Published in final edited form as:

ACM Trans Access Comput. 2018 November ; 11(4): . doi:10.1145/3282434.

Extraction of Emotional Information via Visual Scanning Patterns: A Feasibility Study of Participants with Schizophrenia and Neurotypical Individuals

Joshua Wade,
Department Of Mechanical Engineering, Vanderbilt University

Heathman S. Nichols,
Department Of Psychology, Vanderbilt University

Megan Ichinose,
Department Of Psychology, Vanderbilt University

Dayi Bian,
Department Of Electrical Engineering & Computer Science, Vanderbilt University

Esube Bekele,
Department Of Electrical Engineering & Computer Science, Vanderbilt University

Matthew Snodgress,
Department Of Psychology, Vanderbilt University

Ashwaq Zaini Amat,
Department Of Electrical Engineering & Computer Science, Vanderbilt University

Eric Granholm,
Department Of Psychiatry, University Of California, San Diego

Sohee Park, and
Department Of Psychology, Vanderbilt University

Nilanjan Sarkar
Department Of Mechanical Engineering, Vanderbilt University

Abstract

Emotion recognition impairment is a core feature of schizophrenia (SZ), present throughout all stages of this condition, and leads to poor social outcome. However, the underlying mechanisms that give rise to such deficits have not been elucidated and hence, it has been difficult to develop precisely targeted interventions. Evidence supports the use of methods designed to modify patterns of visual attention in individuals with SZ in order to effect meaningful improvements in social cognition. To date, however, attention-shaping systems have not fully utilized available technology (e.g., eye tracking) to achieve this goal. The current work consisted of the design and feasibility testing of a novel gaze-sensitive social skills intervention system called MASI-VR. Adults from an outpatient clinic with confirmed SZ diagnosis (n=10) and a comparison sample of neurotypical participants (n=10) were evaluated on measures of emotion recognition and visual attention at baseline assessment, and a pilot test of the intervention system was evaluated on the SZ sample following five training sessions over three weeks. Consistent with the literature, participants in the SZ group demonstrated lower recognition of faces showing medium intensity fear, spent more time deliberating about presented emotions, and had fewer fixations in comparison to neurotypical peers. Furthermore, participants in the SZ group showed significant improvement in the recognition of fearful faces post-training. Preliminary evidence supports the feasibility of a gaze-sensitive paradigm for use in assessment and training of emotion recognition and social attention in individuals with SZ, thus warranting further evaluation of the novel intervention.

1 INTRODUCTION

Emotion recognition impairment is a core feature of schizophrenia (SZ), present throughout all stages of this condition, and leads to poor social outcome, but the underlying mechanisms that give rise to such deficits have not been elucidated and hence, it has been difficult to develop precisely targeted interventions [Green et al. 2008; Kohler et al. 2003; Morris et al. 2009]. Emotion recognition is not a unitary construct but consists of multiple underlying perceptual and cognitive mechanisms. One important component of reduced emotion recognition may arise from an inability to direct attention to socially relevant aspects of the environment in order to extract appropriate information in real time [Loughland et al. 2002; Russell et al. 2008]. In neurotypical individuals, affective significance plays a key role in capturing attention, thereby influencing social cognition [Öhman et al. 2001]. Individuals with SZ demonstrate patterns of visual attention that are markedly different from neurotypical individuals in a variety of contexts [Delerue et al. 2010; Loughland, et al. 2002; Quirk and Strauss 2001; Suzuki et al. 2009]. Compared to their neurotypical peers, for example, individuals with SZ tend to display impaired exploratory eye movement [Suzuki, et al. 2009] and direct less attention to emotionally salient components of the faces of others [Loughland, et al. 2002]. Atypical allocation of attention to stimulus areas rich in social information (e.g., eyes, mouth, body posture) may lead to impaired decoding of facial affect, which, in turn, is associated with difficulties in interpersonal interactions and social competence [Bornhofen and McDonald 2008; Hills et al. 2016].

In SZ, social deficits are a major concern because they impede recovery and rehabilitation. One of the most difficult challenges facing people with SZ as they struggle to enter the work force, finish school, or maintain relationships is their inability to exercise social skills in daily interactions [Bellack et al. 2006; Couture et al. 2011]. Because social cognition in SZ is closely associated with functional outcome, the improvement of social cognition in SZ has become a major research initiative in recent years [Green et al. 2008; Tan et al. 2016]. Efforts to address social cognition deficits in SZ have produced a wide range of treatment approaches with some focusing on specific components of social cognition (e.g., Theory of Mind and emotion recognition), while others aim to produce comprehensive change, and the evidence to date demonstrates the malleability of social cognition in response to these varied approaches [Tan et al. 2016]. Researchers have posited that interventions designed to shift users' attention towards emotionally salient facial elements could represent a particularly effective approach to training [Russell et al. 2008; Morris et al. 2009]. However, researchers have not effectively utilized specialized technologies such as eye tracking to develop attention-shaping and individualized interventions.

Using an eye tracking device, attention to social stimuli can be quantified by measuring the visual scanning pattern, or *gaze*, of an individual on representative social content (e.g., images of emotionally expressive faces). Thus, it is possible to infer qualities of social attention from an individual's gaze with regards to stimuli viewed during social interactions. However, tracking gaze information during interactions in naturalistic environments, although possible (e.g., [Fletcher and Zelinsky 2009]), is prohibitively difficult and impractical for social interactions due to the lack of controllability and scalability as well as high cost. Alternatively, Virtual Reality (VR), which is increasingly popular both generally and in mental health research [Freeman et al. 2017], offers a highly controllable, safe, comfortable, repeatable, engaging, and accessible option for simulating social interactions while collecting measures of gaze. In addition, the observation of precise locations of Regions of Interest (ROI), such as faces and facial components, is straightforward in a VR environment, permitting highly accurate gaze measurement with respect to researcher-defined ROI. Moreover, the coupling of VR with eye tracking functionality enables the development of *closed-loop* systems in which gaze information can be used as a complementary input to the system, making possible the adaptation of the system to the individual according to measured aberrations.

Given evidence of the ability to use VR to re-train patterns of attention for a variety of populations and tasks [Achtman et al. 2008; Bruce et al. 2017; Chukoskie et al. 2017], we hypothesize that a system that combines task-related performance feedback and real-time gaze-sensitive feedback may be a particularly effective tool for social skills intervention in individuals with SZ. In the current work, we present the application of such a VR system as well as the results of a feasibility study evaluating individuals with SZ (n=10) and a group of neurotypical controls (n=10). We first compare baseline measures of controls and persons with SZ to gauge the preliminary validity of the novel system and to explore gaze measures captured by the system. Using a pretest/posttest style evaluation, we then assess preliminary training effects of the system on the sample of individuals with SZ (n=9 at posttest). The remainder of this paper is organized as follows. In Section 2, we discuss literature related to social skills training in SZ. Section 3 contains detailed descriptions of the novel and extended software systems used in this research and Section 4 discusses the initial evaluations of these systems. In Section 5, the results of the evaluations are presented and Section 6 concludes the paper with a discussion of the current work's major contributions, and outlines the research limitations and planned future work.

2 RELATED WORK

The literature provides ample evidence of the differences between individuals with SZ and neurotypical individuals on measures of facial recognition, emotion recognition, Theory of Mind, and attention to social stimuli [Green et al. 2008; Kohler, et al. 2003; Loughland, et al. 2002; Morris, et al. 2009]. Despite increased attention of researchers to social cognition training in SZ [Green, et al. 2008; Tan, et al. 2016], as well as calls from researchers to use novel methods to help shape visual attention in individuals with SZ [Morris et al. 2009; Russell et al. 2008], the development of advanced technological systems capable of improving social cognition in individuals with SZ has not been prioritized to date. Here we discuss the state-of-the-art in the assessment and enhancement of emotion recognition skills in individuals with SZ and show the need for further work in this area.

Measurement of emotion recognition skills in SZ has been investigated using a variety of stimuli, including static photographs of people displaying emotions [Comparelli et al. 2013; Kohler et al. 2003] and precisely configurable 3D models of faces [Bekele et al. 2017]. Kohler et al. (2003) presented static photographs of faces displaying five different emotions (anger, fear, disgust, joy, and sadness) at two distinct levels of intensity (mild and extreme) to a group of individuals with SZ and a group of controls. They found that performance as measured by emotion recognition accuracy was significantly poorer in the SZ group across both levels of intensity, but especially so with regards to the recognition of fear, disgust, and neutral expressions [Kohler, et al. 2003]. Diminished recognition of fear in particular is reported consistently in the literature and may have a neurobiological basis [Morris, et al. 2009]. More recent work also provides support for a particular impairment in the recognition of negative emotions [Comparelli, et al. 2013], as well as heterogeneity of emotion recognition skills according to symptom severity and SZ sub-type [Comparelli et al. 2014; Sachse et al. 2014].

