Computers in Human Behavior Reports 17 (2025) 100549

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

MiEmo: A multi-modal platform on emotion recognition for children with autism spectrum condition

Eleonora Chitti [a,*], Rossana Actis-Grosso [b], Paola Ricciardelli [b], Benedetta Olivari [c], Cecilia Carenzi [d], Mariantonia Tedoldi [d], N. Alberto Borghese [a]

[a] *Applied Intelligent Systems Laboratory- Ais-Lab, Department of Computer Science, University of Milan, Milan, Italy*
[b] *Department of Psychology, University of Milan-Bicocca, Milan, Italy*
[c] *Dosso Verde Rehabilitation Center, Milan, Italy*
[d] *Sacra Famiglia Rehabilitation Center, Cesano Boscone, Italy*

A R T I C L E I N F O

Keywords:
ASC
Autism intervention
Emotion recognition
Serious game
Exergame
Social skills
Multi-modal feedback

A B S T R A C T

Introduction: This study presents MiEmo, a multi-modal digital platform designed to improve emotion recognition in children with Autism Spectrum Condition (ASC). The platform integrates serious games with music and color as feedback mechanisms to strengthen emotion understanding in addition to traditional visual interventions such as pictures and videos. The study aims to assess the usability and potential effectiveness of MiEmo in supporting therapy for children with medium- and high-functioning ASC.
Methods: A pilot usability study was conducted in two rehabilitation centers involving 19 children, 8 with medium and 11 with high functioning ASC. Participants engaged with six training activities, or exergames, on the MiEmo platform that implemented multi-modal feedback (music pieces and colored animations associated to the emotion). The System Usability Scale (SUS) and qualitative feedback from therapists were used to evaluate the platform's usability.
Results: The average SUS scores were 86.88 for children with medium-functioning ASC and 96.75 for those with high-functioning ASC, indicating positive usability. Therapists noted that while the platform was well-received, further updates are needed for better adaptation to medium-functioning children. Multi-modal feedback, particularly music and color, was found to enhance emotion recognition, with children responding well to the integration of these sensory cues.
Conclusion: The study demonstrates that MiEmo has significant potential as a tool for socio-emotional training, particularly for high-functioning children. However, limitations such as the small sample size, short intervention duration, and lack of a control group suggest that future studies with larger participant groups are necessary to validate these findings and assess long-term effects.

1. Introduction

The primary aim of this study is to describe the design, implementation and field testing of a novel multi-modal intervention tool, MiEmo, to increase ASC children socio-emotional competences. At its core is an innovative multi-modal feedback and rendering, based on music and colors attuned to emotions, in addition to videos and pictures adopted by classical digital platforms, aimed to largely enhance emotions perception and feeling.

The capacity to effectively communicate emotions, and to recognize them in others, has a strong social value: it allows us to co-construct an emotional space to share information with others and to establish social links and relationships (Scherer, 2005). According to Darwin (Ekman, 2009), humans have developed emotions in the evolution process to better address adaptation problems by selecting the most appropriated reaction (fight or fly) upon interpreting others' emotions.

In this context, Theory of Mind (ToM) plays a fundamental role. It defines to the capacity to understand other people's intentions by ascribing mental and emotional states to them. ToM allows a person to understand if other people are positively approaching him/her or if they

* Corresponding author.
E-mail addresses: eleonora.chitti@unimi.it (E. Chitti), rossana.actis@unimib.it (R. Actis-Grosso), parola.ricciardelli@unimib.it (P. Ricciardelli), alberto.borghese@unimi.it (N.A. Borghese).

https://doi.org/10.1016/j.chbr.2024.100549
Received 19 August 2024; Received in revised form 21 November 2024; Accepted 29 November 2024

have bad intentions (Baron-Cohen et al., 1985). People with Autistic Spectrum Condition (ASC) often lack a ToM and have a reduced ability to recognize and understand other's emotional and mental states. Indeed, emotion competence is at the core of a successful interaction in life. Difficulties in processing emotions can cause in ASC children a reduced ability to show empathy in social contexts (Baron-Cohen et al., 1985), which can interfere with their learning and negatively impact the development of cognitive functions and social interactions with peers (Jaswal & Akhtar, 2019). This may lead to isolation and potential health issues, such as depression (Bauminger & Kasari, 2000).

As the prevalence of ASC diagnosis is rapidly increasing with a diagnosis of 1 in 36 children in United States in 2020 (Maenner et al., 2020), ASC treatment is becoming a high priority of Health Service Providers and a lot of effort has been put in designing effective habilitation protocols.

In the last years, leveraging the natural propension of children to games, several approaches based on serious games have been proposed to improve social skills of ASC children (Serret et al., 2014), (Fridenson-Hayo et al., 2017), (Zoerner et al., 2016). In fact, one of the difficulties in training ASC children is the stress that they may feel during therapy, caused by the direct interaction with the therapist. Digital applications attenuate this, since they shift attention from the therapist to the videogame, possibly played with the therapist on the side to guide the child (Mehlenbacher, 2010). Moreover (Baron-Cohen, 2006), demonstrated that ASC children feel comfortable to interact with a system that has a clear pattern of events that develops over time, and that is based on logical rules.

Such serious games (Tang et al., 2019) are based on pictures, videos and/or graphical animations aimed at eliciting emotions and are largely based on the Facial Action Coding Systems (FACS) (Ekman & V Friesen, 1978) that defines the face micro-movements associated to each emotion. In FACS, emotions are organized into discrete categories (Fig. 1), where primary emotions: joy, sadness, anger and fear, are particularly relevant as they have a survival value for human species. Positive and negative feedback is provided at the end of each game, through standard in-game feedback (Schell, 2019), mainly visual: text, animations, gamification.

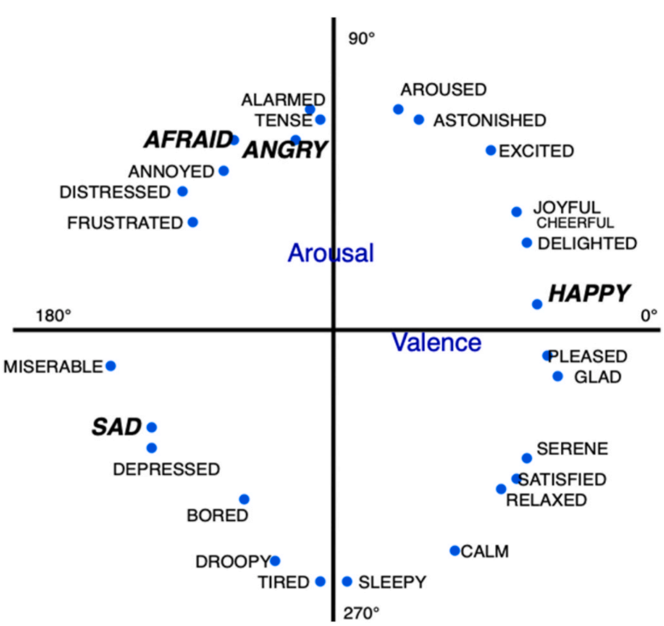

Fig. 1. Representation of emotions inside the 2D space of Valance and Arousal (Russell, 1980), where Valence describes the extent to which an emotion is positive or negative, and Arousal refers to its intensity. The discrete emotions provided by FACS are identified by labelled points: primary emotions of happiness, sadness fear and anger are in capital letters.

However, one of the greatest difficulties is engaging ASC children into therapy, as they tend to distract, and more powerful feedback may help capturing more their attention, that is critical for a successful therapy (Zhang et al., 2021). We resort to recent studies on music (Janzen & Thaut, 2018) and colors (Ismail et al., 2021) perception to improve ASC intervention.

Starting from Ekman's studies, Juslin and Laukka (Juslin & Laukka, 2004), (Juslin, 2019) have shown that similar acoustic patterns are shared by the expression of emotions in music and speech prosody (Koelsch, 2014). To derive a classification of music pieces according to discrete emotions, music is analyzed in terms of tempo variability, dissonance or consonant harmony, minor or major mode, sound level and sharpness, repetitions, number of instruments (Panda et al., 2023), (Cowen et al., 2020) (cf. Section 3.3.1). Recent studies have shown that ASC children do have a music emotional perception similar to neurotypical ones, and that they can easily associate emotional meaning to music pieces (Gebauer et al., 2014), (Quintin et al., 2011). Several studies have also demonstrated that music music-therapy could help to regulate emotions, to reduce stress and to leverage cooperation with peers in free play sessions (Ke et al., 2022). Moreover (Katagiri, 2009), showed that background music can enhance emotional understanding during socio-emotional therapy.

Colors have also been linked to emotions. The studies (Takahashi & Kawabata, 2018), (Nakajima et al., 2017) proposed an association between colors and emotions, that has been used in different fields: it has been explored in human-robot interaction to convey emotions in robotic agents (Casiddu et al., 2021) and it is largely used in educational research for children, in which different colors shown on book pages elicit different emotional reactions congruent with the current scene of the story read (Chang et al., 2018). The different colors are perceived by ASC and neurotypical children as associated to the same emotions (Franklin et al., 2008), although this association is less sharp. Indeed, ASC children benefit from colored overlays in their learning materials (Ludlow et al., 2008).

To the best of our knowledge, the emotional content of music and colors has not been explored yet in serious games for ASC socio-emotional training. We leverage music and color perception to provide here innovative multi-modal feedback inside the serious games to train socio-emotional competences by playing specific music pieces associated to each emotion synchronous with its representation through video, pictures and animation. We complement this also with a color change of the scene and of the game avatar, tuned to that emotion. This novel multi-modal feedback would provide a stronger feeling of the emotion on one side, and more engagement inside the activity on the other, that are key elements for a successful training.

The paper is organized as follow. In Section 2 state-of-the-art is analyzed. In Section 3 method is described, here co-design with ASC therapists is reported, and the guidelines derived are translated into digital activities implementing multi-modal feedback. Results on pilot testing are reported in Section 4 and discussed in Section 5. Conclusions are drawn in Section 6.

2. State of the art

ASC socio-emotional therapy is traditionally based on showing pictures of faces of an actor expressing an emotion and training the child to recognize that emotion (Zhang et al., 2021). Therapists use also stripes portraying stories by which emotions are elicited (Karkhaneh et al., 2010), and Picture Communication Symbol (PCS),[1] visual pictograms representing emotions with stylized faces that contain only eyebrows, eyes and mouth (c.f. Fig. 4). Training starts with the four primary emotions of anger, sadness, happiness and fear (Zhang et al., 2021) and we will focus upon these emotions in this work. The other primary

[1] Creative Commons PCS available at https://arasaac.org.

emotions of surprise and disgust are not considered here. ASC people process disgust in a different manner, since its expression deeply depends on motor reactions, altered in ASC people (Jayashankar & Aziz-Zadeh, 2023). The recognition of surprise needs capability of understanding misbeliefs, often lacking in ASC children (Baron-cohen et al., 1993).

