tool, whether the design was appropriate to the autistic perception, and about using the IVR tool to train social skills with the children. The points that framed the interview is attached in Appendix B.

6.1 Formulating the Questionnaire

Because the VE is developed with MT in mind, assessment methods within the field of MT were investigated for evaluating the impact of the developed IVR tool.

Index	Question	Reference
A(-)	The child displayed stereotyped movements.	Psychologists
B	The child displayed body awareness.	Psychologists
C	The child could maintain focus on the activity.	[13]
D	The child was engaged in the activities.	[13, 72]
E	The child displayed an understanding of the cause-and-effect relationship in the activity.	[13]
F(-)	The child appeared overstimulated.	[13]
G(-)	The child displayed stress or anxiety during the session.	[13] and Psychologists
H	The child had the ability to react with nonverbal communication.	[13, 72]
I	The child was verbalising pertinent (fitting) to the interaction.	[58, 72]

Table 3: *The questions about the child's behaviour and participation during a session. To the right, is the references of the MT assessment method that inspired the formulation. Some questions were added and revised by the psychologists. Question A, F, and G, are 'negatively loaded' (more about this in the text).*

ID	Age	Sex	(ICD-10) Condition	CARS	KABC	Autistic Behaviour	Knowledge of Digital Tools
2	12	M	(F84.0) Childhood autism	47	NVI:49	Hand flapping, tiptoeing, echolalia, hyperventilating	Little
3	11	M	(F84.0) Childhood autism	42.5	MPI:53-68	None	Little
4	12	F	(F84.3) Other childhood disintegrative disorder	39	NVI:35-49	Looking at herself, very tactile	Medium
10	10	M	(F84.0) Childhood autism	41.5	MPI:35-47	Echolalia	Strong
12	10	M	(F84.3) Other childhood disintegrative disorder	41.5	NVI:47-61	Echolalia, sorting things	Little
13	10	M	(F84.1) Atypical autism	32	NVI:72-86	Echolalia in some contexts (i.e., many children around)	Strong
19	10	M	(F84.0) Childhood autism	-	-	Echolalia, clastic crisis (aggressive and destructive behaviours)	Little
20	7	M	(F84.1) Atypical autism	-	MPI:43	None	Little
21	10	M	(F84.1) Atypical autism	-	-	None	Little

Table 4: *ICD-10 is the exact diagnosis [80] given to the children. CARS (Childhood Autism Rating Scale) is a scale indicating the severity of the symptoms [73]. KABC (Kaufman Assessment Battery for Children) is a cognitive profiling for the children's intellectual abilities [51].*

There are several established MT assessment methods [13, 46, 58, 65, 78, 79], however many of these focus on diagnosing, which is not the purpose of this study - or require that the assessment conductor is educated in using the methods. With the change in cooperation from MTs to psychologists (mentioned in segment 4), the MT assessment methods could not be directly used. Instead, a questionnaire was formulated, taking inspiration from the sections of the examined MT assessments dealing with social interaction. Table 3 shows the questions included in the evaluation. The questions were rated 0='not at all', 1='rarely', 2='occasionally', 3='about half of the session', 4='often, but not always', and 5='consistently'. Questions A, F, and G were 'negatively loaded' questions, thus when calculating the total session score for the children, these questions were subtracted from the highest value, 5.

From the assessment profile developed by Carpente [13], additional focus was given to the support that the child needed during interventions. The level of support can create a framework for understanding the recipient's level of independence and how well they understand interpersonal cues, that are provided during an intervention. Support should be given, beginning with the lowest level and be increased when the therapist finds it necessary. The following questions were inspired by Carpente's method; they inquired about the child's need for support to *move around in the environment*, *maintain focus on the activity*, and *interact with the instrument*. These were rated on a scale from 0 to 4, where 0 ='no support', 1='verbal support', 2='visual support' (such as pointing or showcasing how to interact), 3='partial physical support', and 4='full physical support', as in the original method [13].

Specifically for the IVR sessions, an additional section in the questionnaire asked about the child's acceptance of wearing the HMD, if they were annoyed when removing the HMD (negatively loaded question), and whether the system seemed easy to use for them. The questions were inspired by the work done by the colleague in [7] and the System Usability Scale (SUS) [12], and were rated 0='strongly disagree', 1='disagree', 2='neutral', 3='agree', 4='strongly agree', as in the SUS.

Before initiating the IVR interventions, the questionnaire was revised with the psychologists and the colleague (author of [7]). Here, questions were added about the child's mood before and after the intervention, as this could have an impact on their behaviour during the sessions. They asked whether the child was *tired*, *anxious* (both negatively loaded questions), and *happy*, and were rated from 0='not at all' to 3='very'. Because the questionnaire took more than ten minutes to complete, it was shortened to avoid fatigue and to accommodate the daily schedule of the psychologists. This resulted in two times three mood questions, the nine session questions, three support questions, and for IVR sessions, three acceptance questions. It was a total of 18 questions (21 for the IVR sessions) and took about six minutes to complete. The questionnaire is attached in Appendix E.

It was hypothesised that the session scores were higher for the IVR sessions than the non-IVR sessions and data assumed parametric for statistical testing for paired sample t-test.

6.2 Participants

With the aim of the study being primarily about the autistic child's possibility of developing social skills facilitated by the IVR tool, there is a second target group: the therapists, as they also are primary users of the IVR

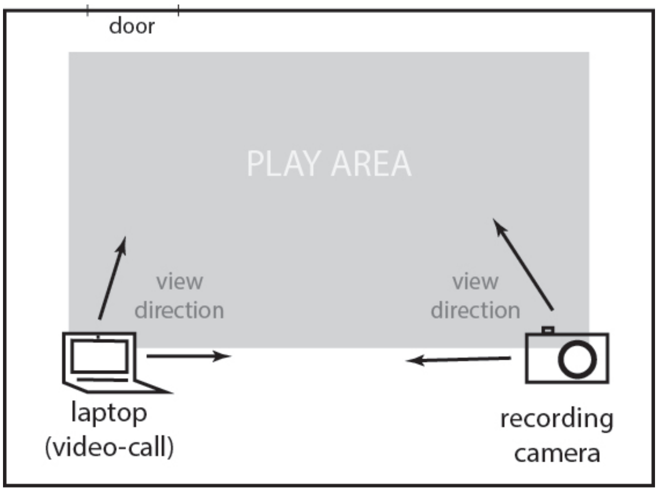


Fig. 9: *Simplified diagram of the room setup, seen from above. Dimensions are not true to the physical room.*

tool. The target groups are outlined here.

6.2.1 The Therapists

The two psychologists who participated in designing the VE and conducted the IVR sessions are autism experts. They use digital tools in their interventions daily, i.e., video games, tablets, and robot mediation. They have previously participated in a study about the use of XR technologies for autism interventions together with the aforementioned colleague [7]. Thus, the connection with the psychologists was established through my professional network.

6.2.2 The Children

The nine participating children were between 7 and 13 years old and were all on the autism spectrum. All children had an IQ under 70, meaning that they have an intellectual disability. Table 4 provides an overview of each child. Child 3, 4, and 13 have little to no spoken language. It should be mentioned that the children, except child 19, 20, and 21, participated in the colleague's study about AR technology in autism interventions [7] and, therefore, have experience with XR technologies.

6.3 Procedure

The non-IVR sessions included theatre mediation, video games therapy, and robot therapy. These were conducted as usual and only observed by the psychologists.

For the IVR sessions, the psychologists were instructed to make the sessions as natural as possible, meaning that if the child did not want to wear the HMD, they should not be forced to do so. However, data from such sessions were not used in the evaluation.