Probing further, other researchers have used eye tracking devices to measure the proportion of time that individuals with SZ spend looking at specific ROI (e.g., eyes, nose, and mouth) during tasks of facial emotion recognition. Typical measures of visual attention include fixations, saccades, and scan paths [Salvucci and Goldberg 2000]. A time series recording of gaze information is characterized by periods of relatively restricted focus (i.e., fixations) separated by rapid changes in fixation location (i.e., saccades). By measuring the gaze of participants with and without SZ, Quirk and Strauss showed that fixation durations on emotionally activating images were longer on average for the SZ sample, which the authors interpreted as an impairment in the ability to efficiently extract meaningful information from presented images [Quirk and Strauss 2001]. A pattern of fewer fixations with longer durations has also been demonstrated with respect to the viewing of images of emotional faces, where, compared to controls, individuals with SZ tend to fixate longer on emotional expressions than on neutral expressions [Delerue, et al. 2010; Loughland, et al. 2002]. Because recognition of emotions is inextricably tied to the allocation of visual attention to key components of the face, researchers have put forth that systems capable of directing gaze to these key components could result in improvements in emotion recognition [Morris, et al. 2009; Russell, et al. 2008]. Despite the clear potential of utilizing real-time gaze tracking in addressing emotion recognition deficits in SZ, researchers have only used eye tracking as a tool for assessment of treatment effects (e.g., [Drusch et al. 2014]) and not as a feedback signal in a closed-loop system.

There is substantial evidence to support the efficacy of intervention tools for improving emotion recognition in individuals with SZ [Morris, et al. 2009; Tan, et al. 2016]. Russell et al. used the Micro-Expression Training Tool (METT) to deliver emotion recognition training to a group of individuals with SZ and a group of controls. METT presents the user with instructional material regarding emotion identification, including videos that contrast commonly confused pairs of emotions such as fear and surprise. They found that participants receiving the full course of training with METT demonstrated significantly larger posttest improvement in emotion recognition compared to individuals who did not receive this level of training [Russell, et al. 2008]. Recent research supports the use of advanced technologies such as immersive VR and physiological sensors to deliver engaging and adaptive computer-based social skills intervention [Bekele et al. 2017; Freeman, et al. 2017; Veling et al. 2016]. Park et al. compared the effects of two role-play based social skills interventions in a large sample of individuals with SZ: one using an immersive VR design and the other using a traditional participant-therapist interaction design. The researchers showed that the VR-based design was more engaging to participants and resulted in larger gains on some outcome measures [Park et al. 2011]. Rus-Calafel et al. evaluated a small sample of individuals with SZ using a series of VR-based social interaction scenarios involving highly realistic avatars. The avatars were capable of displaying several types of emotions, engaging in dialog with the user, and providing constructive feedback to the user. The researchers' results showed a remarkable breadth of improvements post-training, including reductions in negative symptoms and increased emotion recognition; participants also reported strong engagement and enjoyment of the VR system [Rus-Calafell et al. 2014].

Bekele et al. implemented the VR System for Affect Analysis in Facial Expressions (VR-SAAFE) for the purpose of delivering affect-sensitive training of emotion recognition to individuals with SZ. Their preliminary work demonstrated the feasibility of training Machine Learning models on measures of gaze (e.g., fixation counts and scan path lengths) and physiology (e.g., heart rate variability and skin conductance) in order to design a training protocol capable of adjusting to the user's emotional state in real time [Bekele, et al. 2017]. However, to the best of our knowledge, no work has presented the design and evaluation of a gaze-sensitive social skills training paradigm for individuals with SZ. Given evidence supporting the use of training that is designed to shift visual attention towards salient regions of emotion-displaying faces [Morris, et al. 2009], as well as the feasibility of systems that adapt to biofeedback [Bekele, et al. 2017; Chukoskie, et al. 2017], we propose a gaze-sensitive social skills training paradigm as a novel technology for improving recognition of emotions in individuals with SZ.

3 SYSTEM DESIGN

The primary contribution of this paper is the extension and feasibility evaluation of a novel gaze-sensitive training modality within an existing social skills training platform called Multimodal Adaptive Social Intervention in Virtual Reality (MASI-VR). MASI-VR was previously developed and tested as a training tool for teens with Autism Spectrum Disorders (ASD) [Bekele et al. 2016], but has been substantially modified in the current work and has been re-tailored for application in indivdiuals with SZ. In addition, this paper introduces the Emotion Recognition and RATing Application (ERRATA), a tool for fine-grained assessment of emotion recognition and visual attention.

3.1 Emotion Recognition and Rating Application (ERRATA)

A system for presenting tasks of emotion recognition was created for this research. This system, called ERRATA, is largely a modification of a system used in our prior work [Bekele et al. 2014], but introduces finer control of emotion presentation and data acquisition while removing unneeded features of the past work. ERRATA was designed to display faces of a diverse set of highly realistic virtual characters (henceforth, “avatars”), varying by age, gender, race, ethnicity, emotional expression, and emotional intensity. In total, 12 avatars were included in ERRATA: three white males (two adults and one child), three white adult females (one elderly), three non-white adult males (one Asian American, one African American, and one Indian American), and three non-white adult females (one Indian American and two Hispanic Americans). Avatar models were created using a software called *Evolver* (now Autodesk) and then rigged and animated in another software called *Maya* according to the procedures detailed in Bekele et al. 2014. Each avatar was designed to display seven emotional expressions of variable intensity as well as a neutral expression. These seven emotional expressions were proposed by the psychologist, Ekman to be accepted universally and consist of joy, surprise, sadness, fear, disgust, contempt, and anger [Ekman 1993; Biehl et al. 1997]. Four categories of intensity (i.e., the strength of the emotion) were designed for each emotion. Following the conventions of our prior work [Bekele, et al. 2017; Bekele, et al. 2014], these categories consisted of low, medium, high, and extreme intensity and were specified by a 20-point scale (zero corresponding to the absence of emotional display and 20 corresponding to maximum emotional intensity). Note that neutral expressions were not included as a component of testing because they could neither convey transition nor vary by intensity. Although the design of the emotion animations was already validated in our prior work, six neurotypical individuals (three males and three females) were recruited to verify that the newly animated avatars displayed facial expressions consistent with those of our prior implementation [Bekele et al. 2014]. These individuals were shown each of the seven emotions described above at the high intensity level and were then asked to report the emotion perceived. Responses indicated a mean overall accuracy of 87.01%, which far exceeds random chance (i.e., 14.29%) and is consistent with results from our prior implementation as well as with the broader literature [Tottenham et al. 2009], indicating that the presented emotions were recognizable among neurotypical individuals.

ERRATA is an emotion recognition task that uses dynamic avatar stimuli to present a range of emotions (see Figure 1). ERRATA presents the neutral expression of an avatar for 2.5 seconds followed by an instantaneous transition to the emotionally expressive face, which is shown for 2 seconds. The user is then asked to select an emotion description from a list of the seven aforementioned emotions that accurately describes the emotion expressed by the avatar. Finally, the user is asked to indicate how confident he/she is in her/his decision on a scale ranging from “Very unsure” to “Very confident” using a 10-point real-valued scale. The neutral expression at the beginning of each trial was deliberately shown longer to account for initial onset of gaze when faces were presented. Unique configurations of ERRATA’s presentation sequence were designed in order to avoid the potential of practice effects on performance assessment. Given the 336 possible combinations of avatar, emotional intensity, and type of emotion (12 × 4 × 7), it was necessary to optimize presentation duration to allow timely completion of tasks. For this reason, we used three emotional intensities rather than four, omitting the high intensity category while keeping the extreme ends of the scale as well as the medium level. Therefore, 252 combinations of avatar, emotional intensity, and type of emotion were randomly allocated to three distinct configurations of the ERRATA system, resulting in 84 unique combinations within each configuration. We estimated that users would require approximately 10 seconds of deliberation for each face in order to indicate emotion type and confidence, thus yielding an expected total duration of [(2.5 neutral + 2 emotional) + 10 deliberation] × 84 seconds ≈ 20 minutes per assessment (or approximately 1 hour across three unique assessments).

ERRATA collects measures of performance and visual attention. Performance logs detail users' responses to system-generated prompts, which includes perceived emotion (i.e., which emotion the user believes he/she is observing), self-reported confidence in emotion selection, and deliberation time of emotion selection. Accuracy in emotion recognition is computed as the number of instances in which emotions are correctly identified divided by the total number of presented emotions. Accuracy can be further analyzed with regards to individual emotion types and/or levels of emotional intensity. This is useful because evidence indicates that accuracy is unevenly distributed with regards to emotion type [Tottenham, et al. 2009], intensity of emotion [Hoffmann et al. 2010], and population [Bal et al. 2010; Kohler, et al. 2003]. Deliberation time and confidence in emotion selection are exploratory measures that may provide additional insight into performance on the emotion recognition tasks. We expected that these measures would vary according to the intensity of displayed emotions. Specifically, we expected that confidence would be lower for low intensity emotions and higher for medium and extreme intensity emotions. Similarly, we expected that deliberation time would be longer for low intensity emotions and shorter for medium and extreme intensity emotions.

ERRATA interfaces with an eye tracking device to collect information about spatial and temporal aspects of users' gaze. Data logs produced by the system track gaze position and pupil diameter over time. From these logs, higher-order measures of visual attention can be derived, including fixation durations, saccades, and blink rates, all of which provide insight into the regions of the face that users attend to during emotion recognition tasks. In this work, we were interested in measuring fixation duration with respect to salient ROI on avatars' faces that are associated with emotional expression. For each face shown, ERRATA measured the amount of time that users spent looking at an avatar's eyes, nose, mouth, forehead, chin, and cheeks. Based on the literature, we expected that participants in the SZ group would demonstrate fewer fixations than neurotypical peers at baseline evaluation and would show an increase in number of fixations following training with MASI-VR [Loughland et al. 2002; Delerue et al. 2010].