Emotions can be represented as points inside a 2D continuous space, the circumplex model by Russell (Russell, 1980), spanned by Valence (x-axis) and Arousal (y-axis), Fig. 1. Although a strong debate exists on the nature and the role of emotions and their representation, the discrete representation adopted by FACS is considered a good starting point for designing interventions for ASC children (Zhang et al., 2021).

Among the existing interventions based on serious games, **Emotiplay** is one of the most successful (Fridenson-Hayo et al., 2017). This is a web platform that includes interactive lessons, each targeted to one of the primary emotions. Two additional modules address emotions of kindness/unkindness and pride and are devoted to more competent children. Each lesson contains activities such as drawing and videos of actors portraying situations of daily life associated to an emotion, for instance enjoying playing cards or coloring, or getting sad when losing a game. These are accompanied by detailed verbal recorded explanations. At the end of a session, quizzes are provided in written digital form to evaluate the child competence. Positive reinforcement is provided with gamification: the child earns virtual money with which s/he can unlock digital prizes. Here emotion representation is mainly provided through visual information.

Zirkus Empathico (Zoerner et al., 2016) consists of modules of increasing difficulty. Training starts with recognition and verbalization of the avatar emotion. It then asks to recognize emotions portrayed by videotaped facial expressions, shown alone or inside a context. Transfer to daily life is provided through naturalistic video stimuli (facial expressions and context videos) in the app. Evaluation of the emotion by the child is provided by a virtual doll shown on the screen with two sliders on the side one for valence and one for arousal: valence is represented through the shape of the doll's mouth and arousal through the doll upper limbs position. Also, in this situation emotions are represented mainly through visual information.

JeStimule (Serret et al., 2014) targets all the six primary emotions. It is a 3D life-simulator game, in which an adult avatar, controlled by the player through a joypad, lives in a small city and interacts with NPCs (Non-Playable-Characters) inside the game. Every time the avatar interacts with a NPC, a video is played that shows the NPC experiencing a particular emotion in that area of the city. When the video ends, buttons appear in the lower part of the screen and the child is asked to answer which emotion is the NPC feeling in that situation. Before starting to play the user can select what to show on the button: the face of the NPC expressing the emotion or a color that is arbitrary chosen. Colors are therefore exploited as if they were symbols, associating one color with each emotion. The authors exploited also tactile feedback associated to the controller, to help the user to memorize the selected emotion. Different vibration patterns associated to the different emotions were also provided to the user. This is the first study exploring a possible use of color patterns associated to emotions. However, in this case, the aim of the authors is not to implement multimodal feedback with an emotional color encoding, but to give a set of buttons, limiting possible discoveries on emotion-colors association.

Even if all these studies have different limitations, all of them share an important common feature: these activities show to children, in a different manner, the context in which an emotion can be felt. This helps the children to contextualize emotions inside daily life experiences, helping them to recognize emotions expressed by others in daily life, that is the *generalization of skills* (Zhang et al., 2021). Overall, results obtained in these studies suggest that digital games can be an effective support to ASC socio-emotional therapy, reporting a high level of acceptance and engagement.

3. Method

3.1. Co-design process

We co-designed the application with five therapists: three from the Dosso Verde Rehabilitation Center in Milan, referred to as center 1, and two from the Sacra Famiglia Rehabilitation Center in Cesano Boscone, referred to as center 2. We performed three pairs of workshops, one in each center, following the iterative design process described in (Nielsen, 1993). The first pair of workshops was aimed at defining the requirements of the platform, while the additional two pairs of workshops were aimed at co-designing the platform. In each pair of these latter workshops, we showed the prototype developed up to that time and we collected feedback on how to improve the design of the activities, and on how to improve the interface of the application to be simple and effective to be used during therapy sessions.

In the first pair of workshops, we have also identified the children population that could most benefit from MiEmo intervention: children between 5 and 13 years old (primary and high school), from low-verbal with medium-functioning to verbal ones with high-functioning ASC, keen to use technology. It was also required that children were already involved in a socio-emotional therapy.

The key elements of the platform identified (also capitalizing on the experience reported in (Pirovano, Mainetti, et al., 2016)) were the following.

(i) **The application should** support **the generalization of skills**. Generalization is the ability to apply learned skills in new contexts. To achieve this, in traditional interventions therapists vary the content of an exercise, session by session, while maintaining the same aim of the exercise. In this way, the child faces a similar challenge with different content. The activity may vary for instance in the number and type of activity elements. This allows to prevent the child from hacking activities by learning what s/he is expected to do without any learning of the mechanisms that should lead to the answer. This approach is well in line with the principles of good exergames design (Pirovano et al., 2016b).

(ii) **The application should offer a clean and reusable setting.** Children with ASC face difficulties in keeping their attention; the mediation of a digital device is not enough to engage them (Belmonte & Yurgelun-Todd, 2003**)**. The application should be conceived to facilitate focusing on the activity. To this aim, the interface of the digital application should not show/play elements that are not useful for learning purposes. This requirement is perfectly in line with the User Interface design rules described in (Windl & Constantine, 2003): **avoid unnecessary elements** in the interface.

(iii) **Avoid the risk of sensory overload (SO).** Keeping the interface and feedback system simple is fundamental to avoid the risk of SO. In-game feedback (as well as multi-modal feedback) should be appropriately balanced (Cermak et al., 2010). In particular, jingle should be preferred to loud noises, like an applause, that are poorly tolerated by ASC children (Baron-Cohen, 2006).

(iv) **Personalization.** Each child has his/her own level of competence on emotions. Different levels of difficulty should be implemented to match the capacity of each child. Besides making the treatment most effective, it allows also maximizing engagement. When an application turns out to be too easy, the user becomes easily bored, while when it is too difficult, frustration enters into play. Therefore, a proper level of difficulty allows also to keep the child inside what has been defined as the flow channel (Kiili et al., 2012), where the child gets absorbed by the activity and provides his/her maximum improvement.

We relied on the taxonomy from (Rego et al., 2010) to develop the first prototype of the application.

Table showing the design choiches for MiEmo.

Item of taxonomy (Rego et al., 2010)	Design decision for MiEmo
(i) domain area	Socio-emotional training for ASC children
(ii) interaction technology	Tablet or PC through touch screen or mouse
(iii) the serious games level of difficulty should be appropriated to not frustrate and to not bore the user	Several difficulty levels implemented
(iv) the serious game should include appropriate feedback to engage and motivate the player	In-game and multi-modal feedback implemented
(v) the system should contain a module to monitor the patient's progresses	A separate therapist module to see logs was included in the application.
(vi) identify the context in which the platform will be used	During the daily training on social training
(vii) design appropriate game interface	Simple interface implemented (taking into account the guidelines we received from the therapists, avoiding unnecessary elements and balancing visual stimulus (Cermak et al., 2010))
(viii) choose game genre	Puzzle-like (see Subsection 3.4 for activities description)
(ix) choose number of players	1 child with the therapist nearby

The system has been developed to train children on the four primary emotions. Upon therapists' feedback we have introduced in the last prototype the calm emotion as a neutral state, in which valence is positive and arousal is at a minimum, and the joyful emotion, that is close to happiness with higher arousal, to have a better balance between emotions with positive and negative valence. Calm emotion is used in animated stories before the event that elicits a particular emotion and in all the activities when pictures representing a person with a neutral facial expression is shown. Joyful emotion is used only in the activity that require the child to say what emotion s/he is feeling.

The activities were designed with an increasing level of difficulty in mind ((iii) of (Rego et al., 2010)): first, the child has to understand others' emotions, focusing on the primary ones, and afterwards his/her own emotions including the more complex ones. Moreover, we have started from stylized faces that express emotions in a simplified way; we resorted to faces from PCS that allow to focus on the face elements of eyes, mouth and eyebrows, and then move towards real faces (cf. the study by De Capitani, 2019).

3.2. System description

MiEmo has been developed in C# programming language with Unity 3D Game Engine version 2021.2.16f1. The robot virtual avatar was modeled with Blender v.3 and imported in the Unity project. MiEmo is meant to be used with smartphones, tablets or PCs and the interaction can be carried out with mouse or touch screen. Software compatibility among the different devices is enforced by Unity APIs. The application is standalone, and it does not require internet connection. The speech sentences pronounced by the virtual assistant were pre-generated using Speech Synthesis Markup Language (SSML) and Google text-to-speech (TTS)[2] to generate male and female voices. These sentences were generated and stored in the application during the development phase, so that none of them has to be generated at runtime avoiding the requirement of an internet connection to make MiEmo working. No external library was used.

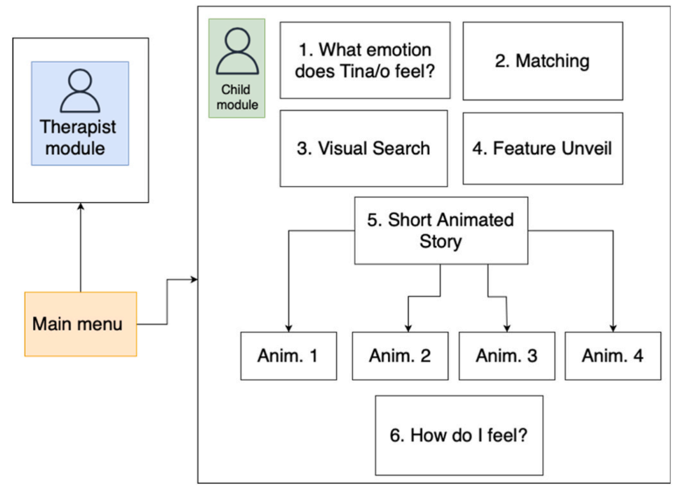

Fig. 2. Shows the MiEmo's structure. The two main modules are shown: one for the therapist and one for the child. The latter allows choosing among six different activities.

MiEmo comprises a module for the child that provides six activities developed for socio-emotion training and a module for the therapist, that allows having an inside out view on the player performance. Fig. 2 shows the application structure. All the activities are accessible from the main menu.

To support an increasing level of difficulty, we have identified different visual supports: (i) stylized pictograms faces; (ii) pictures of Human faces expressing emotions, (iii) videos of a Human face expressing an emotion, (iv) animations of stylized faces, (v) reconstruction of the stylized faces expressing an emotion. For a contextualization, at the end of the level are shown animations of a virtual assistant expressing an emotion in a daily life situation. The activities will be described in more details in Section 3.4.