The IVR sessions were performed in a room about 5x4 meters. Each session lasted 15 minutes, and 10 of those were in IVR. For the IVR session, the psychologists, the colleague, and I were observers. I participated in an online video call (from Denmark) with the therapists and viewed the session from the view of the psychologist's laptop, placed in the corner of the room, as seen in Figure 9. The psychologists and the colleague

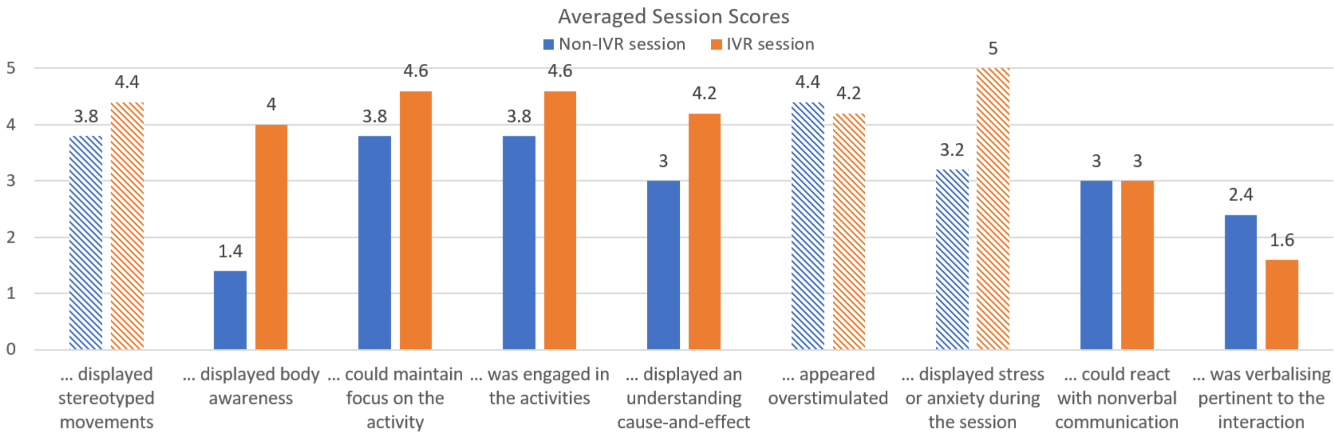


Fig. 10: *The average scores for each question asking about the observations of the children's abilities during the sessions.*

were on-site (in France). In addition, a camera was set in the corner of the room to record the physical view of the sessions. The list with instructions to guide the psychologists through the technical aspects of the session is attached as Appendix C.

The view from the HMD was cast to the laptop so that the child could see the view of the VE before entering IVR. This was done to make the HMD less intimidating and to ease anxious children. The psychologists also recorded the view from the child's HMD.

When the child entered the room, a usual IVR session was as follows:

1. The child was presented to all in the room by the psychologists.
2. One psychologist would prepare the two HMDs for the session, including casting the HMD view to the laptop.
3. The child was shown the VE on the monitor, whereto the other psychologist explained the plan for the session.
4. The child was presented to the HMD and, if accepted to wear it, the HMD was adjusted to fit the child, and the session would begin.

After a day of IVR sessions, a de-brief meeting was held. This was to discuss noticeable behaviours of the children and thoughts about the VE. The findings from the sessions are presented in following section.

7 Results

This section begins by presenting the analysed data from the questionnaire. Hereafter, a presentation of the debriefing interviews focusing on the children's behaviours and the performance of the IVR tool. The section ends by outlining findings from the concluding interview, where one of the psychologists was asked to elaborate on some of the answers from the questionnaire, as well as noticeable behaviours observed during the sessions and in the recordings.

Four of the children were omitted from the data, as they were prevented from participating in some of the sessions. One child did not accept to wear the HMD and was also excluded. Therefore, data from five children could be used for the statistical analysis (child 3, 4, 10, 12, and 13). The sessions were conducted on different days, which is outlined in Table 5. The colleague attended the sessions on the 27th of April and the 11th of May, and I attended by video call on the 11th and 18th of May.

ID	Non-IVR Session	IVR Session	Days in-between
3	08 Apr (Theatre mediation)	27 Apr	19
4	08 Apr (Theatre mediation)	18 May	40
10	06 Apr (Video game therapy)	11 May*	35
12	07 Apr (Robot therapy)	18 May*	41
13	06 Apr (Video game therapy)	11 May	35

Table 5: *The dates of the non-VR session that the psychologists observed, the date of the IVR session, as well as the number of days in-between. All dates are in 2022. (*)An educator was with the child in IVR.*

The data concerning the gaze direction was omitted because the HMD was frequently taken on and off during a session. This was either because the child wanted to see if people were still in the room or because of internet connection issues. With disconnection, the application needed to be restarted to again connect to the online VE. Therefore, the data would not have provided valid information about the direction of the gaze. Furthermore, it had not been made clear to the psychologists that one of

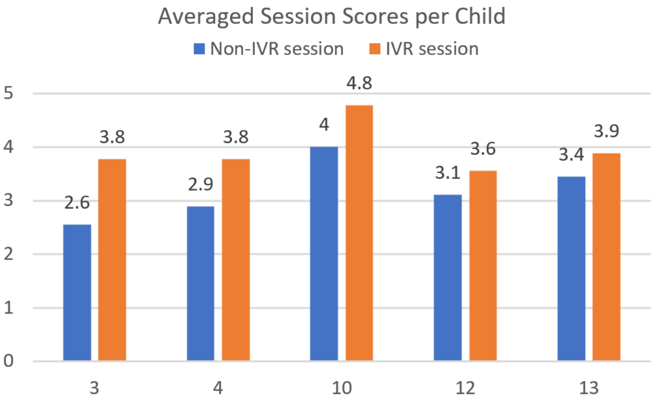


Fig. 11: *The average session score per child.*

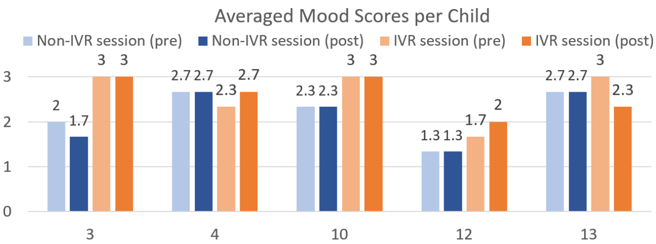


Fig. 12: *The average mood score per child before and after non-IVR and IVR sessions.*

the two HMDs should have been dedicated to the child, for finding tendencies of the child's gaze direction. The recordings from the children's view in IVR are attached in Appendix D.

7.1 Data from the Questionnaire

The results are presented in the categories; session scores, support level, usability and acceptance, and mood scores. The histograms show the contributions of each score, meaning that negatively loaded questions have been subtracted to the highest score possible in its related scale. Negatively loaded questions are visualised as striped columns in the histograms in Figures 10 and 15. The data from the questionnaire is attached in Appendix F.

7.1.1 Session Scores

The averaged scores for the individual session questions are showcased in Figure 10 and each child's score can be seen in Figure 11. The difference between the two data sets (non-IVR and IVR) was evaluated to be significantly non-normal with the Kolmogorov-Smirnov test ($D(10) = 0.35$, $p < .05$), hence parametric testing cannot be performed. Instead, a post-hoc Wilcoxon signed-rank test was utilised two compare the two conditions [22]. It was found that the session scores were significantly higher for the IVR session ($Mdn = 5$) than the non-IVR sessions ($Mdn = 4$), $z = -2.98$, $p < 0.05$, $r = -.94$.

7.1.2 The Mood of the Child

Looking at the change in the child's mood before and after a session, no noticeable difference was seen but a difference in the non-IVR session, which decreased by 0.1 ($mean_{pre} = 2.2$; $mean_{post} = 2.1$). When examining the individual child, there was a trend of the child scoring lower before a non-IVR session than in an IVR session, as seen in Figure 12. Therefore, it was decided to make a post-hoc Spearman's correlation between pre-mood scores and session scores. However, the test showed no significant correlation between the scores ($p > 0.05$).

For child 3 and 10, there were additional comments about the children asking to come to the session, when hearing that they were doing an IVR session.