3.2 Multimodal Adaptive Social Intervention in Virtual Reality (MASI-VR)

MASI-VR is a VR-based social skills training tool that was originally developed for individuals with ASD [Bekele, et al. 2016], and consists of a virtual cafeteria environment populated by avatars with whom the user can engage in conversations about a variety of topics (see Figure 2). The development of MASI-VR was motivated by a logical progression of increasing naturalism in the evaluation of emotion recognition and social interaction skills using VR. A number of researchers have used static images or virtual models of faces to assess emotion recognition accuracy in various populations [Tottenham, et al. 2009]. Researchers later introduced the notion of context into emotion recognition tasks by presenting emotionally-relevant story vignettes alongside emotional expressions (e.g., a virtual character recalls a pleasant memory from a family vacation while showing an expression of joy) [Bekele, et al. 2014]. In two important ways, MASI-VR further expands upon this trend of increasing naturalism. First, while other systems present arbitrarily-ordered sequences of faces showing emotions, virtual avatars in MASI-VR display emotions that are relevant to the context of a conversation that takes place between the user and the avatar. In such conversations, the avatar will, for instance, show an expression of joy when exchanges are reciprocal and on-topic, but will show emotions of negative valence when exchanges veer into areas that are either too direct, too personal, or off-topic. Second, MASI-VR permits the user to explore the virtual environment and to engage in conversations with avatars of their choosing.

Conversations in MASI-VR have a predefined structure, the graph of which resembles a tree. At the root of the tree, the user selects a statement or question from among a finite set of possible statements. Next, the avatar responds appropriately based on the correctness of the user's selection. If the user's opening statement or question is correct (i.e., on-topic based on a predefined structure), then the avatar responds with a topic-appropriate affirmation. If, however, the user's opening line is off-topic or otherwise judged to be incorrect, then the avatar's response may vary from confusion to surprise. The depth of this conversation tree is variable, depending in part on the accuracy of the user in making selections as well as the level of difficulty of the conversation (i.e., greater depth corresponds to greater difficulty). In addition to the feedback that users receive from the avatars during conversations, MASI-VR also produces feedback spoken by a virtual "facilitator." This facilitator inserts constructive feedback between conversational exchanges, which are designed to reinforce appropriate responses while outlining the weaknesses of inappropriate responses. It should be noted that, although the text was designed to be understood by people of average cognitive ability, reading impairments are characteristic of individuals with SZ [Revheim et al. 2006]. Because the interpretation of textual information is key to success in MASI-VR, the text presentation was also accompanied by pre-recorded audio narration produced by human speakers. Examples of conversational exchanges within MASI-VR are given below:

Example Exchange #1

Goal: *Find out how Avatar liked the shopping experience*.

Difficulty: *Easy*

1. User: "I really enjoyed shopping today. How did you like it?" [*Correct choice*]
2. Avatar: "It was fun. I'd like to come again."
3. Facilitator: "Excellent choice. You correctly offered some information about yourself and met your goal of learning how Avatar liked the shopping experience."
4. [*End of exchange*]

Example Exchange #2

Goal: *Find out if Avatar would like to get coffee before work.*

Difficulty: *Medium*

1. User: "Do you want to grab coffee with me tomorrow?" [*Incorrect choice*]
2. Avatar: "I'm not sure, I might be busy."
3. Facilitator: "Very good try. This question might get you the response you're looking for. However, you haven't provided very many details about yourself or where you want to get coffee, so Avatar might not be completely comfortable agreeing to get coffee with you yet. Try again."
4. User: "I always like to go to the Coffee Stop about a block from the bus station. Have you ever been there?" [*Correct choice*]
5. Avatar: "No, I haven't! That sounds like a cool place to get your morning started!"
6. Facilitator: "Excellent choice. You correctly offered some information about yourself and are one step closer to meeting your goal of asking Avatar to get coffee with you before work. This is the best response because it provides information about you, and establishes a place that you and Avatar can go. Now you're ready to ask Avatar to go with you!"

[*Exchange continues*]

As its name implies, key components of MASI-VR are its support for multi-modal input and individualized adaptation. In addition to gaze measurement, MASI-VR was previously integrated with modules for measuring psychophysiological information (i.e., the prediction of user affect from physiological measures such as heart rate variability and electrodermal activity) as well as electroencephalogram [Bekele, et al. 2016]. While past work with MASI-VR included functionality for real-time adaptation of the system in response to physiological and gaze measures, the specific adaptive mechanism used in the current work are entirely new and are described in detail in the next section.

3.3 Extension of MASI-VR for the Current Study

MASI-VR was significantly expanded for the present work, thus warranting a detailed discussion of the modifications and extensions. The most important of these changes was the addition of a new method of gaze-sensitive task adaptation. While the earlier implementation of MASI-VR contained functionality for gaze-sensitivity, the approach used in the past version is wholly different from that of the current work. The past implementation of gaze-sensitivity in MASI-VR relied on a facial occlusion paradigm in which participants could not see the relevant components of an avatar's face until a sufficient proportion of gaze had fallen on the salient areas of the face (e.g., eyes and mouth), thus revealing the face completely. The current system, however, removes the occlusion paradigm and instead utilizes a gaze-sensitive mechanism for progression through conversational exchanges. That is, in order to progress through tasks, users must direct their gaze at least once per conversation to the face of the avatar with whom they are engaged. The face of the avatar remains highlighted with a transparent green mask until the user directs attention to that region (see Figure 4). Once the user looks at the avatar's face, the mask is removed for the duration of the exchange and the user is permitted to proceed with the conversation. The motivation for the new gaze-sensitive functionality is three-fold. First, there is a need to ensure that participants actually view the face of the avatar, because it is possible to complete the tasks never having looked at the avatar (e.g., instead focusing only on the dialog boxes). Second, by establishing appropriate eye gaze at the avatar, this system aims to hone patients' social attention towards the relevant conversation partner. While gaze-sensitive VR systems have been tested in other populations (e.g., [Chukoskie et al. 2017]), to our knowledge, no VR-based social skills training using a gaze-sensitive methodology has been tested in individuals with SZ. Lastly, requiring users to look at the faces of avatars also served to ensure that gaze was consistently tracked, because users' posture can shift over time, thus compromising the quality of measured gaze data.

Another substantial update to MASI-VR concerns the expansion in both the quantity and variety of training content. The study described in this work concerns social skills training that is carried out during five sessions over a period of three weeks. The original implementation of MASI-VR consisted only of a high school cafeteria scene (Figure 2) with 12 conversation topics tailored towards adolescents. Thus, the number and variety of conversations would be insufficient for a multi-session social skills intervention. In order to provide sufficient and diverse content for a series of sessions, MASI-VR was extended to include two new scenes with accompanying conversation scripts. In addition to the original cafeteria scene, a bus stop scene and a grocery store scene were designed and implemented (Figure 3). The new bus stop scene consisted of a city block bounded by two-way streets, a bus stop area populated by avatars, and buses parked on the streets nearby. The models for city streets, intersections, and buildings were generated using the software *CityEngine* (www.esri.com/software/cityengine), and models for vehicles and other objects were obtained from online repositories. The new grocery store scene included rows of shelves with a variety of products, such as canned goods, breakfast cereal, and boxes of other dry goods. Many of the avatars in this scene stood next to shopping carts while browsing food items. Models for this scene were either created using Maya or obtained from online repositories. New conversation scripts were created by researchers with expertise in SZ intervention for each of the three scenes and were appropriately designed for both the environment and population (i.e., adults with schizophrenia).

4 PILOT STUDY

A pilot study involving adults with SZ and a comparison group of neurotypical individuals was conducted in order to (A) evaluate group-level differences with regards to measures of emotion recognition and visual attention on emotionally expressive faces, (B) assess the feasibility and preliminary efficacy of the gaze-sensitive VR social skills intervention, and (C) explore the relationship between quantitative markers of social attention and emotion recognition in individuals with SZ.

4.1 Participants

Ten individuals who met the DSM-5 criteria for SZ were recruited from a private outpatient facility. The SZ group was characterized by a mean age of 45.2 years (SD=6.05), mean educational level of 12.7 years (SD=1.42), and mean duration of condition of 24.78 years (SD=9.72). Clinical symptoms were assessed with the Scale for the Assessment of Positive Symptoms (SAPS) [Andreasen 1984] and the Scale for the Assessment of Negative Symptoms (SANS) [Andreasen 1989]. See Table 1 for a full report of participant characteristics. The SAPS quantifies the severity of positive symptoms such as hallucinations, delusions, and thought disorder, while the SANS quantifies the severity of negative symptoms such as alogia, anhedonia, flat affect, and anergia. The SAPS and SANS yield scores of symptom severity in individuals who have already been diagnosed with SZ. Note that they are not diagnostic instruments with thresholds. The SAPS consists of 34 items while the SANS consists of 25 items, both using a 0–5 Likert-type scale. For both assessments, higher scores correspond to greater severity of symptoms. With regards to interpretation, we note that the symptoms scores reported in Table 1 are within 1 standard deviation of typical patient scores from a multi-site sample [Van Erp et al. 2014].