For all the activities we have designed multi-modal feedback that incorporates also music, color and gestures.

3.3. Design of multi-modal feedback

Feedback is provided in association to the result of each action. Such feedback should be as clear as possible and it is positive when the action is correct or negative when it is not. In **classical in-game** feedback specific sounds associated to these two outcomes are played at the end of the action. Additional feedback elements used can be symbols for success or failure, or voice (compliments or encouragement) (Johnson et al., 2017).

We aim here to go one step further on feedback leveraging the nature of the learning domain, that of emotions. We resort to specific music pieces and colors to stress the emotion that is the current target of learning. We will call this **advanced multi-modal** feedback.

Upon success, classical in-game feedback is played as: a happy jingle, a PCS graphical symbol of "well done", appraisal through voice by the virtual assistant, and a festive animation with confetti (Fig. 4c). A happy jingle was preferred to other classical feedback like applauses or other kinds of "noisy" sounds (Boyd, 2019, pp. 446–451), to match sensory preferences of ASC children. The virtual assistant provides positive feedback through voice, selecting randomly among a set of predefined sentences ("compliments", "well done", "great" and so forth) to avoid stereotypes interaction on one side and to avoid the risk of echolalic speech in the child (Xie et al., 2023).

Upon incorrect action the sound of a buzzer is played, followed by

[2] https://cloud.google.com/text-to-speech/docs/ssml.

the virtual assistant encouraging to try again through voice (Fig. 4b). A brief sound of buzzer (a typical game-show buzzer sound) was selected to respect sensory preferences as it was done for the positive feedback (Boyd, 2019, pp. 446–451). Moreover, we resorted to this sound as it is widely used in both rehabilitation centers of the pilot for negative feedback indicating errors during traditional training, so it was already familiar for the children.

Upon success, advanced multi-modal feedback based on music, colors and animations with gestures are also provided (Fig. 4d). We describe hereafter the advanced multi-modal feedback components.

3.3.1. Music feedback

The music features associated to the primary emotions and to Clam and Joy, according to (Juslin & Laukka, 2004), (Juslin, 2019), are reported in Table 1

In particular, we have considered here classical instrumental music pieces, since they do not have the complexity of the songs with words and can respect sensory preferences of ASC children helping them to focus on the activity during learning sessions as suggested by (Katagiri, 2009). Moreover, such type of music is generally well accepted by ASC children who attend music therapy sessions (Ke et al., 2022). After analyzing several different music pieces, we have chosen the following music pieces selected from the Disney movies Fantasia 1940 and Fantasia 2000, which are particularly appreciated by all children including ASC (Smagorinsky and Smagorinsky, 2016), and that nicely fit the above-mentioned characteristics.

- **Happiness**: The Nutcracker Suite, Op. 71A, Chinese Dance (Fantasia 1940) [3]
- **Anger**: Tin soldier, Piano Concerto No. 2, Allegro Op.102 by Shostakovich (Fantasia 2000) [4]
- **Sadness**: Firebird Suite by Stravinsky (Fantasia 2000) [5]
- **Fear**: Rite of Spring 2/3 (Fantasia 1940) [6]
- **Joyful**: Symphony No. 6 (“Pastoral”), (Fantasia 1940) Op. 68, III·[7]
- **Calm**: Nutcracker Suite, Op. 71A, Dance of The Sugar Plum Fairy (Fantasia 1940).[8]

We have also identified a music piece to be played when the outcome of an activity is not a valid emotion, such when the child reconstructs an PCS face with angry eyes and a smiling mouth. In this case, we play the is Rite of Spring 1/3 of Fantasia 1940[9], a music piece as close as possible to a neutral emotional state, with valence and arousal close to zero.

Table 1
Musical features correlated with discrete emotions in musical expression.

Emotion	Music Features
Fear	Fast tempo, large tempo variability, minor mode, dissonance
Anger	Fast tempo, small tempo variability, minor mode, atonality, dissonance, high sound level
Sadness	Slow tempo, minor mode, dissonance, low sound level
Happiness	Fast tempo, small tempo variability, major mode, medium-high sound level, small sound level variability
Calm	Slow tempo, consonant harmony, repetition, use of few instruments
Joyful	Fast tempo, small tempo variability, major mode with a bright tone, medium-high sound level, small sound level variability, features instruments with a bright timbre as trumpets

[3] https://youtu.be/jfos4IXFqvU from time 0:00
[4] https://youtu.be/5KOtmJLFINo?t=152 from time 2:40
[5] https://youtu.be/cmOBFaIOHVE?t=287 from time 4:47
[6] https://youtu.be/Wylr8g5nyrc?t=304 from time 5:00
[7] https://youtu.be/06spmjjO-zM from time 0:00
[8] https://youtu.be/xUxma9gKAzc from time 0:00
[9] https://youtu.be/4HfvWp9W3nQ from tine 0:00

3.3.2. Design of the virtual assistant

Three main characteristics of the virtual assistant emerged from the workshops with the therapists: (i) the character should show the facial expressions in the same stylized way of the pictograms not to confuse the child, (ii) the character should express the emotions with face and body movements and (iii) the child can easily feel empathy with the virtual assistant.

After the first round of workshops, we opted for a robotic shape (Fig. 3), inspired by (Chevalier et al., 2017) which reported that children with ASC appreciate to interact with robotic entities, better than with humanoid virtual representations. Other studies report also a preference of children for robotic shapes, since robots can convey emotions in a simplified way (and not with all the multi-modal complexity of human agents) and are perceived as predictable agents that behave following rules and standard patterns (Cabibihan et al., 2013).

We have designed a male and female robot and named them Tino and Tina (in Italian language the diminutive of the word robot is “robot-tino” for male and “robot-tina” for female) to foster children empathy. From the main menu the child can choose the preferred gender of the avatar. The robot has mostly white parts, with blue or pink mechanical parts, respectively for male and female versions. We opted for male and female versions to facilitate a possible identification of the child, or one of him/her parent, with the robot. This helps the child to empathize with the avatar that could facilitate socio-emotional learning (Huggins et al., 2020). The robotic avatar expresses the emotions through its stylized face (consistent with PCS), with gestures and with the change of its color (Fig. 3).

Simple gestures associated to the emotions (Mazzoni et al., 2020) have also been implemented (Fig. 3). In the calm state the robot slightly levitates lifting up and down its arms. Happiness is similar to calm with the upper limbs that tend outwards when moving up and down. Anger is shown with a rapid lowering of the arms outstretched with closed fists, and sadness is represented with a slow lowering of the arms at the side of the trunk and collapsed open hands. Fear is represented with arms raised up quickly and stretched. These gestures have been shown to be linked to emotions perception (Della-Torre et al., 2021), (Actis-Grosso et al., 2015).

3.3.3. Color feedback

Based on previous studies (Takahashi & Kawabata, 2018), (Nakajima et al., 2017), (Casiddu et al., 2021), we have associated two colors to each the primary emotion considered: yellow and orange to happiness, two different blue nuances (same hue and saturation, different lightness value) to sadness, two different red nuances to anger, two different purple nuances to fear (cf. Table 2). We associated two different colors since otherwise the face parts would not be easily visible: we choose a darker color for the body and the face parts (eyes, mouth and eyebrows),

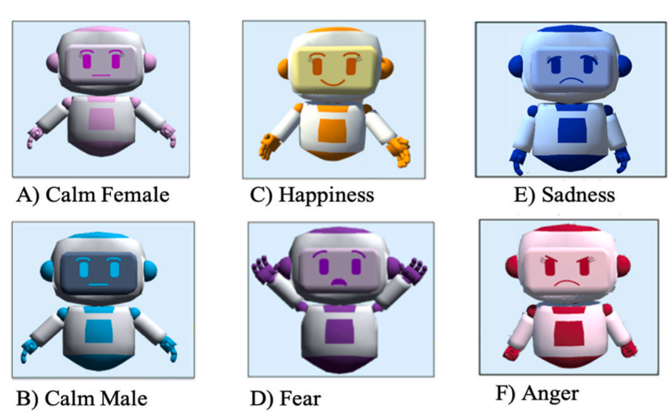

Fig. 3. Tina and Tino expressing the primary emotions with gestures and colors. Video available at https://tinyurl.com/AvatarEmotions.

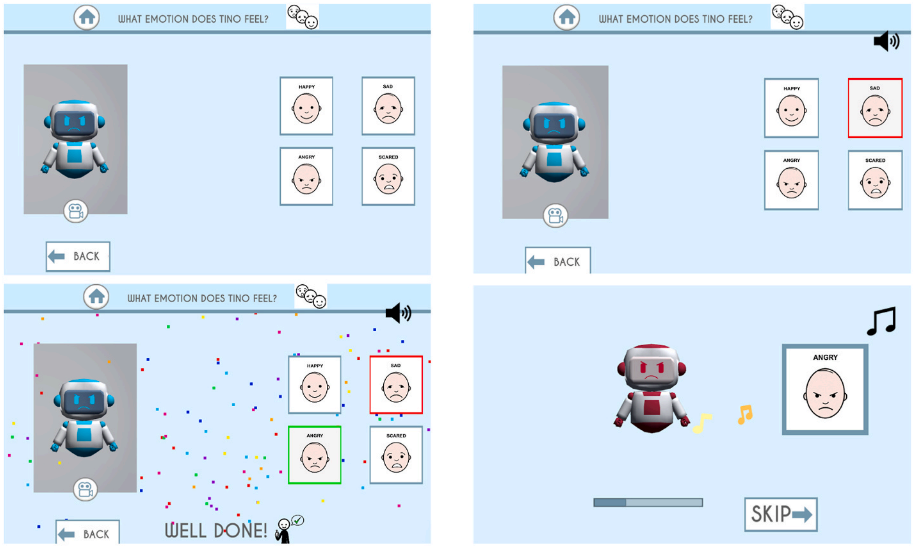

Fig. 4. Example of the Activity "What emotion does Tino/a feel?": (a) On the left part of the screen Tino is expressing anger with gestures and face expression, and on the right part PCSs of happiness, sadness, anger and fear are shown. The child has to click on the correct PCS. (b) Upon incorrect choice a red frame appears around the chosen PCS and negative in-game feedback is provided: a buzzer is sound followed by the virtual assistant voice encouraging to try again. (c) Upon success, a green frame appears around the chosen PCS and positive in-game feedback is played: a happy jingle is played along with the congratulation by the virtual assistant. The text corresponding to the assistant's statement is shown at the bottom of the screen with the PCS of "well done" nearby. (d) Just after the jingle ends, the multi-modal feedback is provided: the assistant is shown in the center of the screen next to the correct pictogram, expressing anger with face expression and gestures. The face and body also gradually change color into red and the Tin soldier music associated to anger is played.