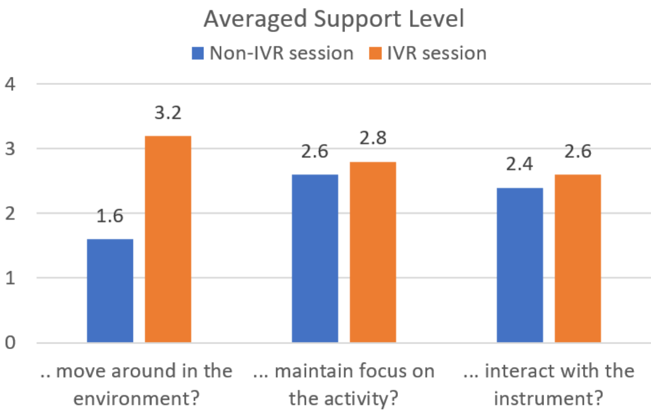


Fig. 13: *The average score for the level of support, the children needed during the session. The higher the score, the higher level of support the child needed.*

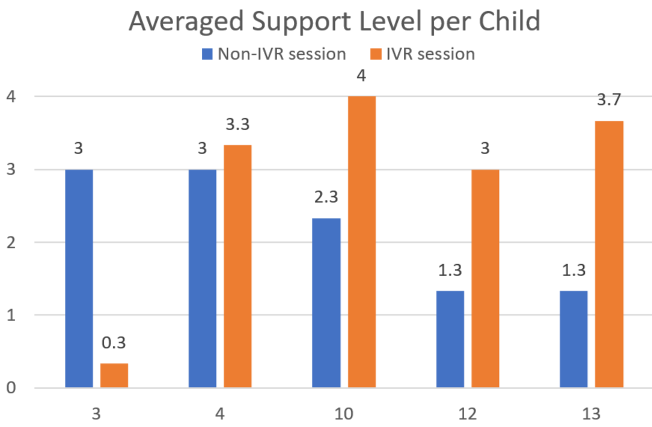
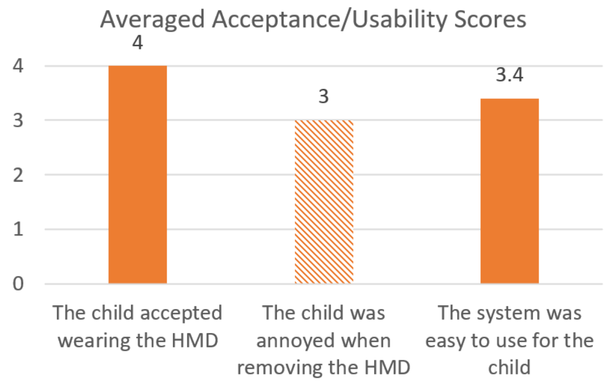


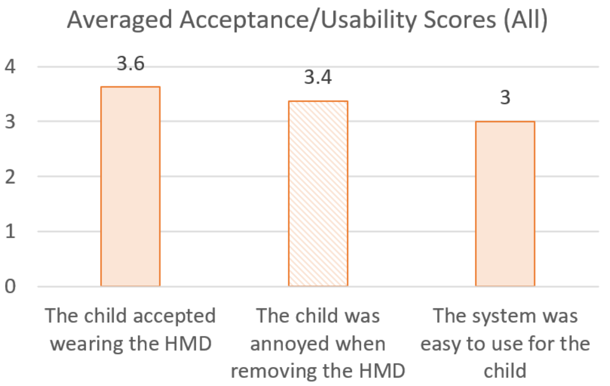
Fig. 14: *The average support level score per child.*

7.1.3 Level of Support

All but one child needed a higher level of support for the IVR session compared to the non-IVR session. The greatest difference between the support scores for non-IVR and IVR sessions was the score for moving around the environment (diff.= 1.6), as seen in Figure 13. The level of support that the children needed in the sessions varied a great deal, looking at the individual scores. Here, the smallest difference between non-IVR and IVR sessions was 0.3 for child 4, and the biggest difference was 2.7 for child 3, who was the child needing less support during the IVR session, as depicted in Figure 14.



(a) Data from the five children that was a part of the comparison between non-IVR and IVR sessions.



(b) Data from all children trying the IVR tool.

Fig. 15: *The average acceptance and usability scores.*

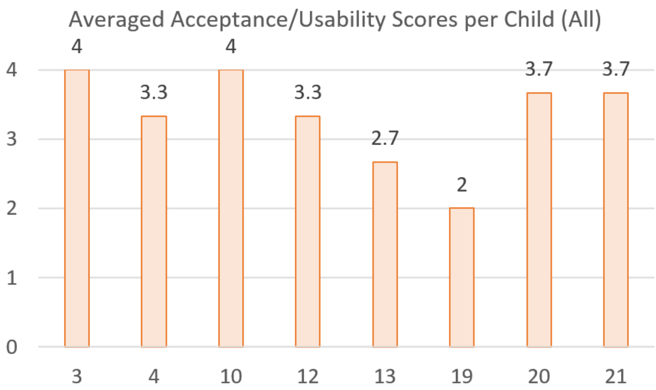


Fig. 16: *The averaged acceptance and usability score for all the children trying the IVR tool.*

7.1.4 Acceptance and Usability

Data from some children were omitted because of a missing non-IVR session, yet did test the IVR tool and, therefore, have data about acceptance and usability. The data from the five children included in the comparative test is visualised in Figure 15a, and for all the children who tested the IVR, data is visualised in Figures 15b and 16. This included all but child 2, who was prevented from participating in an IVR session.

The question of accepting to wear the HMD scored the highest possible (score of 4) for the five children included in the statistical analysis. When including all children who tried the IVR tool, the average scores decreased to 3.6 due to one child (19) who did not accept the HMD and instead participated in the session by looking at the view of the VE, displayed on the laptop.

7.1.5 Observations from the Sessions

This segment will sum up the comments that were written by the psychologists, and the observations made during the sessions. Some comments were written in French and thus were translated to English. The comments are also found in Appendix E together with the data from the questionnaire. Because some of the children participated in the AR study by the colleague [7], there are made some comparisons between the two technologies.

Child 3 seemed curious when told that he was going to try IVR. During the session, he was very active while wearing the HMD. The support he needed (scoring 0.3) was to prevent him from running into a wall. "*He was still very much 'in his bubble' but with quite a bit of sharing*", whereto he mostly used echolalia to express himself.

Child 4 has no spoken language. While wearing the HMD, she did not move more than a meter. She waved 'hello' to the other avatar in the VE and after a lot of trials, she managed to interact with the gates, and catch the whistle, that the psychologist threw to her. "*[She, ed.] makes throat noises, which is usually a sign of investment, relationship, and pleasure*". Typically, when she does not want to participate in an activity, she walks away. Here, she kept the HMD for a longer time before taking it off by herself, compared to the AR study. In terms of support (scored 3.3 of 4), it is suggested that she needs someone beside her during the session to help her interact.

Child 10 had his educator with him. He is usually lonely the first few sessions of a new activity, yet here, he interacted a lot with both the avatar in the VE and the people in the room. He also seemed to understand that people in the room could see on the laptop, what he was doing in the VE. He used the whistle a lot and also blew as if he was using a real whistle. The psychologists wrote that he "*seems to love this activity*" and that "*he is very interested in auditory stimuli*". He also participated in the AR study, which he also enjoyed. At the end of the session, he took off the HMD and "*he showed great surprise when he noticed that [his educator's, ed.] avatar was far away from him but that his [educator, ed.] was near him and that he could [physically, ed.] touch her. He [...] showed a little concern, and was reassured to see her and to understand the difference between the real distance and the virtual distance*".

Child 12 can usually get anxious, and had his educator with him. For the AR study, he "*refused to wear the headset. This time, he easily agrees to wear it*". As he does in real life, he tidied up the room after the activities he had participated in.

Child 13 does not like to be in groups. In IVR, the child interacted a lot with the psychologist. The first time he got the whistle to make a sound, he seemed surprised, but then to understand how to interact with the virtual objects. "*[He was, ed.] so enthusiastic and had to test everything. [He was, ed.] not overstimulated but it lead to a rush*", and was "*laughing a lot during the* *session*". He did not interact with the VRMI as intended but instead tried to test the system's limitations (he tried to 'break' the system) by running through the walls in the VE. It was notable, that the child tried to hit the psychologist's avatar during the session by swinging his arms and doing boxing movements. He also did the AR study and liked it very much.

These five children were included in the statistical analyses. From here, the children did not have data from a non-IVR session and thus were excluded from the comparison.