Ten neurotypical individuals (i.e., controls) with no history of mental illness or neurological disorders were recruited from the community. The mean age of participants in the control group was 43.2 years (SD=10.18), with a mean education level of 15.4 years (SD=2.12). The two groups did not show a significant difference with respect to participant age, $t(18)=-0.53$, $p=.60$. Although the education level (years of education) in the control group was significantly greater than in the SZ group, $t(18)=3.35$, $p=.004$, the groups did not differ significantly on full-scale IQ as measured by the Wechsler Abbreviated Scale of Intelligence (WASI) [Stano 1999], $t(17)=1.84$, $p=.08$. Consistent with the literature, however, participants in the SZ group had significantly lower scores on the verbal component of the WASI [Revheim, et al. 2006]. Despite this difference, both groups possessed normal overall cognitive functioning and were determined to be capable of participating in the study. The experimental protocol was approved by the Institutional Review Board of Vanderbilt University. Written informed consent was obtained after procedures were fully explained. All subjects were paid for their participation as compensation for time and travel.

4.2 Procedures

Figure 4 gives a detailed outline of the study procedures carried out in this work. All participants provided informed consent before engaging in any of the other study activities. Following consent, a battery of clinical assessments was administered by expert personnel trained in the administration of such assessments to individuals with SZ; as previously mentioned, these instruments included SANS, SAPS, and WASI (WASI data were obtained from all but one control). After obtaining consent from all participants and symptoms data from patients, all participants completed an initial evaluation using the novel ERRATA system (i.e., pretest). All participants were exposed to the same sequence of faces as described in Section 3.1 in regards to the three unique configurations of ERRATA. It was initially estimated that participants would complete an entire ERRATA session in approximately 20 minutes. In practice, however, a mean completion time of 25.9 min (SD=5.56 min) across all participants was observed.

For controls, participation in the study was concluded following the pretest evaluation. For members of the SZ group, on the other hand, the first MASI-VR training session ("S1" in Figure 5) was conducted immediately after the pretest evaluation. MASI-VR training sessions typically lasted approximately 20 minutes. On a separate day in the same week as the first training session, participants in the SZ group returned to the lab to complete the second training session ("S2" in Figure 5). In the second week, participants in the SZ group completed—on separate days—the third and fourth training sessions, and in the third week completed—again, on separate days—the fifth training session and ERRATA post-test session. The VR scenes used in the training session were presented in a repeating albeit brief cycle beginning with the grocery store scene, followed by the cafeteria scene, and ending with the bus stop scene. An ERRATA posttest session, using a configuration distinct from that used in the pretest, was conducted at the final visit of the study. In all, the study comprised six visits over three weeks.

5 RESULTS

5.1 Data Analysis

Two categories of analyses are presented. In Section 5.2, we present a comparison of the control group (n=10) and SZ group (n=10) at baseline evaluation. In Section 5.3, we provide a preliminary assessment of the utility of the novel VR system as a tool for social skills intervention in individuals with SZ (n=9; one SZ participant dropped out of the study following the baseline procedures). Mixed variable ANOVAs were used to evaluate group differences and training effects with relevant factors including group (control and SZ), emotion type (joy, surprise, sadness, anger, fear, contempt, and disgust), intensity of displayed emotion (low, medium, and extreme), and time (pretest and posttest). For all such analyses, relevant descriptive statistics, *F*-statistics, degrees of freedom, *p*-values, and effect sizes are reported. Post-hoc Tukey tests were used to perform multiple comparisons following ANOVA procedures and post-hoc analyses are considered significant at the $\alpha<.05$ level. All statistical analyses were conducted using the MATLAB computing environment and gaze metrics were computed using the EyeMMV toolkit for MATLAB [Krassanakis et al. 2014]. Additional non-parametric tests were used to gauge the effects of various factors on recognition of fearful faces due to the specific impairment in recognition of fear in SZ [Kohler, et al. 2003; Morris, et al. 2009].

5.2 Baseline Comparison of Control and SZ Groups

5.2.1 Emotion Recognition Accuracy—A 2 (group) × 3 (intensity of displayed emotion) × 7 (emotion type) mixed ANOVA design was used to assess the effects of several important variables on emotion recognition accuracy. The inclusion of intensity as a factor is supported by literature highlighting the significant effects of intensity on emotion recognition accuracy in a variety of populations [Kohler, et al. 2003]. Similarly, the literature provides evidence of associations between emotion recognition and emotional valence (i.e., positive versus negative) [Comparelli, et al. 2013; Kohler, et al. 2003], warranting the use of emotion type as a factor.

Statistically significant main effects were observed for both intensity of displayed emotion, $F(2,390)=26.41$, $p<.001$, $\eta^2=0.07$, and emotion type, $F(6,390)=40.64$, $p<.001$, $\eta^2=0.32$, but not for group, $F(1,390)=0.58$, $p=.449$, $\eta^2=0$. However, accuracy was nominally higher in the control group (% correct M=57.55, SD=8.17) in comparison to the SZ group (M=52.87, SD=6.26). Note that while these accuracies may seem generally low, they are not uncharacteristic of levels reported in the literature, especially in light of the inclusion of emotions with low intensity and a wide range of negative valence emotions [Kohler, et al. 2003; Tottenham, et al. 2009]. Post-hoc tests revealed that recognition accuracies for faces showing joy or surprise were significantly greater than for all other emotion types. No significant interactions involving group were observed. Tables 2 and 3 present confusion matrices detailing the proportion of correctly identified emotions at all intensity levels for the control and SZ groups, respectively. As shown in these tables, the between-group differences in correctly identified emotions are relatively small except in the case of fear (i.e., a difference of 14.2%), which, although non-significant, is in agreement with the extant literature [Kohler, et al. 2003]. In addition, accuracy of recognition on the *low* intensity emotions (% correct M=40.89, SD=7.64) was significantly lower than for both medium (M=58.89, SD=7.78) and extreme (M=59.82, SD=6.54) intensities.

5.2.2 Self-reported Confidence and Deliberation Time in Emotion Selection— A two-way ANOVA was conducted to explore the potential effects of group and intensity of displayed emotion on self-reported confidence in emotion selection. A statistically significant main effect of intensity was found, $F(2,54)=8.5$, $p<.001$, $\eta^2=0.24$, but there was no group difference, $F(1,54)=0.69$, $p=.411$, $\eta^2=0.01$. There was also no group by intensity interaction, $F(2,54)=0.03$, $p=.974$, $\eta^2=0$. As expected, post-hoc analysis revealed that overall confidence was significantly lower for low intensity emotions (M=6.99, SD=1.28) compared to both medium (M=8.01, SD=1.08) and extreme (M=8.44, SD=1.0) intensity emotions. In similar fashion, a two-way ANOVA was conducted on the same sample to evaluate the effects of group and intensity of displayed emotion on deliberation time (in seconds) for emotion selection. Contrary to our expectations, deliberation time did not vary according to emotional intensity. Results indicate a significant main effect of group, $F(1,54)=5.96$, $p=0.18$, $\eta^2=0.09$, but not of intensity, $F(2,54)=1.37$, $p=.263$, $\eta^2=0.04$, nor of a group by intensity interaction, $F(2,54)=0.55$, $p=.579$, $\eta^2=0.02$. Post-hoc analysis showed that participants in the control group spent significantly less time making selections (M=7.5, SD=2.15) than participants in the SZ group (M=9.2, SD=2.79).

5.2.3 Gaze Metrics—The EyeMMV toolbox was used to identify fixations in time series gaze data. EyeMMV uses a 2-stage dispersion-based algorithm—regarded as one of the preferred algorithms in comparison to velocity- and area-based alternatives [Salvucci and Goldberg 2000]—to first separate fixations from saccades and, second, to optimize identified fixation centroids by excluding outlying and erroneous gaze positions [Krassanakis, et al. 2014]. It should be noted that fixation parameters are both user- and task-dependent, and therefore thresholds for fixation identification should be carefully selected [Mould et al. 2012]. EyeMMV uses three parameters to compute fixations: *t1* is a spatial parameter used for initial separation of fixation clusters, *t2* is a second spatial parameter used to remove erroneous data points, and *min duration* is a temporal parameter used to specify the shortest duration for consideration as a valid fixation. In this work, *t1* was set to approximately 1 degree of visual field [Quirk and Strauss 2001; Williams et al. 1999], *t2* was set automatically by EyeMMV using a standard deviation method, and *min duration* was conservatively set at 200 ms [Delerue, et al. 2010; Williams, et al. 1999]. After identifying all fixations, lists of fixations were reduced so as to include only those relevant to the presentation of neutral and emotionally expressive faces (i.e., excluding fixations collected during other task activities). Note that fixations are exclusively analyzed as they are much better at characterizing patterns of visual attention than saccades [Mould et al. 2012].