Table 2
Colors coded as hue; saturation lightness used for emotion feedback

Emotion	Body color (including eyes, eyebrows and mouth)		Head color	
Happiness	H:40, S:100, L:55		H:60, S:100, L:80	
Anger	H:350, S:78, L:40		H:350, S:78, L:90	
Sadness	H:231, S:77, L:38		H:231, S:77, L:90	
Fear	H:283, S:62, L:42		H:283, S:62, L:90	

and a lighter color for the head. Colors have been selected in the HSL (Hue, Saturation and Lightness) space so that different nuances could be easily identified keeping the same value of Hue and Saturation and changing the value of lightness. The colors selected are part of the colors positively accepted by ASC children, to respect their sensory preference (Wardani & Mustikasari, 2023). Regarding colors selected, often black and dark are associated to fear in addition to purple, but in accordance with therapists we opted for one color, the purple, to support children's memorization (Wardani & Mustikasari, 2023).

It is here worth noticing that, although colors are largely considered in interfaces to support ASC people to memorize information (Ludlow et al., 2008), (Wardani & Mustikasari, 2023), their use as a reinforcement element (particularly in association with emotions) has not been used yet. Thus, this study is the first attempt to investigate the possible role of colors in reinforcing emotional feedback.

3.4. Description of the activities

Activities are structured for an increasing difficulty: the simplest exercises make use of PCS, afterwards pictures of real faces are shown; these have been extracted from the pictures provided by the FACES dataset (Ebner et al., 2010) that offers a large choice of different actors and emotions. Moreover, the pictures have been taken all from the same point of view and background, reducing variability.

We designed the interface of each activity with a well-defined structure: the interacting elements are located always in the same position of the User Interface to not disorient the child. Every time an activity is loaded the elements are randomly chosen and located in one of a set of predefined positions. In this way the child is confronted each time with a different disposition of the elements on the screen. This is implemented to prevent from "cheating", i.e., to avoid that the child learns where the winning element is positioned in each game instead of what the correct answer is and it promotes generalization of acquired abilities (Zhang et al., 2021). Finally, MiEmo implements accessibility with both text to speech voice audio and PCS symbols that are paired with each text statement (including the title of the activity) to facilitate visual learning of children with ASC that prefer visual feedback to written instructions (Grandin, 2018).

To offer the most suitable challenge, each activity implements three levels of difficulty: **easy, medium, and hard,** according to the number of elements proposed. We neither offer levels of difficulty for the first and the last activities and neither for the animated stories in which the child has the role of observer (a screen capture of a full play session is available at https://tinyurl.com/MiEmoGameplay). A detailed description of the activities is reported hereafter.

3.4.1. "What emotion does tina/tino feel?" activity

This activity is **mandatory**, and it is performed once for each activities' training session. The robot avatar is shown on the left part of the screen and expresses an emotion from the set with both gestures and face expression. On the right part four buttons appear showing the four primary emotions through PCS, and the child has to click on the correct PCS (Fig. 4a). Notice the sprite besides the activity title (common to all activities) that is the PCS representation of the current activity. The aim of this exercise is to help the child understanding which face micro-movements are associated to a given emotion.

Upon incorrect choice, the negative classical in-game feedback is provided: a buzzer sound is played, the avatar spells the emotion selected by the child and then says that it is not the correct answer and then encourages the child to play again. (Fig. 4b).

When the child picks-up the correct PCS face, the classical positive in-game feedback is provided (Fig. 4c). Just afterwards, multi-modal feedback is presented to the child for the first time: the assistant expresses the correct emotion with gestures and face expression, while the corresponding music piece is played and the body and face of the virtual assistant change the color (Fig. 4d). The child can click on the skip button if annoyed by feedback. Video available at: https://tinyurl.com/MiEmoWhatEmotionDoesTinaFeel.

3.4.2. Matching activity

One or more pictures of the face of an actor, portraying an emotion, is shown on the right part of the screen while the PCS pictograms associated to those emotions are shown on the left part. The child should drag and drop each picture of the actor's face below the correct PCS. In this activity the child learns which parts of a real face to look at to identify an emotion (associating the facial features of the pictogram with those of a human face). At the **easy level** a single picture and a single PCS are shown on the screen (no possibility of error); at the **medium level** two pictures of the same actor expressing two different emotions and the corresponding two PCSs are shown (Fig. 5a); at the **hard level** two pictures of different actors are shown.

Upon success, classical positive in-game feedback is provided followed by the multi-modal one (Fig. 5b): the pairs PCS-picture are highlighted in sequence (with an arrow indicating the current pair), while the music piece, corresponding to the current pair, is played. When the wrong choice is made by the child, classical in-game negative feedback is played (Fig. 5c). Just afterwards, the two pictures are automatically moved back to their initial position, the assistant says that the answer was not correct and that "the correct answer is", then the two pictures are moved automatically under the correct PCS, and afterwards the multi-modal one is played (as after victory). The forward button appears only at the end of the multi-modal feedback. The child can also decide to repeat the exercise by clicking on the reload arrow in the center-bottom of the screen. Video available at: https://tinyurl.com/MiEmoMatching.

3.4.3. Visual Search activity

The PCS of an emotion is shown on the top of the screen, and the child has to click on the correct picture of a Human face expressing that emotion, among several pictures shown on the bottom.

We designed this exercise to offer a level of difficulty higher than matching, since in this activity the child has to look at the facial features in each picture to find the correct one. At the **easy level** three different pictures of the same actor are shown (Fig. 6a); at the **medium level** three pictures of different actors; at the **hard level** a set of five pictures of different actors are shown.

If the picture chosen is wrong, a red frame appears around the chosen picture, a buzzer sound is played, and the virtual assistant spells the emotion chosen by the child, and the music associated to the emotion selected by the child is played. Afterwards the virtual assistant encourages the child to try again. Upon success, classical positive in-game feedback is provided. Just afterwards, the multi-modal feedback is played: the PCS of the emotion is shown centered on top of the screen, the video of the actor from a neutral expression to that of the correct emotion is shown under the PCS (Fig. 6b) while the virtual assistant spells the emotion name, and the associated music piece is played. The video of the actor expressing an emotion is obtained by morphing his/her picture of the neutral expression with that expressing the emotion considered. Morphing was created through the Face-morph library (Campbell, 2018). The forward button appears only after the correct picture has been chosen. Video available at: https://tinyurl.com/MiEmoVisualSearch.

3.4.4. Features Unveil activity

Only the outline of a PCS face is shown in the center of the screen without eyes and mouth and the child has to fill it with the correct eyes and mouth of a given emotion. To this aim s/he has to drag and drop the correct components, displayed at the side of the face, in their proper place (Fig. 7a). When done, the child will push the Check button to verify the built PCS face. The goal of this activity is to show what shape the various parts of the face take in the expression. It offers a higher level of difficulty than previous activities, as the child needs memory and the abstraction abilities to imagine the stylized version of mouth and eyes of a face expressing an emotion. At the **easy level:** eyes and mouth of three possible emotions are provided (one is the correct emotion and the other two are randomly chosen); at the **medium level**, the elements of the four primary emotions are shown; at the **hard level**, a pair of eyes associated to neutral face and a pair that is not congruent with any emotion are added. In the latter case, the not congruent pair has one eye associated to sadness and the other to anger.

Upon success the positive reinforcement is played, and afterwards the animation of the PCS moving from a neutral expression to that of the correct emotion is shown, while the virtual assistant names the emotion, and the associated music piece is played (Fig. 7c).

If the wrong emotion is reconstructed, the virtual assistant spells the name of the emotion reconstructed by the child and the associated music piece is played. If the reconstructed PCS does not represent any emotion, the virtual assistant says, "no emotion", the buzzer sound is provided, and the music piece associated to the neutral emotional state is played. As this is the most difficult exercise, a hint is provided to the child upon

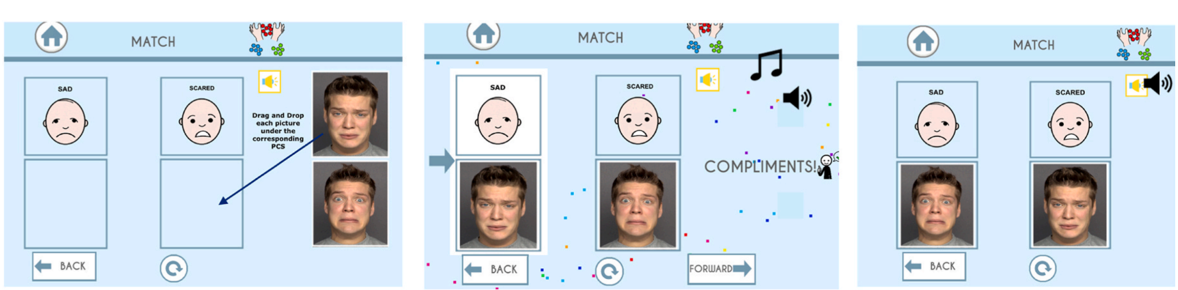

Fig. 5. Example of the "Matching" activity at the medium level **(a)** The child will drag&drop each picture under the corresponding PCS, sad and scared in in this example. **(b)** Afterwards, positive in-game feedback is played: compliments by the avatar, jingle and confetti animation. Afterwards, the pairs of PCS-picture are highlighted in sequence (with an arrow indicating the current pair), the avatar names the emotion, and the corresponding music piece is played. **(c)** If the child makes an error placing both pictures under the wrong PCS, the buzzer sound is played. Afterwards, the pictures are moved under the correct PCS; they are then highlighted in sequence (with an arrow indicating the current pair), the avatar names the correct emotion, and the corresponding music piece is played. No classical positive in-game feed-back is played as it was in (b). The child can also decide to repeat the exercise by clicking on the reload arrow in the bottom center of the screen.