Child 19 is usually anxious when introduced to new activities; "*he always needs a little time to discover a new medium*". Here, he did not accept to wear the HMD, yet did look inside of it. Instead, the psychologist entered IVR, which the child then could follow by looking at the laptop. As he was looking back and forth between the psychologist in IVR and the laptop, it seemed he made a connection between the two. The psychologists wrote that he has "*very little verbal language, [and uses, ed.] echolalia that are not necessarily adapted to the situation*".

Child 20 was as usually "*very cheerful*". In the VE, he was "*very present in his body, but moves very little, and observes a lot around him*". He seemed to understand what the psychologist showed him in the VE. The psychologists wrote that he was "*not too stimulated. Properly*". After the session, he was "*very happy and said so*".

Lastly, **child 21** was "*rarely happy currently*" and had his educator attending the session. For this session, there was an issue with the height of the avatar, which made the child and psychologist perceive that they were much taller than their physical height. The child was challenged in using teleportation and needed support to move in the space (had an average support score of 3). During the session, the child "*said a few times 'it's like real life'*", and after the session, the child verbalised that "*'I did video games!!' to his educator*", and said that he "*really liked it*".

7.1.6 Debriefing Interviews

After the sessions, the debriefing interviews were held. Here, the psychologists gave additional feedback about conducting the IVR sessions and from some of the educators of the day hospital.

It was observed that haptic feedback had helped some children understand what should be perceived as salient in the VE, such as the VRMI and the components of the other user's avatar. Furthermore, the children had a general interest for the whistle, where they could make a sound from their 'own' body.

The psychologists mentioned that IVR sessions are more difficult to handle than AR sessions because they cannot see what the child sees. Therefore, they recommend that two therapists are needed for the sessions, and during these evaluations where there is more equipment to handle (casting to a computer, recording with a camera, video-call, writing observations, etc.), at least three people are needed.

From speaking with the educators, the psychologists mentioned that the children had talked about their experiences on the playground. This resulted in multiple educators wanting to include other children in the IVR sessions, which they did not after the AR sessions.

7.2 Concluding Interview with a Psychologist

After examining the video recordings and data from the questionnaire, a concluding interview was conducted with one of the psychologists. He was inquired about some of the written responses and prominent behaviours of the children. The interview lasted 15-minutes via video call.

First, the question about one of the children, who had been boxing the psychologist's avatar. This behaviour was explained as being a positive behaviour, as this is the child's way of initiating interaction. However, as this behaviour is not condoned by society, it is also not accepted at the day hospital. The psychologist mentioned that it was, therefore, good that the child can express themselves freely in the VE without the consequences of real life. In this way, they can work on the behaviour instead of immediately shutting it down, which can be frustrating for the child, as they then find it difficult to make contact.

Generally, the psychologist had noticed that the children would interact more with them (the psychologists) and other people in the room, compared to non-IVR sessions. Also, compared to AR, the children interacted more with others. He suggested that because the child was wearing an HMD, visual and partly auditory stimuli from the outside were limited, resulting in fewer disturbances and that the child could focus more on the social interaction. He adds that it was important to see, that many children either took off the HMD or called the psychologists, to check if they were still in the room. With more practice, he would like to see the children reach a level of comfort where they do not feel the need to do so. He mentions that for future iterations, it would be interesting if the application could change between IVR and AR to progressively expose the child to more complex stimuli. The most important was that it seemed as if the children could do this kind of session for multiple weeks.

In terms of the stimuli that the VE provided, there was an appropriate amount of stimulation. This was specifically seen for child 13, where the psychologists had written about a 'stimulation rush', which was positive feedback. He mentioned that the auditory stimuli were an important aspect of the IVR tool. He asked for an additional feature, where a sound would mark the end of the sessions, which would help the children understand this better. Now, many of the children did not want to leave the session.

Concerning the features of the IVR tool, the psychologist had noticed that no child interacted with the pitch rotator and that all children wanted to grab the locomotive (which was not possible). He proposed tutorials to introducing the children to the different concepts of the IVR tool. He expressed his vexation about the internet connection issues and that the objects were difficult to grab, if multiple users interacted with the same object.

The psychologist emphasised that it is important to acknowledge that the VE is a tool for the therapist and that the therapist, therefore, should know how to utilise the tool properly. The interview concluded with him mentioning that the IVR sessions will continue at the day hospital to see how the children develop the following month.

7.3 Key Findings

The key findings from this section are outlined here:

- Five autistic children were included in the statistical analyses.
- A significant difference was found between the session scores for the non-IVR sessions and the IVR session.
- The mood score did not significantly correlate with the session scores.
- The children needed higher levels of support, especially to move around in the environment.
- All children but one accepted to wear the HMD.
- The psychologists observed that the children
  - children were excited to try IVR,
  - interacted with others more than usual,
  - and could distinguish between the real and virtual world.
- The VE had included an appropriate amount of stimuli.

The following section will discuss the findings as well as the IVR tool's ability to facilitate social skills.

8 Discussion

This section includes a discussion of the results presented in section 7, as well as a general discussion about the design of the VE and the impact, that the IVR tool may have on the children's social skills. Lastly, a discussion about the implementation of the IVR tool in the therapeutic practice.

8.1 Discussion of Results

As it was hypothesised, the children scored significantly higher for the IVR sessions compared to the non-IVR sessions. With a small sample size, it can be a coincidence that the participating children did particularly well, compared to how the target population would actually perform. However, with the general findings of the acceptance and impact of IVR tools within the field of autism research [72], this is implausible.

The children displayed notably higher levels in, especially, body awareness, understanding of the cause-and-effect relationship, and lower levels of stress and anxiety. Body awareness is essential for self-regulation and relates to both external sensory stimuli and interoception, which pertains to emotions [32]. Therefore, the noticeable higher level of body awareness can be a result of the focused stimuli that IVR technology provides, hence fewer disturbances, which then leads to lower levels of stress and anxiety. This is in compliance with the children scoring higher in the question about their ability to focus on the activity, compared to the non-IVR session. Additionally, understanding a cause-and-effect relationship and reacting pertinent to a situation means that the children had high levels of engagement (as mentioned in 3.3), which is important in the context of learning [63].

The two questions, where the IVR session scored lower than the non-IVR session, were about pertinent verbal communication and about the child appearing overstimulated. A connection between the two can be made, as echolalia is usually associated with sensory overload [24]. Albeit, child 13, who had a score than indicated a considerably level of overstimulation during the IVR session, was also the child who had experienced a 'stimulation rush'. The psychologists later explained this as positive and that the level of stimuli had been appropriate to challenge the child, thus suggesting that a low score did not necessarily mean that the children had reacted negatively. This was further supported by the repeated feedback about the appropriate levels of stimulation in the VE.

It was suggested that the mood of the child previous to the session could be an influencing factor for the outcome (segment 6.1). Yet, no significant correlation was found between the mood score and session score. It is suggested that the mood scores were lower before non-IVR sessions, because these were conducted in groups, and that the children prefer individual sessions with the psychologists. It was experienced that a child was stressed or anxious due to a negative interaction with another child from the day hospital.

The level of support is said to indicate the level of independence and how well the child understands the interpersonal cues (mentioned in 6.1). However, as the IVR tool was a new medium for the children, and with the removed visual contact with the psychologists, it is not surprising to see a higher score in support. Thus, it is not seen as a decreased level of independence, when looking at how well the children did in terms of social interaction. This idea is supported by data, where the child who needed the most support (scoring 2.7) and the child needing the least support (scoring 0.3), both scored 3.8 of 5 in the session scores.

Concerning the acceptance and usability scores from the questionnaire, the psychologists did rate the children to be annoyed when removing the HMD at the end of a session (the average score was 3.4 out of 4). In segment 7.2, the psychologist mentioned that adding a sound to mark the end of the session, would lead to minimal annoyance by the children, as they would recognise its meaning. He suggested a 'gong' sound, which the children were familiar with from the colleague's work [7]).

The one child, who did not test the IVR tool was mentioned to always be cautious about new kinds of mediums. The child displayed curiosity by looking inside the HMD and participating by looking at the IVR view that was cast onto the laptop. The child did have little experience with digital tools, which can explain the anxiety. Albeit, not every child with the same level of experience had the same precautions. Hence, this is seen as an exception.