From the procedures described above, measures of total number of fixations and mean fixation duration were obtained for each participant. Similar to other analyses in the literature [Delerue, et al. 2010; Loughland, et al. 2002; Williams et al. 2003], we evaluated the effects of group (i.e., control and SZ) and expression condition (i.e., neutral and emotionally expressive) on these measures. Data points were excluded from analysis if the number of fixations was less than the total number of trials presented (i.e., 84); using this criterion, only one control subject was excluded from analysis due to poor tracking of the eyes. A two-way ANOVA revealed a statistically significant main effect of group for total number of fixations, $F(1,34)=5.7$, $p=.023$, $\eta^2=0.14$. As expected, post-hoc analysis showed that the SZ group demonstrated significantly fewer fixations (M=131.95, SD=54.53) than the control group (M=174.28, SD=51.35). No significant effects were observed for expression condition, $F(1,34)=0$, $p=.961$, $\eta^2=0$, nor for expression condition by group interaction, $F(1,34)=0$, $p=.974$, $\eta^2=0$. With regards to mean fixation duration, no statistically significant effects were observed. Though not statistically significant in the current work, $F(1,34)=2.91$, $p=.097$, $\eta^2=.08$, with greater statistical power, we expect that there will be a main effect of expression condition for both groups based on the extant literature (i.e., longer fixations on neutral faces compared to emotionally expressive faces [Delerue, et al. 2010; Loughland, et al. 2002; Quirk and Strauss 2001]). Lastly, Figure 6 gives a qualitative comparison of visual attention by group and expression condition for a male child face displaying medium-intensity joy; the group disparity visible in this figure recalls the stark contrast in scanpath patterns between SZ and comparison subjects reported in previous studies (e.g., [Loughland, et al. 2002]). Specifically, the gaze heatmaps depict patients' aberrant focus on potentially less socially-informative facial regions, such as decreased attention to the eyes.

5.2.4 Follow-up Analyses Concerning Fearful Faces—

Because of the special status of impaired recognition of fear in individuals with SZ [Kohler, et al. 2003; Morris, et al. 2009], we decided to conduct follow-up analyses comparing the two groups specifically on the recognition of fear. We used a non-parametric approach because of the observed skewness of the independent variables. A Mann-Whitney *U*-test revealed that recognition of medium-intensity fear by the SZ group (% correct, *Median*=37.5) was significantly lower than in the control group (*Median*=75, $U=132$, $p=.040$), which is in agreement with the literature [Kohler, et al. 2003]. No statistically significant effects related specifically to fearful faces were observed with regards to either confidence, deliberation time, or fixation duration.

5.3 Preliminary Assessment of Training Effects

5.3.1 Changes Following Training with MASI-VR—Preliminary training effects were evaluated using a variety of mixed ANOVAs, all of which included a time factor with levels *pretest* and *posttest*. One participant in the SZ group dropped out of the study following the baseline procedures, thus data are reported for n=9 participants. Additionally, gaze data from one participant in the posttest were excluded due to poor acquisition using the same criterion as described in Section 5.2.3. Table 4 presents the results of these analyses for each of the recorded measures. Although significant effects were observed, none were observed across time, likely due in part to the limited power of this sample. Consistent with results at baseline evaluation, a significant effect of emotion intensity was observed with post-hoc analysis revealing that low intensity emotions were recognized with significantly less accuracy (M=41.73%, SD=6.74%) than both medium (M=57.09%, SD=9.68%) and extreme (M=62.56%, SD=7.75%) intensity emotions. Relatedly, self-reported confidence in emotion selection was significantly greater for extreme intensity emotions (M=8.5, SD=1.2) than for low intensity emotions (M=7.19, SD=1.87). A significant effect was observed with regards to the number of fixations on specified regions of the face; overall, the number of fixations on the eyes (M=115.5, SD=84.69) was significantly greater than for the nose (M=66.78, SD=44.72) and mouth (M=23.06, SD=44.72).

5.3.2 Follow-up Analyses Concerning Fearful Faces—As in Section 5.2.4, follow-up analyses of changes in performance from pretest to posttest were conducted with respect to fear using a non-parametric approach. A paired-sample Wilcoxon signed-rank test revealed that recognition of fear was significantly higher at posttest (% correct, *Median*=58.33) than at pretest (*Median*=50, *Z*=−2.21, *p*=.027). This represents an overall improvement in the recognition of fear by 16.67% post-training. No statistically significant fear-specific differences between pre- and posttest measurements were found for confidence, deliberation time, or fixation duration.

6 DISCUSSION AND CONCLUSION

We achieved our primary goal of completing a feasibility study of individuals with and without SZ using two novels systems—ERRATA for emotion recognition measurement and MASI-VR for social skills improvement via a previously untested gaze-sensitive training methodology. Many of our hypotheses were supported by the preliminary results which are described in detail below.

A pilot test of the novel emotion recognition system revealed a number of promising results supporting the feasibility of ERRATA as a tool for the collection of performance and gaze measures during tasks of emotion presentation. Although analyses using ANOVA did not reveal a significant main effect of group with regards to emotion recognition accuracy—likely attributable to limited statistical power—the largest disparity between the SZ and control groups was in the recognition of fear, which is in firm agreement with the extant literature [Kohler, et al. 2003; Morris, et al. 2009]. However, a follow-up, non-parametric analysis did reveal a significant group difference with respect to recognition accuracy for medium-intensity fearful faces. Specific attention to individual emotion types, as well as to emotional valence more generally, is justified by the evidence to date, which points to complex rather than monolithic differences between individuals with and without SZ on measures of emotion recognition [Comparelli, et al. 2013; Comparelli, et al. 2014; Kohler, et al. 2003].

Significant effects of intensity and emotion type were also observed. As expected, higher intensity emotions were more accurately recognized than lower intensity emotions and specific emotions such as joy and sadness were correctly identified more often than anger and fear—results that replicate results from past literature [Kohler, et al. 2003; Morris, et al. 2009]. As expected, confidence in emotion selection across all participants was significantly greater for higher intensity emotions than for lower intensity emotions, but no group difference was seen. Deliberation time in emotion selection, however, was significantly longer in the SZ group compared to the control group, which seems to be in agreement with a need for additional processing time (and perhaps greater effort) in order to identify presented emotions in people with SZ. With respect to gaze metrics, a significant effect of group was observed for total number of fixations; consistent with the literature, participants in the SZ group demonstrated significantly fewer fixations than participants in the control group [Bekele, et al. 2017; Delerue, et al. 2010; Loughland, et al. 2002]. A significant effect of group was not observed for mean fixation duration, but we expect that such an effect would be present with a larger sample and thus remains a task for future work.

Next, we conducted a preliminary evaluation of the intervention by exposing individuals in the SZ group to sessions of MASI-VR over a period of five visits while using ERRATA to collect pre- and posttest measures of accuracy and visual attention. The sample was unfortunately affected by the dropout of one participant, thus diminishing statistical power. Encouragingly, however, significant effects still emerged despite this limitation. Although the overall measure of emotion recognition accuracy improved only nominally post-training, recognition of fear in particular showed significant improvement, rising by 16.67% across all intensity levels. Consistent with the results of baseline evaluation, accuracy and confidence were again greatest for higher intensity emotions in comparison to lower intensity emotions (Table 4). With regards to gaze metrics, no significant effects of time were observed, although the nominal changes in means hint at potentially positive changes in visual attention. Specifically, an observed decrease in fixation duration and increase in number of fixations may indicate a movement towards patterns of attention more characteristic of neurotypical individuals [Delerue, et al. 2010; Loughland, et al. 2002], though additional work will be necessary to determine this. A significant effect of ROI was found for number of fixations and further analysis showed that, consistent with other studies [Delerue, et al. 2010], eyes were viewed with greater frequency than the nose and mouth across all assessments.

Based on the success of comparatively low-tech solutions (e.g., [Russell, et al. 2008]) in improving emotion recognition skills, we are confident that the novel approach of the current work is capable of improving this skill. The system was well-tolerated by individuals with SZ and our preliminary results show that the recognition of fear improved considerably post-training. In all, we believe that the feasibility of the novel approach has been shown, thus warranting further investigation with an adequately-powered sample. While this preliminary work yielded encouraging results, a few primary limitations must be addressed in future research. First, while the sample size evaluated (i.e., N=20; 10 neurotypical controls, and 10 individuals with SZ evaluated at two different time points) is consistent with the scope of a feasibility study to assess the usability of a gaze-sensitive training paradigm, a larger clinical study will be required to robustly gauge the utility of the novel approach. Second, the period of exposure to the intervention system may need to be extended in future work. We trained participants with SZ at five visits over the course of three weeks, but more exposure will likely be warranted; a period of exposure of 5–10 weeks or longer is common in the literature [Tan, et al. 2016]. Third, fixation means and counts were used exclusively to gauge patterns of visual attention because fixations are much more explanatory of attention in comparison to saccades which are often discarded from analysis [Mould et al. 2012]; however, future work should consider additional measures of attention in order to obtain a clearer picture of attention patterns in this population. Finally, as with other studies aiming to address the issue of social skills training in SZ, the ecological validity of the system must be shown in order to justify use of the novel approach. For example, despite observed improvements in the recognition of fearful faces within the MASI-VR system, future work should compare these effects with changes on an established social functioning measure or scale Bell Lysaker Emotion Recognition Task [Pinkham, et al. 2015].

Based on the success of this pilot work, we will follow-up with an extended longitudinal study while also employing a larger, adequately powered recruitment sample of adults with SZ. Importantly, lessons learned from this pilot evaluation may help to optimize the task content in future evaluations. For example, because people with SZ appear to shown evidence of competent performance with regards to the recognition of positive valence emotions, future deployments of the training system may be modified to focus only on the emotion types and intensities that are most problematic for individuals with SZ. With this enhanced study design, we will be able to determine the efficacy of the proposed gaze-sensitive training paradigm, ultimately seeking to deploy a maximally effective intervention tool for use by researchers and clinicians.