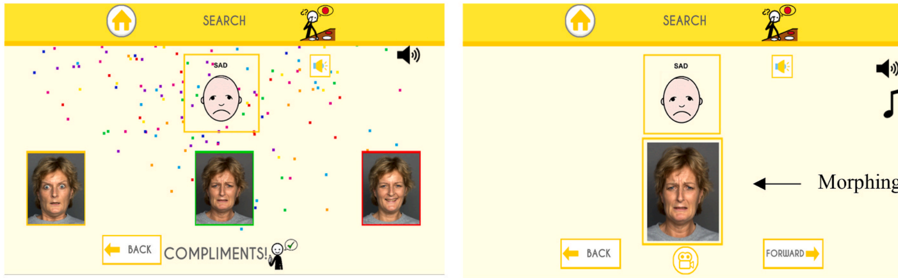

Fig. 6. Example of "the Visual Search" activity at the easy level (a) The child has to click on the picture of the actor expressing the same emotion of the PCS. If the answer is correct, positive in-game feedback is played: a green frame appears around the correct picture, the confetti animation is shown, and a jingle is played. (b) After positive feedback, the video of the actor moving from the neutral expression to that of the correct emotion is shown under the PCS, while the virtual assistant names the emotion, sadness in this case, and the associated music piece, "Firebird suite" is played. The child can click on the button with a camera to watch the video again.

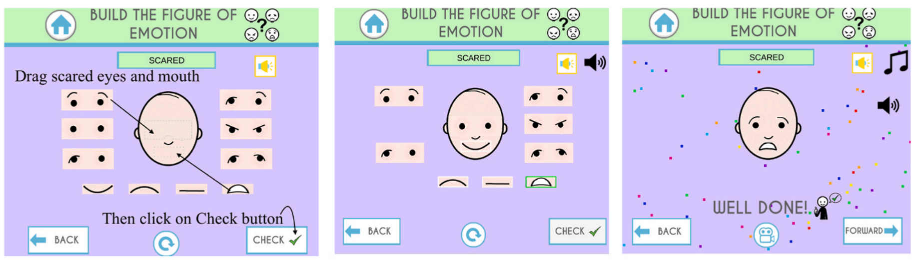

Fig. 7. Example of "Feature Unveil" activity at the hard level. (a) The child has to choose the scared eyes among the six possible pairs of eyes and the scared mouth among the four possible mouths to reconstruct the PCS of Fear. When the child is satisfied, s/he can click on the check button to receive feedback. (b) Here the child has chosen happy instead of scared: the music piece associated to happiness is played while the assistant names "happy", and a buzzer sound is played. Afterwards, the afraid mouth is framed in green, followed by the voice of the virtual assistant that encourages to try again and to look at the hint given. Just afterwards, the eyes and the mouth of happiness are relocated in their original position and the child can try again. (c) Upon success, after the classical-in-line positive feedback, the animation of the PCS from the neutral expression to fear is shown in the center of the screen, the virtual assistant names the emotion and the congruent music piece is played. The child can click on the button with the camera to watch the animation again.

error: one correct element randomly chosen between mouth and eyes -pair is highlighted in green, the virtual assistant encourages the child to look for the hint and to try again (Fig. 7b). The child has a maximum of three attempts to build the right face. Afterwards, the assistant says that the attempts are finished, the eyes and the mouth are located again in their starting position, and an animation starts in which eyes and mouth in sequence move into the correct position. Afterwards the virtual assistant says "the correct answer is" and the multi-modal feedback associated to the correct answer is played. Video available at: https://tinyurl.com/MiEmoFeatureUnveil.

3.4.5. *Animated stories activity*

We designed them as situations that the children could have experienced in everyday life and that can be easily associated to an emotion. We have realized four short animations designed to elicit a given emotion; these stories represent four situations in which the child will observe the emotions felt by the 3D virtual assistant.

At the end of each story, the 3D assistant expresses the emotion felt with advanced multi-modal feedback. This requires that the child put him/herself in other's shoes, that is the basis of ToM (Baron-Cohen et al., 1985) and it supports the generalization of the acquired skills to everyday life. We structured the stories following Vollm's strips structure (Völlm et al., 2006), composed of the three basic sequential states: initial state (calm or happy depending on the story), event and final emotion. We designed (i) a story of happiness, in which Tina brings a gift to Tino (or vice versa) Fig. 8a, (ii) a scaring story, in which Tina/Tino is scared by a spider coming down from the ceiling Fig. 8b, (iii) a sad or angry story, in which in which Tina/Tino drops an ice cream on the floor, and another (iv) sad or angry story, in which Tina/Tino accidently breaks a toy car. Stories iii) and iv) can evolve dynamically with a variable ending (Fig. 9, Fig. 10): the child can attribute to the robot either the angry or sad emotion through a pop-up with two buttons, each one with a PCS of one of the two emotions.

The animations are without speech since this distract the child, but they are accompanied by the music piece associated to the emotion felt by the robot (Katagiri, 2009). When the transition between the initial emotion and the final one takes place, music transition between the two pieces associated to the two emotions is achieved through fading. In same time interval, the color of the face and of the body of the virtual assistant change color according to Table 2, and so does the virtual furniture of the scene. Finally, the virtual assistant produces the gestures congruent with the animation.

To drive the attention of the child on the virtual character, all the moving elements of the scene are concentrated around the avatar. To this aim, we have computed the Saliency Maps of the scene[10] that have demonstrated that the player does focus on the avatar.

3.4.6. *"How do I feel?" activity*

This activity is meant as the last activity in a habilitation session. The aim of this exercise is to help the child to name her own emotion and associate it with a PCS. This activity is different from the initial one as it requires the child to express the emotion felt by him/her. This activity does not have correct or wrong answers, so in-line classical positive and negative feedback here are avoided. A set of buttons with the PCS of the emotions are shown on screen (Fig. 11a). The emotions are balanced: three with positive valence: calm, happiness, joy and three with negative valence: sadness, anger and fear. Once the child selects the emotion, the corresponding PCS is shown on the center of the screen, the virtual assistant names the emotion and the music piece is played (Fig. 11b).

[10] https://github.com/ivanred6/image_saliency_opencv

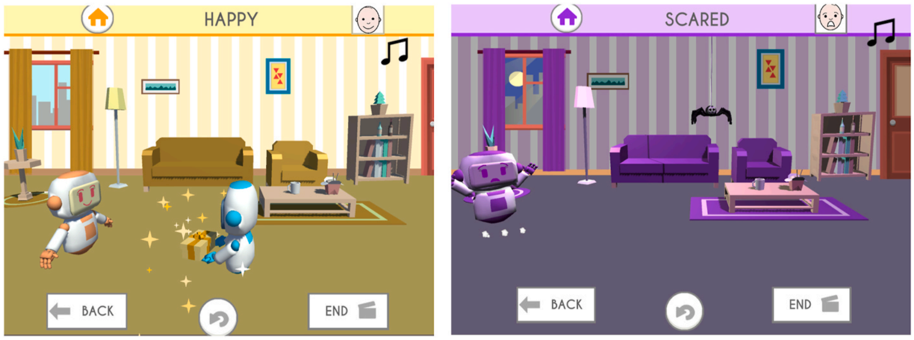

Fig. 8. (a) Happy story: Tina/o receives a gift. When Tina sees the gift brought by Tino, colors start to change: Tina color changes to yellow and the same happens to the assets in the scene. Moreover, Tina expresses happiness with gestures and face expression, and "The nutcracker suite - Chinese Dance", associated to happiness, is played. **(b)** Scaring story: a spider comes down from the ceiling at night. When Tina looks up and notices the spider, colors start to change: Tina color changes to purple and the same happens to the assets in the scene. Moreover, Tina expresses fear with gestures and face expression. Tina also runs in a scattered manner from a part to the other of the room, and "Rite of spring", associated to fear, is played.

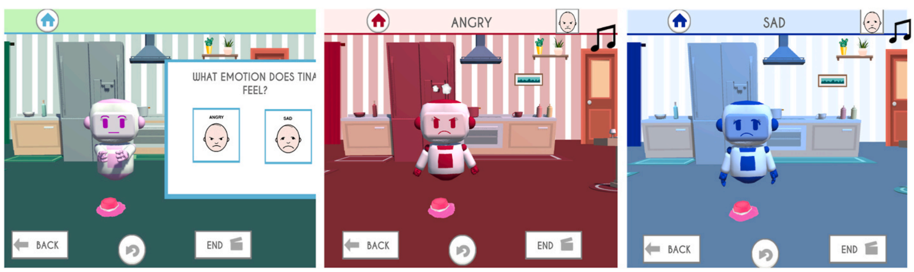

Fig. 9. Anger and sad possible endings for the story of Ice Cream Falls (a). The child has to click either on the PCS of the anger emotion (b) or on that of sadness (c). Just after, the animation starts: the avatar expresses that emotion with gesture and face, at the same time the avatar changes its color together with the assets in the scene and the corresponding music piece is played.

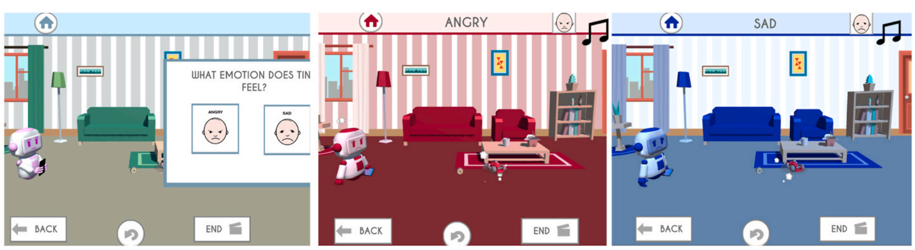

Fig. 10. Anger and sad possible endings for the story of Toy Car Broken (a). Also in this case the child will click on the PCS of the emotion chosen to end the story: between anger (b) and sadness (c).

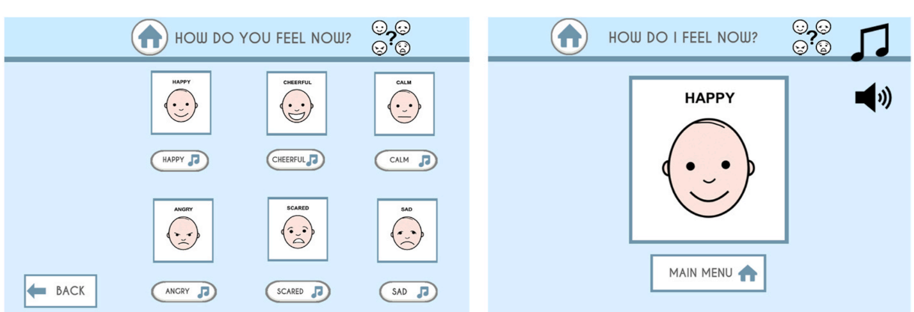

Fig. 11. Screenshots of How do I Feel? activity
(a) The child can push on the PCS or on the label under it to select the emotion s/he is feeling, from left top to bottom right: happy, joyful, calm, angry, scared, sad. **(b)** Once selected the emotion felt: the corresponding PCS appears in the center of the screen with the button to go back to the main menu under it, in this case the assistant names "happy" and the music piece associated is played.