From the video recordings, it was observed that the whistle was a preferred object to interact with. Looking at the differences between the whistle and the VRMI, these can be categorised as 'active' and 'passive' VRMIs, respectively. The railway VRMI is passive because it does not require constant and repeated interaction for making sound, whereas, with the whistle, the user must make the act of moving the whistle near their head. It is, therefore, suggested that additional 'active' instruments should be included in the VE. This is in compliance with the findings from the previous study [52], where one of the MTs acknowledged the importance of including an instrument that a child could immediately recognise - such as the xylophone in Figure 2.

It should be mentioned, that with the frequent internet connection issues, the IVR tool's full potential may not have been exploited. The connectivity issues resulted in undesired disturbances and positions of objects not updating for all users, who then could not interacted with said objects. Therefore, for this multiplayer application, a stable internet connection is a requirement.

The next segment will discuss the methodology from which these outcomes emerged.

8.2 Discussion of Methodology

With the obstructed plan of utilising mixed methods, this evaluation relies on the subjective answers from the two psychologists. Whether a bias of novelty or a conflict of interest was present cannot be assured. Future studies are advised to incorporate eye-tracking for data triangulation, as well as acquiring more precise data. Eye-tracking was not supported by the HMD used for this study.

With eye-tracking, it would be interesting to investigate where the autistic children would direct their attention. This could be compared to non-autistic children, to see if they could equally skilful recognise social cues made by the therapist during a session. A study about attention capture is conducted by Scheerer et al. [64], where they found that trains captured the children's attention more strongly than face stimuli for both non-autistic and autistic children. Because the participants were 7-10 years, they suggested that the children may not yet have been conditioned to recognise faces. Therefore, to investigate a difference, the study by Scheerer et al. should be replicated with an age group similar to the participants in this article (generally, 10-12 years).

When examining assessment methods for this study, the AQR (Analysis system to evaluate the Quality of Relationship) was considered. This method inquire about the quality of the relationship between the therapist and recipient, which is important for the development of social skills [40, 44, 71]. As with other excluded assessment methods, this required the conductor to be trained in utilisation. Thus, for a longitudinal study, this evaluation method should be considered. As MT is recognised as a time-consuming process [26], studies are suggested to follow the development of recipients of MT for more than five months, if not years for seeing a significant difference in development (mentioned in segment 2.2.1). It should be made clear that assessment methods, that are deduced from studies about N-AIs, are not applicable in autism research due to the notable different comprehensions of social interaction. This was also discovered in segment 3.3.

Child-led sessions (recommended in section 2.2) provided the possibility to discover how the child expresses themselves without being limited by the consequences that there usually are in the real world. In multiple cases, the psychologists expressed their surprise at the children's abilities to interact with others, as well as distinguish what was the real world and what was mediated. Therefore, the IVR session can be a safe space for the child to initiate interaction in their own way - such as boxing the therapist or hugging others very frequently. Typically, are these behaviours immediately 'appropriated' and, as a result, the child may feel frustrated. With the IVR tool, the undesired behaviour can instead be addressed differently without suppressing the child's way of self-expression.

8.3 Consideration for Developing IVR Tools for Autism Interventions

When designing for people on the autism spectrum, it is important to recognise the importance of including autistic individuals and stakeholders to acquire a proper understanding of the autistic lifeworld.

For this project, the IVR tool was designed in cooperation with MTs and psychologists, taking an interaction design approach. As an alternative, participatory design could have been used, as it is deemed particularly effective when designing technological solutions for autistic individuals as it incorporates the ideas behind the double empathy problem [77]. However, this method has also been criticised as it can create a false image of autistic participation. It is rare, that autistic individuals are actual contributors in the process, that instead is dominated by the coordinators of the project [49].

The design process resulted in three iterations of the IVR tool for autism intervention. Different from other studies [1, 23, 27], this study incorporated music therapeutic elements and developed a multiplayer tool, where also the therapist is present in the VE. Besides the design considerations outlined in section 3, from this process, the most prominent design considerations are:

1. Expectations for the level of stimuli complexity should be aligned with stakeholders and, if possible, autistic participants.

2. The IVR tool should introduce the child progressively to the concepts incorporated in the VE.

3. As a starting point, the VE should only contain objects with a function.

4. Haptic feedback can be utilised to enhance interaction and emphasise important aspects of the VE (such as the other user's avatar).

5. The IVR tool should allow the therapist to mark the end of a session (e.g., a button that emits a 'gong' sound).

6. There should be both passive and active instruments.

Some studies further propose that individualised technological solutions are necessary to be effective in autism interventions [72]. Albeit, a tailored tool can be very expensive and thus be prohibitive for many in the autistic community - both for the therapists, other stakeholders, and autistic individuals [49]. Therefore, designing for versatility should be in focus.

In the concluding interview (segment 7.2), the psychologist emphasised that the therapist should know how to properly exploit the possibilities of the IVR tool. This standpoint is in coherence with Parsons et al., stating that for therapists and parents, "*training to use technologies can also be a crucial determinant of successful uptake of the technology*" [49] and that such a tool should not be classified as a toy. Therefore, cooperation between developers and therapists, similar to the previous study [52], is advised for the sake of creating a useful tool - both for the therapist and recipient. The function of the developer is to present the technology's capabilities properly to exploit its full potential. With this, also comes the knowledge about the tool's limitations. It was advised in segment 2.1.2 that training social skills should focus on arousal regulation and relaxation skills. Nonetheless, the children were not observed to be relaxed with the developed IVR tool. Albeit, there exist other therapy methods for inducing relaxation, such as Snoezelen [7]. Instead, IVR therapy can engage the child in the session and progressively expose the child to more complex stimuli, potentially facilitating the ability of self-regulation in more busy environments.

8.4 Representation of Players' Bodies

It was interesting that the test participants scored high in body awareness even with the simple avatar. This relates to the findings about simple VEs eliciting engagement (in segment 3.3. Also, Serafin et al. [66] mention that less realistic avatar hands provide a higher level of agency in the user. Thus, a simple avatar design is adequate for social skills training. Yet, future research should examine this concept in isolated studies, as significant improvements in social skills are found when using highly-realistic VEs and avatars [28]. To this, Feng et al. [21] already concluded that the uncanny valley effect is not exhibited in autistic children.

One issue with a simple avatar, is the lack of facial expressions and precise body language, that the therapist usually relies on for practising, i.e., turn-taking (mentioned in segment 2.2). Facial expressions are yet a challenge for IVR technology, however, eye-tracking technology exists [53], just not in standalone devices. It can be argued that facial expressions may also be more important for the therapist than the children, due to the challenges autistic individuals have with processing facial information, as mentioned in segment 2.1.3. Again, the level of realism of the avatar can be investigated, focusing on different possibilities of generating facial expressions and their impact on the child's social abilities as well as the therapist's abilities to correctly read the child's expressions. For more precise body language, inverse kinematics could be implemented for added expressiveness in the body language of the avatars. Here, additional gear is used to track the position of the child's feet and hip, however, this moves away from the initial requirement in segment 3.1, of the technological solution being standalone.

8.5 Generalisation of Learned Skills

With an IVR tool, it can be discussed, whether learned skills can be generalised in other situations. Because outside disturbances, the child can focus on the session and the therapist can receive more attention to train social skills with them. However, when placing the child in a different situation, there is not the same exclusion of disturbing stimuli, hence the child with have more difficulties in focusing in applying learned skills. Therefore, increasing complexity of stimuli is imperative for ecological validity.

As the cost of IVR displays decreases, it becomes more available for the general public, and the possibility to reinforce learned skills at home increases. In that way, parents can support their child's desire for social participation (as mentioned in segment 2.2). – PsI relates both to the exteroceptive and interoceptive sensations, which was found different for autistic individuals compared to N-AIs (segments 2.3.2 and 2.1.3). Yet, a child did express suspension of belief, meaning that the IVR may have a similar effect on autistic individuals as N-AIs in terms of being under the place illusion, PI.