ACKNOWLEDGEMENT

We acknowledge funding support from National Institute of Health grants 5R21 MH106748–02 and 1R01 MH091102–01A1 and National Science Foundation grant 0967170. The authors wish to thank all of the participants for their commitment of time to this research. We also thank Park Lab members Laura Adery and Lénie Torregrossa, as well as Alena Gizdic for continuous support in conducting experiments.

REFERENCES

Achtman R, Green C and Bavelier D 2008 Video games as a tool to train visual skills. Restorative Neurology and Neuroscience 26, 435–446. [PubMed: 18997318]

Andreasen NC 1984 Scale for the assessment of positive symptoms (SAPS). University of Iowa Iowa City.

Andreasen NC 1989 Scale for the Assessment of Negative Symptoms (SANS). The British Journal of Psychiatry.

Bal E, Harden E, Lamb D, Van Hecke AV, Denver JW and Porges SW 2010 Emotion recognition in children with autism spectrum disorders: Relations to eye gaze and autonomic state. Journal of Autism and Developmental Disorders 40, 358–370. [PubMed: 19885725]

Bekele E, Bian D, Peterman J, Park S and Sarkar N 2017 Design of a Virtual Reality System for Affect Analysis in Facial Expressions (VR-SAAFE); Application to Schizophrenia. IEEE Transactions on Neural Systems and Rehabilitation Engineering 25, 739–749. [PubMed: 27429438]

Bekele E, Crittendon J, Zheng Z, Swanson A, Weitlauf A, Warren Z and Sarkar N 2014 Assessing the utility of a virtual environment for enhancing facial affect recognition in adolescents with autism. Journal of Autism and Developmental Disorders 44, 1641–1650. [PubMed: 24419871]

Bekele E, Wade J, Bian D, Fan J, Swanson A, Warren Z and Sarkar N 2016 Multimodal adaptive social interaction in virtual environment (MASI-VR) for children with Autism spectrum disorders (ASD). In Proceedings of Virtual Reality (VR) IEEE, 121–130.

Bellack AS, Green MF, Cook JA, Fenton W, Harvey PD, Heaton RK, Laughren T, Leon AC, Mayo DJ and Patrick DL 2006 Assessment of community functioning in people with schizophrenia and other severe mental illnesses: a white paper based on an NIMH-sponsored workshop. Schizophrenia Bulletin 33, 805–822. [PubMed: 16931542]

Biehl M, Matsumoto D, Ekman P, Hearn V, Heider K, Kudoh T and Ton V 1997 Matsumoto and Ekman’s Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences. Journal of Nonverbal behavior 21, 3–21.

Bornhofen C and McDonald S 2008 Emotion perception deficits following traumatic brain injury: A review of the evidence and rationale for intervention. Journal of the International Neuropsychological Society 14, 511–525. [PubMed: 18577280]

Bruce C, Unsworth C, Dillon M, Tay R, Falkmer T, Bird P and Carey L 2017 Hazard perception skills of young drivers with Attention Deficit Hyperactivity Disorder (ADHD) can be improved with computer based driver training: an exploratory randomised controlled trial. Accident Analysis & Prevention 109, 70–77. [PubMed: 29040873]

Chukoskie L, Westerfield M and Townsend J 2017 A novel approach to training attention and gaze in ASD: A feasibility and efficacy pilot study Developmental Neurobiology.

Comparelli A, Corigliano V, De Carolis A, Mancinelli I, Trovini G, Ottavi G, Dehning J, Tatarelli R, Brugnoli R and Girardi P 2013 Emotion recognition impairment is present early and is stable throughout the course of schizophrenia. Schizophrenia Research 143, 65–69. [PubMed: 23218561]

Comparelli A, De Carolis A, Corigliano V, Di Pietro S, Trovini G, Granese C, Romano S, Serata D, Ferracuti S and Girardi P 2014 Symptom correlates of facial emotion recognition impairment in schizophrenia. Psychopathology 47, 65–70. [PubMed: 23796958]

Couture SM, Granholm EL and Fish SC 2011 A path model investigation of neurocognition, theory of mind, social competence, negative symptoms and real-world functioning in schizophrenia. Schizophrenia Research 125, 152–160. [PubMed: 20965699]

Delerue C, Laprévote V, Verfaillie K and Boucart M 2010 Gaze control during face exploration in schizophrenia. Neuroscience Letters 482, 245–249. [PubMed: 20667499]

Drusch K, Stroth S, Kamp D, Frommann N and Wölwer W 2014 Effects of Training of Affect Recognition on the recognition and visual exploration of emotional faces in schizophrenia. Schizophrenia Research 159, 485–490. [PubMed: 25248938]

Ekman P 1993 Facial expression and emotion. American Psychologist 48, 384. [PubMed: 8512154]

Fletcher L and Zelinsky A 2009 Driver Inattention Detection based on Eye Gaze—Road Event Correlation. The International Journal of Robotics Research 28, 774–801.

Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B and Slater M 2017 Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychological Medicine 47, 2393–2400. [PubMed: 28325167]

Green MF, Penn DL, Bentall R, Carpenter WT, Gaebel W, Gur RC, Kring AM, Park S, Silverstein SM and Heinssen R 2008 Social cognition in schizophrenia: an NIMH workshop on definitions, assessment, and research opportunities. Schizophrenia Bulletin 34, 1211–1220. [PubMed: 18184635]

Hills PJ, Eaton E and Pake JM 2016 Correlations between psychometric schizotypy, scan path length, fixations on the eyes and face recognition. The Quarterly Journal of Experimental Psychology 69, 611–625. [PubMed: 25835241]

Hoffmann H, Kessler H, Eppel T, Rukavina S and Traue HC 2010 Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. Acta Psychologica 135, 278–283. [PubMed: 20728864]

Kohler CG, Turner TH, Bilker WB, Brensinger CM, Siegel SJ, Kanes SJ, Gur RE and Gur RC 2003 Facial emotion recognition in schizophrenia: intensity effects and error pattern. American Journal of Psychiatry 160, 1768–1774. [PubMed: 14514489]

Krassanakis V, Filippakopoulou V and Nakos B 2014 EyeMMV toolbox: An eye movement post-analysis tool based on a two-step spatial dispersion threshold for fixation identification. Journal of Eye Movement Research 7.

Loughland CM, Williams LM and Gordon E 2002 Visual scanpaths to positive and negative facial emotions in an outpatient schizophrenia sample. Schizophrenia Research 55, 159–170. [PubMed: 11955975]

Morris RW, Weickert CS and Loughland CM 2009 Emotional face processing in schizophrenia. Current Opinion in Psychiatry 22, 140–146. [PubMed: 19553867]

Mould MS, Foster DH, Amano K and Oakley JP 2012 A simple nonparametric method for classifying eye fixations. Vision Research 57, 18–25. [PubMed: 22227608]

Öhman A, Flykt A and Esteves F 2001 Emotion drives attention: detecting the snake in the grass. Journal of Experimental Psychology: General 130, 466. [PubMed: 11561921]

Park K-M, Ku J, Choi S-H, Jang H-J, Park J-Y, Kim SI and Kim J-J 2011 A virtual reality application in role-plays of social skills training for schizophrenia: a randomized, controlled trial. Psychiatry Research 189, 166–172. [PubMed: 21529970]

Pinkham AE, Penn DL, Green MF and Harvey PD 2015 Social cognition psychometric evaluation: results of the initial psychometric study. Schizophrenia Bulletin 42, 494–504. [PubMed: 25943125]

Quirk SW and Strauss ME 2001 Visual exploration of emotion eliciting images by patients with schizophrenia. The Journal of Nervous and Mental Disease 189, 757–765. [PubMed: 11758659]

Revheim N, Butler PD, Schechter I, Jalbrzikowski M, Silipo G and Javitt DC 2006 Reading impairment and visual processing deficits in schizophrenia. Schizophrenia Research 87, 238–245. [PubMed: 16890409]

Rus-Calafell M, Gutiérrez-Maldonado J and Ribas-Sabaté J 2014 A virtual reality-integrated program for improving social skills in patients with schizophrenia: a pilot study. Journal of Behavior Therapy and Experimental Psychiatry 45, 81–89. [PubMed: 24063993]

Russell TA, Green MJ, Simpson I and Coltheart M 2008 Remediation of facial emotion perception in schizophrenia: concomitant changes in visual attention. Schizophrenia Research 103, 248–256. [PubMed: 18565733]

Sachse M, Schlitt S, Hainz D, Ciaramidaro A, Walter H, Poustka F, Bölte S and Freitag CM 2014 Facial emotion recognition in paranoid schizophrenia and autism spectrum disorder. Schizophrenia Research 159, 509–514. [PubMed: 25278104]

Salvucci DD and Goldberg JH 2000 Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the 2000 Symposium on Eye Tracking Research & Applications ACM, 71–78.

Stano J 1999 Wechsler abbreviated scale of intelligence. San Antonio, TX: The Psychological Corporation.