Video available at: https://tinyurl.com/MiEmoHowDoIFeel.

4. Results: a first test of playing with MiEmo as a case study

A pilot study involved 19 children; it was carried out with at least one therapist present along with one experimenter. The study was aimed at testing the usability (Sánchez et al., 2012) of MiEmo and the acceptability of it by ASC children, with the possibility to collect from the therapists' further suggestions and possible criticisms.

The study was run in two rounds (i.e. R1 and R2). **R1** was held in the rehabilitation center n.1, with 8 children, two females and six males diagnosed with low-verbal, medium-functioning autism (Mean age = 8.875, SD = 3.218, cf. **Table 3**). **R2** was held in rehabilitation center n.2, with 11 children, 1 female and 10 males diagnosed with verbal, high-functioning autism (Mean age = 7.818, SD = 1.336, cf. **Table 3**). The study was approved by the Ethical Committee with identification "University of Milan, ID 98/22".

The inclusion criteria were set such that children who could maximally benefit from the platform could be included; they were: (i) Age≥5 years and≤13; (ii) Appreciation to use digital technology (defined by the therapist); (iii) Attending Social skills therapy sessions; (iv) Diagnosis of medium or high functioning autism based on DSM-2013 (DSM-5, 2013) (iv) Children with basic reading skills, since MiEmo does not require high level of reading skills as it implements accessibility with both text to speech voice audio and PCS symbols along with written text. **The exclusion criteria were:** (i) Diagnosis of a severe form of autism; (ii) Being not comfortable with the use of technology; (iii) Children who do not use verbal communication at all.

The children were chosen by therapists involved in the study. Each therapist selected among the children s/he was training the ones that fit inclusion and exclusion criteria. Once the children were selected, their parents were informed on the possibility to participate to the study with both informed consent and a meeting with the therapists and the computer scientist to full inform them. Also, children were informed on the study with a simplified version of the informed consent with images appropriated for their age. The informed consent versions, for parents and for children, were approved in the Ethics Committee session. All the parents and children contacted agreed to participate.

4.1. Testing procedure

19 children were involved (CH1 … CH19). **Table 3** shows demographic data of the children participating in the study. The MiEmo platform was tested inside a slot of 1 h of regular therapy session, one child at a time with one therapist and the experimenter present in the room. Two A7 Samsung tablets with Android 12 were distributed with MiEmo installed on, one for each center. The application trial lasted no more than 20 min of play, during which the child played all the activities of a chosen level in sequence.

Three therapists participated in R1 and five in R2. In both R1 and R2 each therapist cared for a maximum of three of the children involved in the study. The pilot with MiEmo lasted a week and every day different children played with the application. Every child involved played with MiEmo once.

All the children played at the level of difficulty their therapist considered suitable for them. After this, each child could continue playing up to 20 min selecting the preferred difficulty level, in case s/he wanted to do it, to explore freely the application. The testing procedure is described in Fig. 12. At the end of the 20 min session the child returns to participate in the regular training session, and once concluded the session the therapist filled the questionnaires in paper forms, the therapists were also given with blank paper for notes to be used in the debriefing. The objective of the study was to deepen usability of the application, leaving as most as possible the children to explore use of the application to collect their acceptance and their perception of MiEmo.

The application was integrated inside daily sessions of training, and to reduce possibility of bias between the two rounds in both R1 and R2, all the child played with the therapist in a separated area for 20 min with MiEmo, and the therapist observed the behavior of the child filling usability questionnaires described below. A low bias can be reasonably ascribed to the therapists' opinion in filling out the questionnaires, given the low number of children associated to each therapist, who knew very well the children s/he cared. In fact, the therapists had already built a relationship of play and trust with each child, and they knew their unique tastes, behaviors (to signal something stressful or something appreciated) and needs. Each therapist's opinion on MiEmo usability focused primarily on the individual child's perception and the suitability of the application to be used during sessions with her/him. Most of the children involved had basic reading abilities with exception of the two children aged 5 in R1 (CH3 and CH5), since MiEmo does not require profound reading skills as it implements accessibility with both text to speech voice audio and PCS symbols that accompany each written statement.

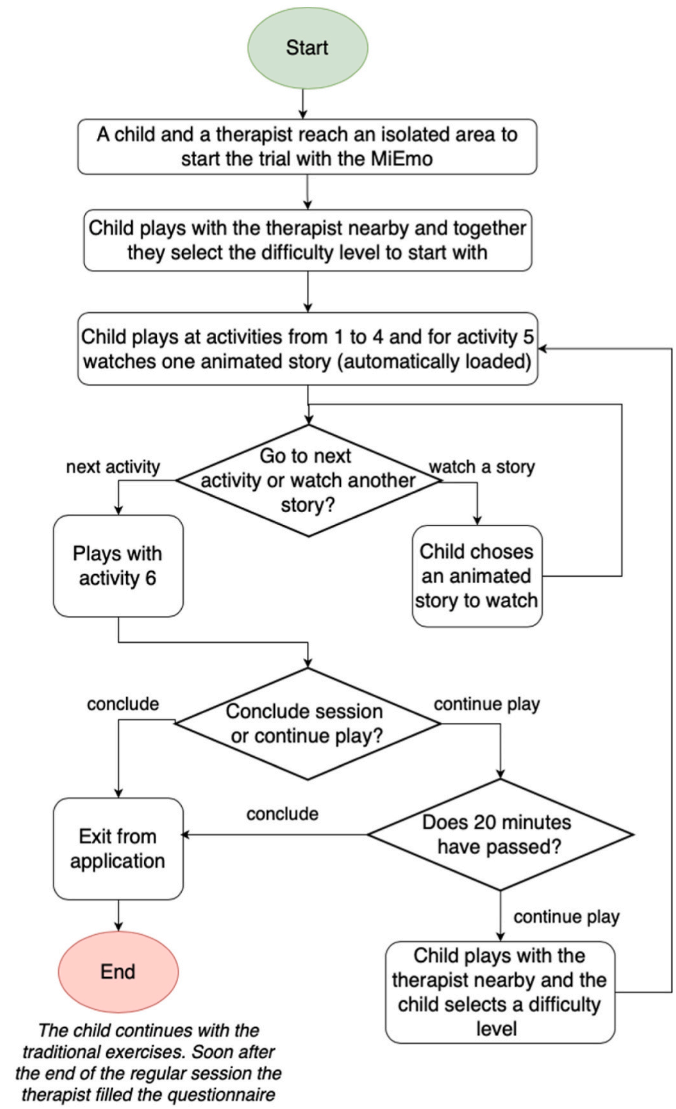

Fig. 12. Flowchart describing the usability study procedure, each child will play a session of 20 min maximum with the therapist nearby.

All the children involved both in R1 and R2 attended social skills training with the therapist also involved in the study Mi Emo was integrated in these social skills sessions. The socio-emotional training was held with a similar protocol in both the rehabilitation centers, and it included a phase of socialization and play with other children and a phase on emotion recognition training from pictures. However, each protocol was personalized and tailored to the child needs. In **R1** all the children did a session of singular training with the therapist, it could happen that different therapists along with their child shared large common spaces such as the gym. In **R2** training was similarly organized, except for children CH14 - CH15 that were trained in pair, but each child was followed by her/him personal therapist. In this case, to ensure low bias and to avoid that a child could influence the other one, the two children played with MiEmo in turn, sequentially: while a child continued traditional therapy in the usual room, the other one (along with her/him therapist) moved far away in a separated area for the 20 min training with MiEmo and vice versa.

4.2. Evaluation methodology

Evaluation has been carried out through questionnaires administered at the end of each trial. Each questionnaire has two different sections: one for the therapists (questions 1–30) and one for the children (questions 31–48); for the latter the therapist compiled the questionnaires for the children. Qualitative observations were collected through a brief debriefing with the therapists after each session in the form of a free format interview. The debriefing stars soon after the therapist filled the questionnaires: s/he will use questionnaire answers and paper notes to report in free form to the computer scientist all the observations noticed during the use of MiEmo with the child. The free format interviews always started with questions to the therapist on open items Q37-Q40, that were used as a starting point to begin the debriefing. Fig. 13 shows the testing and evaluation procedure phases.

Each item of the questionnaires, except open questions, was scored with a Likert scale from 1-("strongly disagree") to-5 "strongly agree". The questionnaires were administrated in Italian Language. The following questionnaires were used (specific questions are given in the Appendix).

1. (Q1-Q10) *System Usability Scale SUS* (Brooke, 1996), that was already positively evaluated in the ASC domain (Junaidi et al., 2022).
2. (Q11-Q16) to assess the Perceived Usefulness (PU) of the application (Holden & Karsh, 2010).
3. (Q17-Q22) for the Perceived Ease of Use (PEU) to assess the perceived mental/physical effort while using the application (Holden & Karsh, 2010).
4. (Q23-Q30) on the evaluation of the activities included in the application.
5. (Q31- Q36) on children's acceptance of the application.
6. (Q37-Q39) Three open questions on the evaluation of advanced multi-modal feedback.
7. (Q40) One open question to collect more in-depth feedback: "Do you have feedback on the application?".

We added a small questionnaire (Q41-Q52) on children's acceptance of multi-modal feedback only for R2, so that we could collect specific information from high-functioning verbal children, in particular their impressions on music and color feedback.

At the end of the study, we had different sets of data. First, we analyzed quantitative data from questionnaires. We used a non-parametric U Mann Whitney test to compare the results of R1 and R2, and to assess differences between the two populations. Secondly, we analyzed qualitative data, which included both the data obtained during the debriefing at the end of each individual session and those collected by on-field observations of the child while playing. Given that questionnaires were filled by the therapist also for the sake of the child, the therapist had the chance to ask the child a qualitative evaluation on the quality of the interaction during each session. These qualitative data helped us to better understand the quantitative scores given in each session.

Finally, the debriefing after each session gave very valuable feedback, together with the on-field observation made by the experimenter. These resulted in the feedback reported in Table 4 which summarizes both the main outcomes and directions for future development.

4.3. Questionnaire analysis

In general, the application was well accepted by all the children (Q31: R1 - IQR = 0.25, Median = 4.0, Mean = 4.125 | IQR = 0.0, Median = 5.0, Mean = 5.0, see next paragraph). One child of R2 was excluded from the analysis because he was showing anxiety already before the application was launched on the tablet saying that he did not want to see the "blue color of sadness" and "the face of anger". The therapist insisted on playing the game together, to show to the child that anger can be faced and that there was nothing to be scared about. However, this behavior, which is interesting from a qualitative point of view, resulted in an incongruent compilation of the questionnaires by the therapist. For this reason, data from this child were not included in the analysis.