A slow introduction of the different aspects within the VE, is suggested to better prepare the child for more social activities in a busy environment [16]. As the psychologists suggested, this could further lead to introducing more complex stimuli, which is in accordance with [7] and [72]. From here, the social training can move to simply being in the physical world, when the child is ready. Also, utilising VR for training - and then AR, which has shown promising results for partly incorporating the physical world [7, 72]. The success of VR compared to AR, may be caused by the challenges of inhibiting cognitive interference, as mentioned in segment 2.1.1. And then finally; moving to practise in the physical world. this is in accordance with the statements made by Tarantino et al. in segment 3.3.

8.5.1 Worries about the Digitisation of Therapy

A general concern is expressed by parents and some professionals about the digitisation of therapy, as they fear an increase of social withdrawal of their autistic relation [23]. To this, Parsons [49] advocate that the autistic movement has since the 1990s, where the personal computers became more mainstream, developed through online communities. Technology is seen as a gathering point for autistic people and a way to participate in societal discussions. With the increased popularity of social media, it became a powerful tool for enabling communication. Parsons continues the discussion about that academia focusing on validating technologies and methodology without involving the target group; autistic individuals [49].

9 Conclusion and Future Work

This study aimed to develop a tool, that facilitates social skills training for autistic children, and thus increase the likelihood of independence and improved life quality. Here, IVR technology's capability of simulating realistic stimuli was exploited to prompt valid responses in the user, as this provides high ecological validity. The basic principles of MT constituted the foundation for the developed IVR tool, as MT can facilitate social communication skills and interaction. The tool was developed in cooperation with autism experts, including MTs and psychologists working daily with autistic children.

The IVR tool was successful in facilitating social interaction and was further deemed appropriate for autistic perception. Future evaluation will focus on assessing the development of the participants' social abilities. The generalisation of the learned skills may be noticeable after several IVR sessions.

Acknowledgments

I would like to thank Professor in Music Therapy Ulla Holck and Music Therapist Anne Cathrine Hulthin Andersen for the great guidance, that shaped this study. Many thanks to psychiatrist and physician director Marie-Noëlle Clément at the Association CEREP-PHYMENTIN, Day hospital André Boulloche, for letting me conduct my studies at the day hospital. I would like to extend my deepest gratitude to psychologist and researcher Olivier Duris, clinical psychologist and Ph. D. student Charlotte Labossière, who provided invaluable insights into autism and constructive criticism for developing the IVR tool. A special thanks to my supervisor Ali Adjorlu, for his encouragement, and to ph.D. student Valentin Bauer (University of Paris-Saclay, CNRS, LISN), who shared his work and a great amount of assistance throughout the project.

Appendices

- **Appendix A:** Pedersen et al., "Exploring the Possibilities of Music Therapy in Virtual Reality: Social Skills Training for Adolescents with Autism Spectrum Disorder", unpublished work (*.pdf*).
- **Appendix B:** Follow-up Semi-structured Interview (*.pdf*).
- **Appendix C:** To-do list for the IVR sessions (*.pdf*).
- **Appendix D:** Recordings from the HMD of child 3, 4, 10, 12, and 13 (*.mp4*, zipped).
- **Appendix E:** Questionnaire for the sessions (*.pdf*).
- **Appendix F:** Data from Questionnaire (*.xlsx*).
- **Appendix G:** *Music Quest* - the evaluated IVR tool (*.apk*).
- **Appendix H:** Unity Project (zipped folder).

Videos

- **Video link:** First iteration test: https://youtu.be/J1LRAL_cARs
- **Video link:** Second iteration test: https://youtu.be/NCp9D0dHxt0

References

[1] Ali Adjorlu, Nathaly Belen Betancourt Barriga, and Stefania Serafin. "Virtual Reality Music Intervention to Reduce Social Anxiety in Adolescents Diagnosed with Autism Spectrum Disorder". en. In: *16th Sound and music computing conference*. ISBN: 9788409085187 Publisher: Zenodo. Sound and Music Computing Network. May 2019. DOI: 10.5281/ZENODO.3249339.

[2] Ahmad Yaser Alhaddad et al. "Robotic Trains as an Educational and Therapeutic Tool for Autism Spectrum Disorder Intervention". en. In: *Robotics in Education*. Ed. by Wilfried Lepuschitz et al. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing, 2019, pp. 249–262. ISBN: 978-3-319-97085-1. DOI: 10.1007/978-3-319-97085-1_25.

[3] American Psychiatric Association and American Psychiatric Association, eds. *Diagnostic and statistical manual of mental disorders: DSM-5*. en. 5th ed. Washington, D.C: American Psychiatric Association, 2013. ISBN: 978-0-89042-554-1.

[4] Felicity Baker et al. "Therapeutic Songwriting in Music Therapy". In: *Nordic Journal of Music Therapy* 17.2 (2008), pp. 105–123. DOI: 10.1080/08098130809478203.

[5] Leya Breanna Baltaxe-Admony et al. "Exploring the Creation of Useful Interfaces for Music Therapists". en. In: *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion*. Wrexham United Kingdom: ACM, Sept. 2018, pp. 1–7. ISBN: 978-1-4503-6609-0. DOI: 10.1145/3243274.3243307.

[6] Simon Baron-Cohen. “The extreme malebrain theory of autism”. en. In: *Trends in cognitive sciences* 6.6 (June 2002), pp. 248–254. DOI: 10.1016/S1364-6613(02)01904-6.

[7] Valentin Bauer, Tifanie Bouchara, and Patrick Bourdot. “Designing an Extended Reality Application to Expand Clinic-Based Sensory Strategies for Autistic Children Requiring Substantial Support: Participation of Practitioners”. en. In: *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. Bari, Italy: IEEE, Oct. 2021, pp. 254–259. ISBN: 978-1-66541-298-8. DOI: 10.1109/ISMAR-Adjunct54149.2021.00059.

[8] Scott Bellini. “The Development of Social Anxiety in Adolescents With Autism Spectrum Disorders”. en. In: *Focus on Autism and Other Developmental Disabilities* 21.3 (Aug. 2006), pp. 138–145. ISSN: 1088-3576, 1538-4829. DOI: 10.1177/10883576060210030201.

[9] Lucja Bieleninik et al. “Effects of Improvisational Music Therapy vs Enhanced Standard Care on Symptom Severity Among Children With Autism Spectrum Disorder: The TIME-A Randomized Clinical Trial”. en. In: *JAMA* 318.6 (Aug. 2017), p. 525. ISSN: 0098-7484. DOI: 10.1001/jama.2017.9478.

[10] Kristen Bottema-Beutel et al. “Avoiding Ableist Language: Suggestions for Autism Researchers”. In: *Autism in Adulthood* (Mar. 2021), pp. 18–29. DOI: 10.1089/aut.2020.0014.

[11] J. Bradt, D. S. Burns, and J. W. Creswell. “Mixed Methods Research in Music Therapy Research”. en. In: *Journal of Music Therapy* 50.2 (June 2013), pp. 123–148. ISSN: 0022-2917, 2053-7395. DOI: 10.1093/jmt/50.2.123.

[12] John Brooke. “SUS: A ’Quick and Dirty’ Usability Scale”. In: *Usability evaluation in industry* 189.3 (1996).

[13] John A. Carpente. *The Individual Music-Centered Assessment Profile for Neurodevelopmental Disorders (IMCAP-ND): A Clinical Manual*. eng. Regina Publishers, 2013. ISBN: 9780989179003.

[14] Nicholas Chown. “More on the ontological status of autism and double empathy”. In: *Disability & Society* 29.10 (2014), pp. 1672–1676. DOI: 10.1080/09687599.2014.949625.

[15] Nicholas Paul Chown. “The mismeasure of autism: a challenge to orthodox autism theory”. In: *Autonomy, the Critical Journal of Interdisciplinary Autism Studies* 1.2 (2013). ISSN: 2051-5189.

[16] Franceli L. Cibrian et al. “A Step towards Identifying the Sound Preferences of Children with Autism”. en. In: *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*. New York NY USA: ACM, May 2018, pp. 158–167. ISBN: 978-1-4503-6450-8. DOI: 10.1145/3240925.3240958.

[17] Sue V. G. Cobb et al. “Virtual Reality-Induced Symptoms and Effects (VRISE)”. In: *Presence: Teleoper. Virtual Environ.* 8.2 (Apr. 1999), pp. 169–186. ISSN: 1054-7460. DOI: 10.1162/105474699566152.