Stichter JP, Laffey J, Galyen K and Herzog M 2014 iSocial: Delivering the social competence intervention for adolescents with high functioning autism. Journal of Autism and Developmental Disorders 44, 417–430. [PubMed: 23812663]

Suzuki M, Takahashi S, Matsushima E, Tsunoda M, Kurachi M, Okada T, Hayashi T, Ishii Y, Morita K and Maeda H 2009 Exploratory eye movement dysfunction as a discriminator for schizophrenia. European Archives of Psychiatry and Clinical Neuroscience 259, 186–194. [PubMed: 19165524]

Tan B-L, Lee S-A and Lee J 2016 Social cognitive interventions for people with schizophrenia: a systematic review. Asian Journal of Psychiatry 35, 115–131. [PubMed: 27670776]

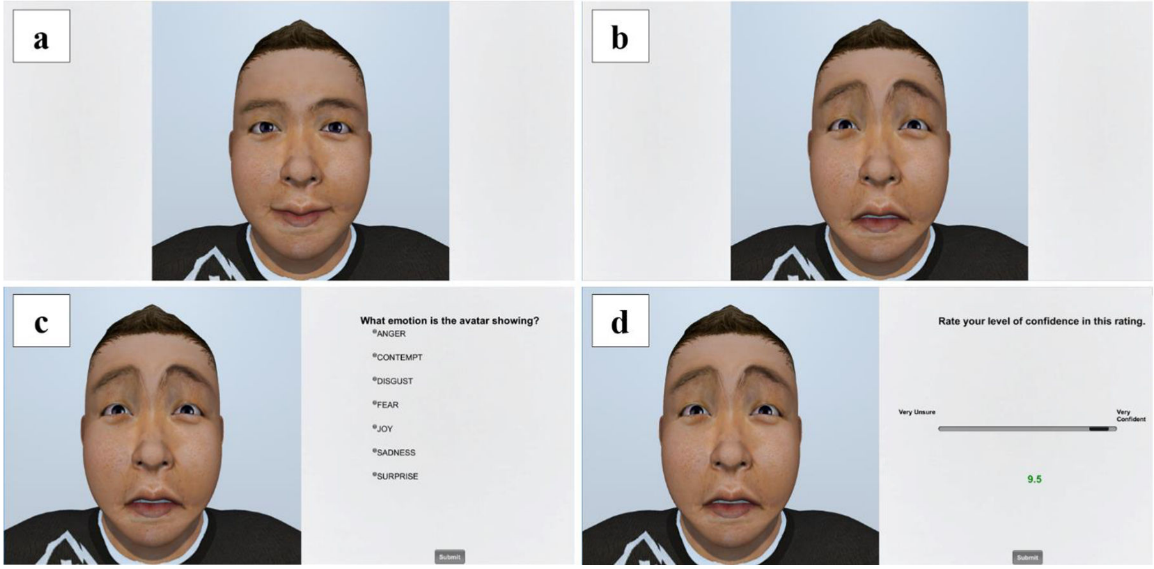
Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, Marcus DJ, Westerlund A, Casey B and Nelson C 2009 The NimStim set of facial expressions: judgments from untrained research participants. Psychiatry Research 168, 242–249. [PubMed: 19564050]

Van Erp TG, Preda A, Nguyen D, Faziola L, Turner J, Bustillo J, Belger A, Lim KO, Mcewen S and Voyvodic J 2014 Converting positive and negative symptom scores between PANSS and SAPS/SANS. Schizophrenia Research 152, 289–294. [PubMed: 24332632]

Veling W, Pot-Kolder R, Counotte J, Van Os J and Van Der Gaag M 2016 Environmental social stress, paranoia and psychosis liability: a virtual reality study. Schizophrenia Bulletin 42, 1363–1371. [PubMed: 27038469]

Williams LM, Loughland CM, Gordon E and Davidson D 1999 Visual scanpaths in schizophrenia: is there a deficit in face recognition? Schizophrenia Research 40, 189–199. [PubMed: 10638857]

Williams LM, Loughland CM, Green MJ, Harris AW and Gordon E 2003 Emotion perception in schizophrenia: an eye movement study comparing the effectiveness of risperidone vs. haloperidol. Psychiatry Research 120, 13–27. [PubMed: 14500110]



What emotion is the avatar showing?

○ANGER

○CONTEMPT

○DISGUST

○FEAR

○JOY

○SADNESS

○SURPRISE

Rate your level of confidence in this rating.

Fig. 1.

ERRATA task sequencing: (a) the non-expressive face of an avatar is shown for 2.5 seconds, (b) the emotionally expressive avatar face is shown—sadness in this case—for 2 seconds, (c) the user is prompted to identify which of the 7 emotions he/she believes is shown, and (d) the user is prompted to report confidence in his/her selection.

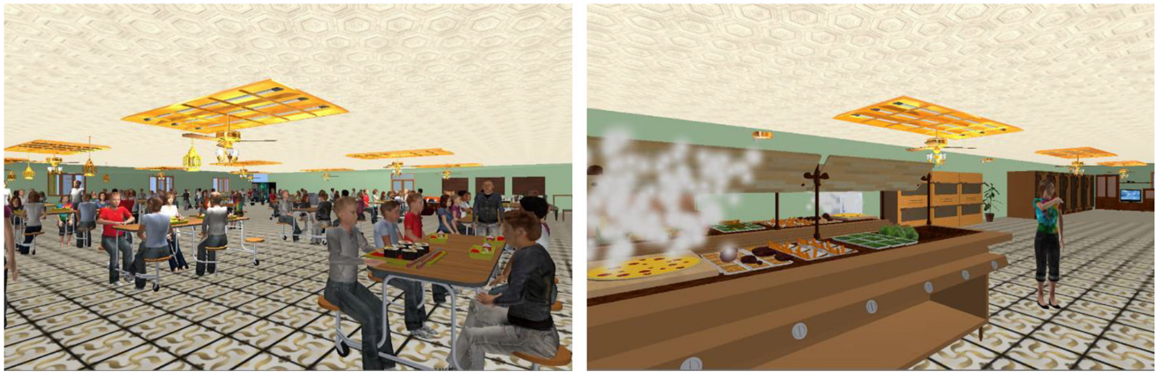


Fig. 2.
The cafeteria scene within MASI-VR.

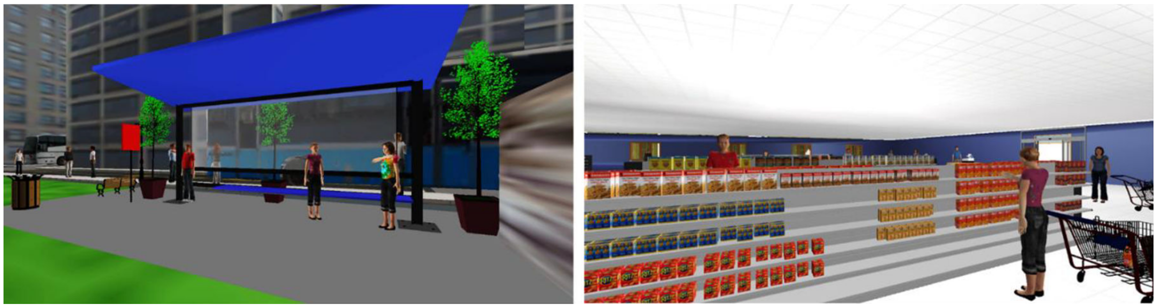
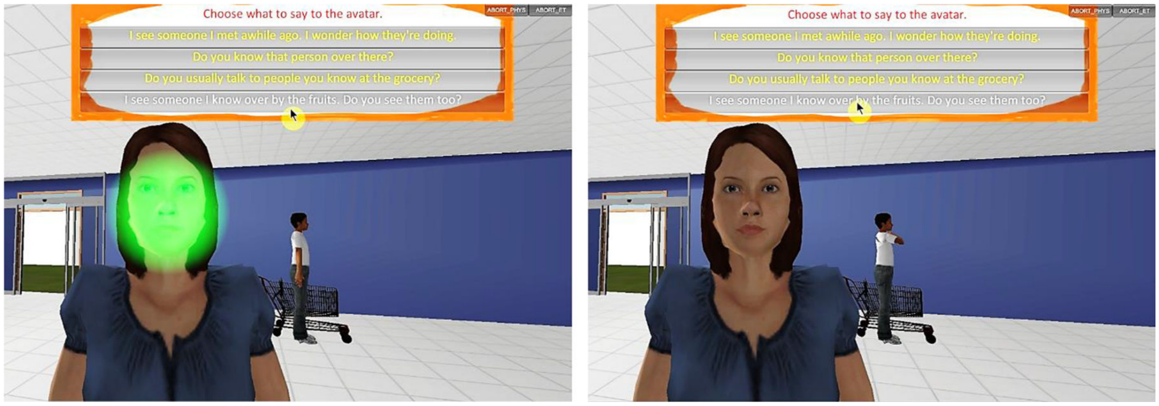


Figure 3.
The bus stop (left) and grocery store (right) scenes in MASI-VR.



Choose what to say to the avatar.

I see someone I met awhile ago. I wonder how they're doing.

Do you know that person over there?

Do you usually talk to people you know at the grocery?

I see someone I know over by the fruits. Do you see them too?

Choose what to say to the avatar.

I see someone I met awhile ago. I wonder how they're doing.

Do you know that person over there?

Do you usually talk to people you know at the grocery?

I see someone I know over by the fruits. Do you see them too?

Fig. 4.

MASI-VR communication interface. The transparent green mask is visible at the start of a conversation (left) until the avatar's face is observed at which point the mask is removed (right).