No particular issue was collected for the following activities: "What emotion does Tina feel?", "Matching", "Visual Search", "How do I feel?". Children in R1 found "Feature unveiled" activity very challenging.

Regarding the difficulty level, in R1, the 4 youngest children (CHI1,

Table 3
SUS score.

R1 – Medium functioning ASC				R2 – High functioning ASC			
Child	Age	Sex	SUS score	Child	Age	Sex	SUS score
CH1	7	M	75	CH9	7	M	90
CH2	12	M	92.5	CH10	6	M	97.5
CH3	5	M	82.5	CH11	8	M	95
CH4	11	M	90	CH12	8	F	95
CH5	5	F	97.5	CH13	7	M	97.5
CH6	12	M	85	CH14	8	M	95
CH7	13	M	92.5	CH15	6	M	97.5
CH8	6	F	80	CH16	8	M	100
				CH17	11	M	100
				CH18	8	M	100
				CH19	9	M	72.5 - outlier score

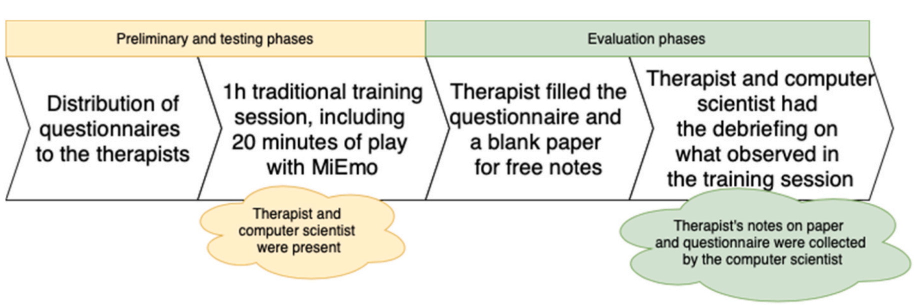

Fig. 13. Graph showing phases of the Evaluation Procedure from preliminary preparation to data collection.

Table 4
Lessons learned.

Feedback collected	Action required to improve MiEmo
Under-representation of positive emotions	It would be appropriate to add a larger number of positive emotions to balance the negative ones.
More animated stories	More animated stories should be added, in particular on positive emotions. Also, a story on the calm state should be included.
Ability to upload pictures to the app	The therapist could upload pictures of the child and use them in the activities instead of the pictures of the actor.
Single interaction mode	Low-medium autistic children should have a single interaction mode: either through drag & drop or click. The application may also allow to choose the preferred modality of interaction for all the activities.
Personalization of speech speed and of animation speed	For low-medium autistic children, the therapist should personalize for each child the speech speed of the virtual assistant and of its animation.
Sound Feedback customization	For low-medium autism, the therapist should personalize positive and negative sound feedback.
Feature Unveil activity update	For low and medium autism, PCS positive/negative feedback should be played automatically (removing the verify button).

CH3, CH5 and CH8) played at the easy level, CH2 played at the medium level and 3 children (CH4, CH6, CH7) at the difficult level, the therapists choose the difficulty level most appropriated for children's age and their current training. The 10 children of R2 all started at the easy level and progressed up until the most difficult one.

As can be seen from Table 3, the application has been considered very easy to be used in both rounds. The SUS average score was 86.88 for R1 and 96.75 for R2, well above the minimum score of 68, above which the system usability is evaluated positively (Lewis & Sauro, 2018). We did first compute the results pooling together the children of the two groups. This way we obtained very high values (for example the SUS score for the full group was 92.21 ± 7.32). However, we noticed that we were losing the differences between the two populations. For instance, children in R1 had some difficulties in switching from Drag&Drop interaction modality to Click and vice versa and they found the activity of Feature Unveil very challenging because of many distracting elements (cf. Table 4).

Given the small sample size and the likelihood that the data may not be normally distributed, we opted for the Mann-Whitney U test, which does not require normality and is suitable for ordinal data. Moreover, each child involved in the study completed the tasks independently.

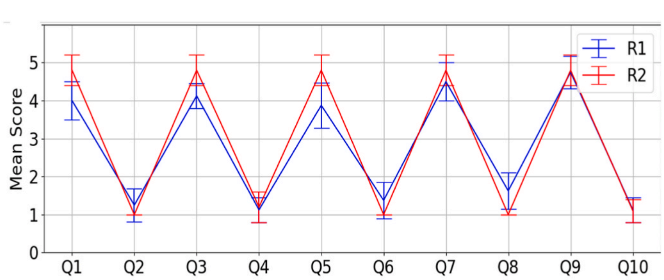

Fig. 14. Mean and Standard Deviation of SUS scores, R1 is plotted in blue and R2 in red.

Null hypothesis: There is no difference in scores between the two groups. The Mann-Whitney U test on SUS scores ($U = 8.0$, $p = 0.00468$) demonstrates a statistically significant difference between the two populations, with R1 showing a lower mean (Figs. 14 and 15). Mann-

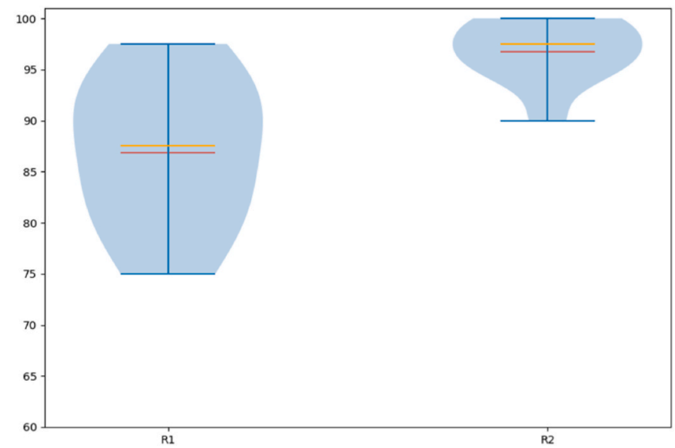

Fig. 15. Distribution of SUS scores for R1 and R2 described by violin plots: mean is plotted in red and median in yellow.

Whitney U was computed using SciPy function .[11] This difference is mainly attributed to Feature Unveil activity, that was considered too difficult by children of R1, who, being at a lower functioning level with respect to those in R2, found the game stressful (this is consistent with the answer to Q14 in Fig. 16a that assesses the change in stress level of the child using MiEMO). Consequently, the therapist gave a lower score for children CH1, CH3 and CH6. Moreover, this exercise required the therapist to intervene helping the child in the interaction.

The scores given to the other usability questionnaires are consistent with the SUS ones (Fig. 16). Indeed, in **R2** a score, averaged on all children, greater or equal 4 to each positive item and lower or equal to 1.5 to each negative item was given. In R1 although the system was rated as useable, scores were lower with average scores to positive items in the range [3–4.5] and on negative scores in the range [1–2.5]. All the Mean and Standard deviation values were computed using Numpy[12] and plotted with Matplotlib,[13] as well as the violin[14] and box plot.[15]

Both in R1 and in R2 the therapists gave a low score to the appropriateness of the difficulty levels (item 30, Fig. 16c), for two different reasons (R1 - IQR = 2.0, Median = 3.0, Mean = 3.0 |R2 - IQR = 1.0, Median = 4.0, Mean = 4.1). If in R1, the low score was mainly due to Feature Unveil being considered too difficult, in R2, this was due to the terminology used: one therapist gave score 1 as she considered inappropriate the levels' denomination: "easy, medium, hard" as this could induce performance anxiety to the child. She suggested to denominate the three levels simply "level 1, level 2 and level 3".

Finally, regarding children's acceptance (Fig. 16d), in R1 few therapists gave a low score to specific items: three therapists scored 3 to the complexity of use (Q32) leading to an overall score of Mean = 2.125, Median = 2.0, IQR = 1.25. These therapists would have liked that, as soon as the child chooses the correct answer, the positive reinforcement is played without pressing the verify button. In R2 all the items, including Q32, received positive scores, suggesting that the children found the application simple to use.

The 10 children of R2, after completing the easy level, decided to continue to play, feeling comfortable in using the app (Q36: IQR = 0.0, Median = 5.0, Mean = 5.0). 8 children could complete the hard level inside the time allocated for free play, and 2 children could reach the

[11] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mannwhitneyu.html.
[12] https://numpy.org/doc/stable/reference/generated/numpy.mean.html.
[13] https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.errorbar.html.
[14] https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.violinplot.html.
[15] https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.boxplot.html.

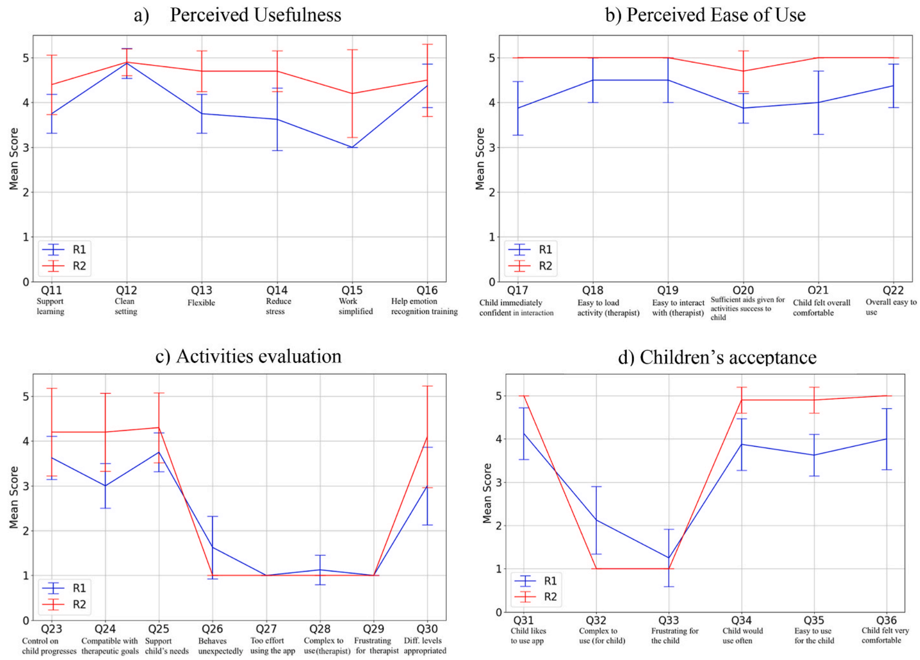

Fig. 16. Mean and Standard Deviation of questionnaires scores. In Blue the scores from R1 and in Red the scores from R2.