[18] Simon Davis, Keith Nesbitt, and Eugene Nalivaiko. “A Systematic Review of Cybersickness”. In: *Proceedings of the 2014 Conference on Interactive Entertainment*. IE2014. Newcastle, NSW, Australia: Association for Computing Machinery, 2014, pp. 1–9. ISBN: 9781450327909. DOI: 10.1145/2677758.2677780.

[19] Hanne De Jaegher. “Embodiment and sense-making in autism”. In: *Frontiers in Integrative Neuroscience* 7 (2013). ISSN: 1662-5145. DOI: 10.3389/fnint.2013.00015.

[20] Jelle van Dijk and Caroline Hummels. “Designing for Embodied Being-in-the-World: Two Cases, Seven Principles and One Framework”. In: *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*. TEI ’17. New York, NY, USA: Association for Computing Machinery, Mar. 2017, pp. 47–56. ISBN: 978-1-4503-4676-4. DOI: 10.1145/3024969.3025007.

[21] Shuyuan Feng et al. “The uncanny valley effect in typically developing children and its absence in children with autism spectrum disorders”. In: *PLOS ONE* 13 (Nov. 2018), e0206343. DOI: 10.1371/journal.pone.0206343.

[22] Andy Field. *DISCOVERING STATISTICS USING SPSS*. 3rd ed. Los Angeles: SAGE Publications, 2009. ISBN: 978-1-84787-906-6.

[23] Eynat Gal, P. L. (Tamar) Weiss, and Massimo Zancanaro. “Using Innovative Technologies as Therapeutic and Educational Tools for Children with Autism Spectrum Disorder”. In: *Virtual Reality for Psychological and Neurocognitive Interventions*. Ed. by Albert “Skip” Rizzo and Stéphane Bouchard. New York, NY: Springer New York, 2019, pp. 227–246. ISBN: 978-1-4939-9482-3. DOI: 10.1007/978-1-4939-9482-3_10.

[24] Shaun Gallagher. “Understanding Interpersonal Problems in Autism: Interaction Theory as An Alternative to Theory of Mind”. en. In: *Philosophy, Psychiatry, & Psychology* 11.3 (2004), pp. 199–217. ISSN: 1086-3303. DOI: 10.1353/ppp.2004.0063.

[25] M. Geretsegger et al. “Common Characteristics of Improvisational Approaches in Music Therapy for Children with Autism Spectrum Disorder: Developing Treatment Guidelines”. en. In: *Journal of Music Therapy* 52.2 (June 2015), pp. 258–281. ISSN: 0022-2917, 2053-7395. DOI: 10.1093/jmt/thv005.

[26] Monika Geretsegger et al. "Music therapy for people with autism spectrum disorder". en. In: *Cochrane Database of Systematic Reviews* (June 2014). Ed. by Cochrane Developmental, Psychosocial and Learning Problems Group. ISSN: 14651858. DOI: 10.1002/14651858.CD004381.pub3.

[27] Parisa Ghanouni et al. "Social Stories for Children with Autism Spectrum Disorder: Validating the Content of a Virtual Reality Program". en. In: *Journal of Autism and Developmental Disorders* 49.2 (Feb. 2019), pp. 660–668. ISSN: 0162-3257, 1573-3432. DOI: 10.1007/s10803-018-3737-0.

[28] Jorge Fernández Herrero and Gonzalo Lorenzo. "An immersive virtual reality educational intervention on people with autism spectrum disorders (ASD) for the development of communication skills and problem solving". en. In: *Education and Information Technologies* 25.3 (May 2020), pp. 1689–1722. ISSN: 1360-2357, 1573-7608. DOI: 10.1007/s10639-019-10050-0.

[29] Inês Hipólito. "On Autism and interaction theories of the self". In: *Altered Self and Altered Self-Experience* (2014), p. 263.

[30] Ulla Holck. "Turn-Taking in Music Therapy with Children with Communication Disorders". en. In: *British Journal of Music Therapy* 18.2 (Dec. 2004), pp. 45–54. ISSN: 1359-4575, 2059-9773. DOI: 10.1177/135945750401800203.

[31] Ulla Holck. "Music therapy research-children with an autism spectrum disorder". In: *Tidsskrift for Dansk Musikterapi* 8.2 (2011), pp. 27–35.

[32] "Interoception in Autism Spectrum Disorder: A review". In: 52 (2016), pp. 104–111. ISSN: 0736-5748. DOI: https://doi.org/10.1016/j.ijdevneu.2016.05.001.

[33] Daniel Johnston, Hauke Egermann, and Gavin Kearney. "Innovative computer technology in music-based interventions for individuals with autism moving beyond traditional interactive music therapy techniques". en. In: *Cogent Psychology* 5.1 (Dec. 2018). Ed. by Teppo Särkämö, p. 1554773. ISSN: 2331-1908. DOI: 10.1080/23311908.2018.1554773.

[34] Andreas Junker et al. "Exploring Social Skills Training Possibilities in an Open-ended Virtual Reality Application for Adolescents with Autism Spectrum Disorder". In: CHI2020 Social VR Workshop: A New Medium for Remote Communication & Collaboration. Apr. 2020.

[35] Ami Klin et al. "The enactive mind, or from actions to cognition: lessons from autism". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358.1430 (Feb. 2003). Publisher: Royal Society, pp. 345–360. DOI: 10.1098/rstb.2002.1202.

[36] Blythe LaGasse. "Social outcomes in children with autism spectrum disorder: a review of music therapy outcomes". en. In: *Patient Related Outcome Measures* Volume 8 (Feb. 2017), pp. 23–32. ISSN: 1179-271X. DOI: 10.2147/PROM.S106267.

[37] Joseph J. LaViola. "A discussion of cybersickness in virtual environments". In: *ACM SIGCHI Bulletin* 32.1 (Jan. 2000), pp. 47–56. ISSN: 0736-6906. DOI: 10.1145/333329.333344.

[38] Catherine Lord et al. "Autism spectrum disorder". en. In: *Nature Reviews Disease Primers* 6.1 (Jan. 2020). ISSN: 2056-676X. DOI: 10.1038/s41572-019-0138-4.

[39] Elysa J Marco et al. "Sensory Processing in Autism: A Review of Neurophysiologic Findings". en. In: *Pediatric research* 69.8 (2011), pp. 48–54.

[40] Amparo V. Marquez-Garcia et al. "Music Therapy in Autism Spectrum Disorder: a Systematic Review". en. In: *Review Journal of Autism and Developmental Disorders* (Feb. 2021). ISSN: 2195-7177, 2195-7185. DOI: 10.1007/s40489-021-00246-x.

[41] Meta. *META QUEST 2*. https://store.facebook.com/dk/en/quest/products/quest-2/. (accessed: 12.05.2022).

[42] Damian E. M. Milton. "On the ontological status of autism: the 'double empathy problem'". In: *Disability & Society* 27.6 (2012), pp. 883–887. DOI: 10.1080/09687599.2012.710008.

[43] Peter Mitchell, Sarah Cassidy, and Elizabeth Sheppard. "The double empathy problem, camouflage, and the value of expertise from experience". English. In: *Behavioral and Brain Sciences* 42 (2019). DOI: 10.1017/S0140525X18002212.

[44] Karin Mössler et al. "The Therapeutic Relationship as Predictor of Change in Music Therapy with Young Children with Autism Spectrum Disorder". en. In: *Journal of Autism and Developmental Disorders* 49.7 (July 2019), pp. 2795–2809. DOI: 10.1007/s10803-017-3306-y.

[45] Niels Nilsson, Rolf Nordahl, and Stefania Serafin. "Immersion Revisited: A Review of Existing Definitions of Immersion and Their Relation to Different Theories of Presence". In: *Human Technology* 12 (Nov. 2016), pp. 108–134. DOI: 10.17011/ht/urn.201611174652.

[46] Paul Nordoff and Clive Robbins. *Therapy in music for handicapped children*. eng. 3rd ed. London: Victor Gollancz, 1971. ISBN: 0575007559.

[47] NORMAL INC. *Normcore*. https://normcore.io/. (accessed: 12.05.2022).