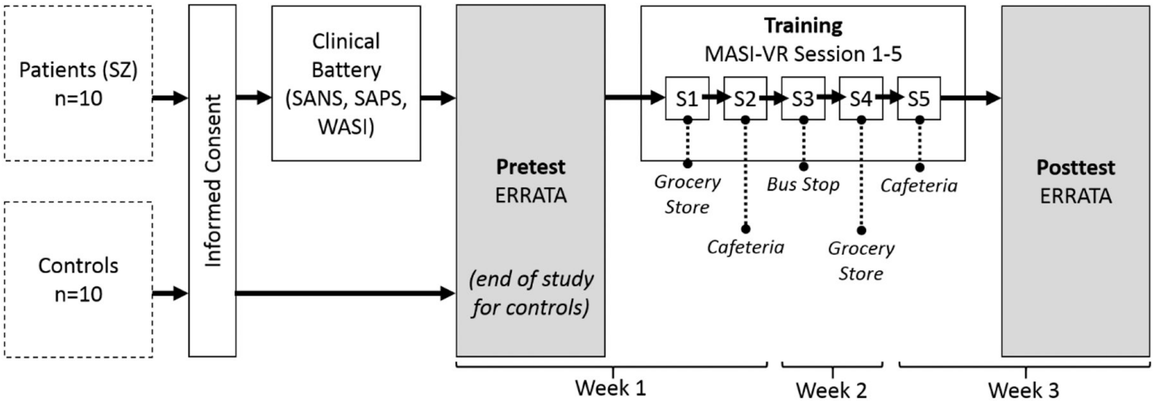


Fig. 5.
Study Procedures

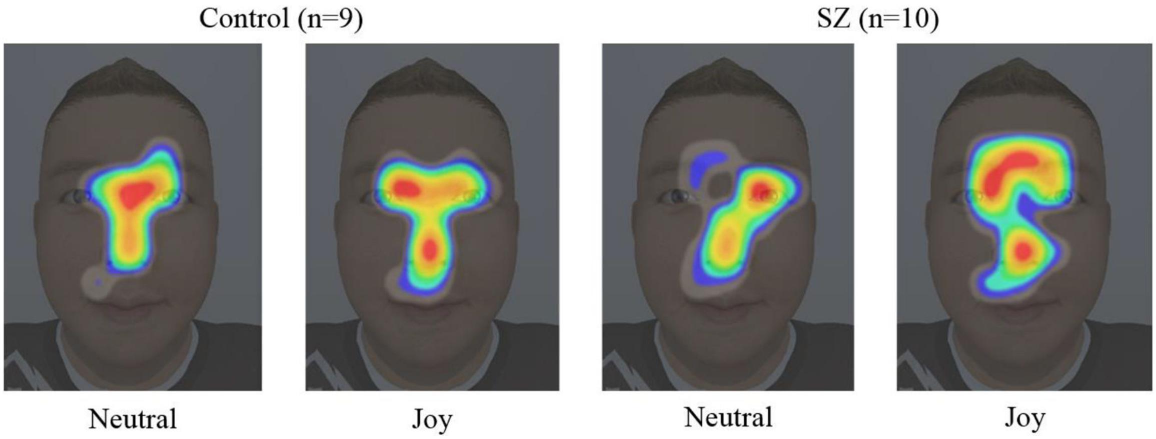


Fig. 6.

Gaze heatmaps for male child showing medium level of joy. Heatmaps were generated using the EyeMMV toolbox function “heatmap_generator” with parameters *grid spacing* = 75 (pixels), *kernel size for Gaussian filtering* = 15, and *sigma for Gaussian filtering* = 5.

Table 1.

Participant Characteristics

Group M (SD)		
	Control	**SZ**
	(n=10)	**(n=10)**
Gender (% female)	50%	40%
Age (years)[a]	43.2 (10.18)	45.2 (6.05)
Years of Education[b]	15.4 (2.12)	12.7 (1.42)
WASI Full-scale IQ[c]	104.78 (12.65)	95.7 (8.65)
WASI Verbal IQ[d]	104.44 (9.9)	94 (10.38)
WASI Performance IQ[e]	103.89 (15.61)	98.1 (7.80)
SANS[f]	-	33.5 (15.46)
SAPS[g]	-	22.1 (17.23)
Duration of Condition (years)	-	24.78 (9.72)

[a] No between-group difference in age, *t*(18)=−0.53, *p*=.60.

[b] Participants in the control group had significantly more years of education than participants in the SZ group, *t*(18)=3.35, *p*=.004.

[c] *WASI* Wechsler Abbreviated Scale of Intelligence; No between-group difference on full-scale IQ, *t*(17)=1.84, *p*=.08.

[d] Verbal IQ component of WASI; Participants in the control group had significantly higher verbal IQ than participants in the SZ group, *t*(17)=2.24, *p*=.039.

[e] Performance IQ component of WASI; No between-group difference on performance IQ, *t*(17)=1.04, *p*=.313.

[f] *SANS* Scale for the Assessment of Negative Symptoms

[g] *SAPS* Scale for the Assessment of Positive Symptoms

Table 2.

Confusion Matrix: Control Group Emotion Recognition Accuracy (%) (All Intensities)

	Joy	Sadness	Surprise	Fear	Anger	Contempt	Disgust
Joy	75.8	1.7	8.3	3.3	0	10	0.8
Sadness	1.7	55	8.3	13.3	0.8	10	10.8
Surprise	8.3	0.8	80.8	8.3	0.8	0.8	0
Fear	0.8	2.5	30	57.5	1.7	3.3	4.2
Anger	4.5	21.8	0.9	2.7	36.4	21.8	11.8
Contempt	15	5	15.8	0	0.8	35.8	27.5
Disgust	5.8	4.2	3.3	5	35	11.7	35

Table 3.

Confusion Matrix: SZ Group Emotion Recognition Accuracy (%) (All Intensities)

	Joy	Sadness	Surprise	Fear	Anger	Contempt	Disgust
Joy	81.7	0	9.2	1.7	0	6.7	0.8
Sadness	2.5	58.3	10.8	7.5	0.8	10	10
Surprise	14.2	0	78.3	5	0	1.7	0.8
Fear	0	3.3	33.3	43.3	0	8.3	11.7
Anger	8.2	8.2	2.7	0.9	32.7	31.8	15.5
Contempt	27.5	0	3.3	1.7	0	40	27.5
Disgust	10	2.5	2.5	5.8	35.8	13.3	30

Table 4.

Preliminary Assessment of Training Effects

Response Variable	Pretest		Posttest		Mixed ANOVA Statistics			
	M	SD	M	SD	Effect	*F*-statistic	*p*	η^2
***Performance Measures** (based on n=10 at pretest and n=9 at posttest)*								
Emotion Recognition Accuracy (%)	52.29	3.78	55.29	5.76	Time Emotion Intensity[a] Time × Emotion Intensity	*F*(1,51)=2.01 *F*(2,51)=33.38 *F*(2,51)=0.82	.162 <.001*** .447	0.02 0.55 0.01
Confidence in Emotion Selection (0–10 scale)	7.93	1.16	7.86	1.82	Time Emotion Intensity[b] Time × Emotion Intensity	*F*(1,51)=0.02 *F*(2,51)=3.30 *F*(2,51)=0.12	.880 .045* .891	0 0.11 0
Deliberation Time (sec)	9.20	2.79	8.98	2.77	Time Emotion Intensity Time × Emotion Intensity	*F*(1,51)=0.08 *F*(2,51)=0.57 *F*(2,51)=0.15	.781 .567 .865	0 0.02 0.01
***Gaze Measures** (based on n=10 at pretest and n=8 at posttest)*								
Fixation Duration by Neutral v. Emotional face (ms)	375.5	60.11	373.9	54.36	Time Emotion Condition Time × Emotion Condition	*F*(1.32)=0.01 *F*(1,32)=2.00 *F*(1,32)=0.04	.933 .167 .835	0 0.06 0
Number of Fixations by Neutral v. Emotional face	132	54.53	153.4	81.45	Time Emotion Condition Time × Emotion Condition	*F*(1,32)=0.84 *F*(1,32)=0.11 *F*(1,32)=0.11	.366 .737 .747	0.03 0 0
Fixation Duration by ROI (ms)	353.1	103.9	331.3	122.4	Time ROI Time × ROI	*F*(1,64)=0.67 *F*(3,64)=1.82 *F*(3,64)=0.29	.416 .153 .832	0.01 0.08 0.01
Number of Fixations by ROI	65.98	55.75	76.69	70.35	Time ROI[c] Time × ROI	*F*(1,64)=0.68 *F*(3,64)=8.74 *F*(3,64)=0.44	.413 <.001*** .727	0.01 0.29 0.01

Time: pretest, posttest; **Emotion Intensity**: low, medium, extreme; **Emotion Condition**: neutral, expressive; **ROI**: eyes, nose, mouth, other face.

* p<.05,

** p<.01,

*** p<.001

[a] Low intensity emotions were recognized with significantly less accuracy (M=41.73%, SD=6.74%) than both medium (M=57.09%, SD=9.68%) and extreme (M=62.56%, SD=7.75%) intensity emotions.

[b] Self-reported confidence in emotion selection was significantly greater for extreme intensity emotions (M=8.5, SD=1.2) than for low intensity emotions (M=7.19,SD=1.87).

[c] Overall, the number of fixations on the eyes (M=115.5 ms, SD=84.69 ms) was significantly greater than for the nose (M=66.78 ms, SD=44.72 ms) and mouth (M=23.06 ms, SD=44.72 ms).