[48] Sarah Parsons and Sue Cobb. "State-of-the-art of virtual reality technologies for children on the autism spectrum". In: *European Journal of Special Needs Education* 26.3 (2011), pp. 355–366.

[49] Sarah Parsons et al. "'Whose agenda? Who knows best? Whose voice?' Co-creating a technology research roadmap with autism stakeholders". en. In: *Disability & Society* 35.2 (Feb. 2020), pp. 201–234. ISSN: 0968-7599, 1360-0508. DOI: 10.1080/09687599.2019.1624152.

[50] Thomas D. Parsons. "Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences". en. In: *Frontiers in Human Neuroscience* 9 (Dec. 2015). ISSN: 1662-5161. DOI: 10.3389/fnhum.2015.00660.

[51] Pearson. *Kaufman Assessment Battery for Children — Second Edition Normative Update.* https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Kaufman-Assessment-Battery-for-Children-%7C-Second-Edition-Normative-Update/p/P100009004.html. (accessed: 21.05.2022).

[52] Linnea Bjerregaard Pedersen, Ali Adjorlu, and Valentin Bauer. "Exploring the Possibilities of Music Therapy in Virtual Reality: Social Skills Training for Adolescents with Autism Spectrum Disorder". unpublished. N.D.

[53] Linnea Bjerregaard Pedersen and Rolf Nordahl. "Experiencing Presence in a Virtual Reality Music Video". In: *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2021, pp. 86–89. DOI: 10.1109/VRW52623.2021.00023.

[54] Thiago M Porcino et al. "Minimizing cyber sickness in head mounted display systems: design guidelines and applications". In: *2017 IEEE 5th international conference on serious games and applications for health (SeGAH)*. IEEE. 2017, pp. 1–6.

[55] Jennifer Preece, Helen Sharp, and Yvonne Rogers. "The Process of Interaction Design". In: *Interaction Design: Beyond Human-Computer Interaction*. Wiley, 2015. Chap. 9.

[56] David Premack and Guy Woodruff. "Does the chimpanzee have a theory of mind?" In: *Behavioral and Brain Sciences* 1.4 (1978), pp. 515–526. DOI: 10.1017/S0140525X00076512.

[57] Danielle Provenzano. "Music Therapy for Communication in Children with Autism Spectrum Disorder". en. In: Academic Festival, Event 7. Digital Commons, 2020, p. 19.

[58] A. Raglio, D. Traficante, and O. Oasi. "A Coding Scheme for the Evaluation of the Relationship in Music Therapy Sessions". en. In: *Psychological Reports* (2006), pp. 85–90. DOI: 10.2466/pr0.99.1.85-90.

[59] Allison B. Ratto et al. "What about the girls? Sex-based differences in autistic traits and adaptive skills". In: *Journal of autism and developmental disorders* 48.5 (2018), pp. 1698–1711. DOI: 10.1007/s10803-017-3413-9.

[60] Brian Reichow and Fred R Volkmar. "Social skills interventions for individuals with autism: Evaluation for evidence-based practices within a best evidence synthesis framework". In: *Journal of autism and developmental disorders* 40.2 (2010), pp. 149–166.

[61] A. E. Reschke-Hernandez. "History of Music Therapy Treatment Interventions for Children with Autism". en. In: *Journal of Music Therapy* 48.2 (June 2011), pp. 169–207. ISSN: 0022-2917, 2053-7395. DOI: 10.1093/jmt/48.2.169.

[62] Dan Saffer. "The Four Approaches to Interaction Design". In: *Designing for Interaction: Creating Innovative Applications and Devices*. 2nd ed. New Riders, 2009. Chap. 2.

[63] Bob Rinse Schadenberg. English. PhD thesis. Netherlands: Human Media Interaction, Apr. 2021. ISBN: 978-90-365-5164-9. DOI: 10.3990/1.9789036551649.

[64] Nichole E. Scheerer et al. "Attention capture by trains and faces in children with and without autism spectrum disorder". In: *PLOS ONE* 16.6 (June 2021), pp. 1–28. DOI: 10.1371/journal.pone.0250763.

[65] Karin Schumacher and Claudine Calvet-Kruppa. "The "AQR" – an Analysis System to Evaluate the Quality of Relationship during Music Therapy: Evaluation of interpersonal relationships through the use of instruments in music therapy with profoundly developmentally delayed patients". en. In: *Norsk Tidsskrift for Musikkterapi* 8.2 (July 1999), pp. 188–191. ISSN: 0803-9828. DOI: 10.1080/08098139909477974.

[66] Stefania Serafin et al. "Virtual Reality Musical Instruments: State of the Art, Design Principles, and Future Directions". en. In: *Computer Music Journal* 40.3 (Sept. 2016), pp. 22–40. ISSN: 0148-9267, 1531-5169. DOI: 10.1162/COMJ_a_00372.

[67] Mojtaba Shahab et al. "Utilizing social virtual reality robot (V2R) for music education to children with high-functioning autism". en. In: *Education and Information Technologies* 27.1 (Jan. 2022), pp. 819–843. ISSN: 1360-2357, 1573-7608. DOI: 10.1007/s10639-020-10392-0.

[68] Elizabeth Sheppard et al. "How easy is it to read the minds of people with autism spectrum disorder?" In: *Journal of autism and developmental disorders* 46.4 (2016), pp. 1247–1254. DOI: 10.1007/s10803-015-2662-8.

[69] Mel Slater. "A note on presence terminology". In: *Presence connect* 3.3 (2003), pp. 1–5.

[70] Mel Slater. "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535 (2009), pp. 3549–3557.

[71] Neta Spiro, Giorgos Tsiris, and Charlotte Cripps. "A Systematic Review of Outcome Measures in Music Therapy". en. In: *Music Therapy Perspectives* 36.1 (Apr. 2018), pp. 67–78. ISSN: 0734-6875, 2053-7387. DOI: 10.1093/mtp/mix011.

[72] Laura Tarantino et al. "Immersive applications: what if users are in the autism spectrum?" en. In: *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*. Brisbane QLD Australia: ACM, Nov. 2019, pp. 1–7. ISBN: 978-1-4503-7002-8. DOI: 10.1145/3359997.3365696.

[73] The Children's Hospital of Philadelphia. *CHILDHOOD AUTISM RATING SCALE, 2ND EDITION (CARS2)*. https://www.carautismroadmap.org/childhood-autism-rating-scale/. (accessed: 21.05.2022).

[74] Unity Technologies. *Unity 2020.3.18*. https://unity3d.com/unity/whats-new/2020.3.18. (accessed: 12.05.2022).

[75] Unity Technologies. *XR Interaction Toolkit*. https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@1.0/manual/index.html. (accessed: 12.05.2022).

[76] Petra Hendrika Johanna Maria Vlamings et al. "Basic abnormalities in visual processing affect face processing at an early age in autism spectrum disorder". In: *Biological psychiatry* 68.12 (2010), pp. 1107–1113.

[77] Thijs Waardenburg et al. "Design Your Life: User-Initiated Design of Technology to Support Independent Living of Young Autistic Adults". In: *arXiv:2105.12370 [cs]* 12780 (2021). arXiv: 2105.12370, pp. 373–386. DOI: 10.1007/978-3-030-78224-5_26.

[78] Eric G Waldon et al. *Music Therapy Assessment: Theory, Research, and Application*. eng. London: Jessica Kingsley Publishers, 2018. ISBN: 9781785922954.

[79] N. Wasserman et al. "A Music Therapy Evaluation Scale and Its Clinical Application to Mentally Retarded Adult Patients". en. In: *Journal of Music Therapy* 10.2 (June 1973), pp. 64–77. ISSN: 0022-2917, 2053-7395. DOI: 10.1093/jmt/10.2.64.

[80] World Health Organization. *ICD-10 Version:2019*. https://icd.who.int/browse10/. (accessed: 18.05.2022).

[81] Robyn L. Young and Miriam Posselt. "Using The Transporters DVD as a Learning Tool for Children with Autism Spectrum Disorders (ASD)". en. In: *Journal of Autism and Developmental Disorders* 42.6 (June 2012), pp. 984–991. ISSN: 0162-3257, 1573-3432. DOI: 10.1007/s10803-011-1328-4.