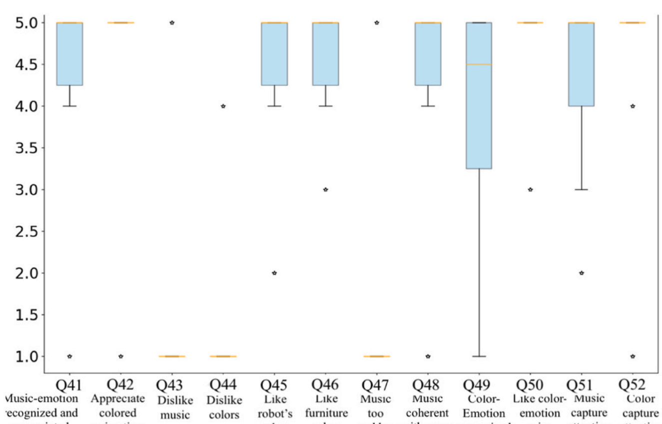

Fig. 17. Plot of the scores on Multi-modal feedback evaluation collected in R2.

end of the medium level. CH12, decided to spend the final part of her free session, rewatching all the Animated Stories first with Tina and then switching to Tino. Multi-modal feedback (i.e., color and music) was evaluated very positive (Q41-52) by all children of R2, who had the chance to experience it (Fig. 17). Only the answers to Association of color to emotion (Q49) exhibits a large dispersion: CH12 and CH15 correctly associated anger, fear and sadness to the corresponding color but could not associate happiness, and CH14 did not noticed colors at all.

4.4. Qualitative analysis

The use of the tablet turned out to be particularly useful for the children with medium-low functioning as they could bring the tablet around while playing. Moreover, some of these children were already accustomed to tablet as they were already using it in traditional training, as some non-verbal autistic children often use it to communicate (for example with applications such as Voca). Reports suggest providing a single modality of interaction with all the activities for children with low-verbal medium-functioning autism as using Drag&Drop (for Matching and Feature Unveil) and click (for all the other activities) may frustrate these children who may try to interact with the application using the wrong modality. All children appreciated "Animated Stories" activities very much, requiring a richer set of stories.

The therapists reported that personalization might improve compliance especially for children with medium-functioning ASC. This could be obtained displaying pictures of the child him/herself instead of pictures of the actor in all the activities, modulating the avatar speech and gesture and personalizing feedback sound effects, since in the pilot CH3 and CH5 liked so much the buzzer sound that they deliberately made mistakes during the play.

The therapists reported high acceptance of music feedback and in particular music feedback associated to change of color in the animated stories. The general impression was that these elements facilitated capturing child's attention, in particular of the youngest children. Colored animations and music feedback were assessed by the therapists in R2 asking to children with verbal high-functioning ASC questions like: "Did you noticed the [current color of Tino]"? or "What do you think about this music?". Most of the children recognized the association between the music, colors and emotion portrayed; for example, one child reported that "this is really the music of anger", other children noticed that "Tino is blue as he is sad now", another one noticed how even the curtains assumed the color of fear in the spider animated story. Moreover, children perceived with great emotional intensity the video of the actor's face expressing an emotion accompanied by the music in Visual Search exercise.

Two children were hyperacustic CH17 and CH14. CH17 asked to turn off the music (with low scores on music items: Fig. 17). The therapist asked if he preferred another music, but he answered "no music". Nevertheless, he did not have any problem with color feedback in automated stories. CH14 was not annoyed by music as he was engaged while playing but later on, when he heard another child playing with the music turned on (in another area of the same room), he asked to turn his music down or to move away. Only one child was not engaged by the colors and barely noticed them (which explain the low scores on color items, Fig. 17); he was instead engaged by music and correctly associated each music piece with the emotion.

The possibility of choosing the dynamic ending of animated stories was appreciated by the therapists, so that each child was able to choose between anger and sadness, which emotion s/he considered most pertinent. Interestingly, CH14 commented that animations were similar to commercials since both are short and attention-catching. For all the children the catchiest animation was the one with a spider falling from the ceiling that scares Tina/o (Fig. 8b). This animation elicited two different reactions: amusing or scaring. For example, one CH9 was amused, and he decided to turn on the music at high volume and to show this animation to the therapist to scare her, while another child found the gestures that Tino was making while running away with its arms raised engaging, saying that "Tino is very scared". CH17 said that he was afraid to look at this animation because he fears spiders and did understand why Tino ran away. Indeed, therapists found that this animation represented an opportunity to talk with all children about what makes them scared. In general, the animated stories have been an important starting point to allow children to talk about themselves to the therapist, leading them to tell more naturally what makes them angry at school or what makes them sad, making new confidences. CH10 said that he felt angry and sad sometimes because his classmates treated him badly and make him angry. CH18, after watching the story of the ice-cream falling (that he considered a sad ending), said that he felt like Tino, because he had to change the school class (which implies no longer seeing his best friend) and he thought that it was all his fault. The same child had difficulty also seeing the video of the actor expressing anger in the Visual Search (Fig. 6b) and he skipped the video: the therapist explained that the child tried to ignore the bad emotions instead to elaborate them. This therapist took the opportunity to deepen this topic with the child and work on it. He also suggested him to go out and have an ice cream with his best friend and that he will see the friends outside school.

The animation on happiness was very enjoyed by making the children dream about what makes them happy. CH10 said that he likes to make jokes and that playing with jack-in-the-box makes him happy (Fig. 8a). CH18 dreamed on what could be in the gift represented in the animation that Tina gives to Tino as "a Maserati car" or a "brand new iPhone". All the children asked to see more animations on positive emotions. As animated story express three emotions with negative valence (anger, fear or sadness) and only one with positive valence (joy), it was suggested by therapists of both R1 and R2 to increase the number of stories expressing positive emotions.

Two children, CH3 and CH5, 5 and 6 years old respectively, were making the same face expression of the protagonist of the animated stories (for example they frowned when Tino became sad).

Finally, most of the children in R2 focused on the morphing animations in the Visual Search (Fig. 6b) for a long time, reloading them and commenting with phrases such as "this guy expresses fear with great intensity" or "this is the face of when you for the first time see all 4 in the school report card [fear] (cf., Fig. 18) and you have to tell to the parents", instead "this is the face of when you see all nine on the school report card [happiness]". CH9 said he liked reloading the morphing animation to see it again the expression of emotions since "it's my favorite thing".

This kind of intervention can also have a role in improving children socialization. CH8 (aged 6) was playing with the application on a gym-mat while a peer was doing gymnastics in a distant area of the gym. When the girl started watching the Animated Stories, the other child attracted by the music, approached the girl curious. When he saw that she was watching an Animated Story, he asked her if he could watch it with her. She agreed and they both sat on the gym-mat, sharing the tablet and watching all the animations together.

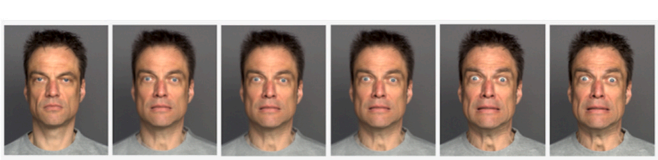

Fig. 18. Example of fear animation of a male actor with morphing (subset of frames).

5. Discussion

We have explored the usability strengths and weaknesses of digital training with MiEmo application, evaluating the usability of with novel digital exercises designed for emotion recognition, combined with multi-modal feedback.

In fact, music perception is currently investigated in ASC domain, providing evidence that music is highly appreciated by children with ASC, that benefit from music therapy (Ke et al., 2022). We show here how integration of music tuned to emotions inside ICT mediated intervention has the potentiality to largely increase the feedback, making the intervention more effective. This was even more increased by integrating also change in ambient colors tuned to current emotions and gestures of the avatar (Ludlow et al., 2008).

Finally, different studies on school children demonstrated that short video could enhance learning (e.g. (Altinpulluk et al., 2020),), reducing the cognitive load, with respect to the long-lasting video lessons. In line with these results, we kept the duration of animated stories inside less than 1:30 min realizing the equivalent of short animation-pills, similar to animations on television, as also stated by a CH14 that compared Animated Stories to engaging commercial spots. This approach differs from that of (Fridenson-Hayo et al., 2017) where the video was designed as lesson to explain how to recognize social cues.

5.1. Emotion representation in MiEmo

The difference in children population in R1 and R2 produced slightly different outcomes in terms of usability, acceptability, and engagement. The application was deemed useable and effective for children with high-functioning autism, with activity difficulty rated as appropriate for this group. However, therapists of children with medium-functioning autism suggested updates to better suit their needs. Some activities, like "Feature Unveil" were considered too difficult, even at the easiest level. Over time, the use of other people's photos could be introduced. Therapists also emphasized the importance of incorporating negative emotions like anger, fear, and sadness to help children navigate these feelings. However, they suggested a better balance of positive and negative emotions, as the current version underrepresents positive emotions, which may increase stress in some children, as seen with CH19, who found the training too intense after an argument with his mother.

5.2. Music feedback

Indeed, studies on the role of music in both videogames and serious games have demonstrated the importance of music in (a) enhancing the emotion felt by the user (e.g. enjoyment or shock, (Millet et al., 2021), (Klimmt et al., 2019)) and (b) improving learning skills. Different pedagogical studies (Holmes, 2017), (Green, 2017) have established that music (both with and without words) improve intellectual performance (especially memory), and spatial-temporal abilities. Here, the emotional content of music is aligned to learning focus, that is emotion itself, that is an innovative use of musing feedback.

This was used here in two modalities. In the first modality, it is aimed at increasing feedback provided in all the activities except the Animated Stories. Indeed, children recognized the match between the emotion portrayed on screen and the specific music piece by naming the emotion to the therapist. In particular, high functioning children in R2 perceived easily the emotion portrayed by the video of the actor's face accompanied by the music at the end of the Visual Search. This result is interesting because usually ASC children have a difficulty in recognizing faces with emotions (Uljarevic & Hamilton, 2012), especially when those are very realistic. Even though this is a simple experimental observation (which should be supported by a dedicated study) it somehow fits with previous studies demonstrating similarities between visual and acoustical structures in music and the visual arts (Actis-Grosso et al., 2017), (Limbert & Polzella, 1998). As second modality, the same music piece tuned to the portrayed emotion is played as soundtrack in Animated Stories (Section 4.4).

This study preliminary findings fits with studies on emotion recognition in music that is often not impaired in ASC. Indeed, the introduction of music during socio-emotional can be powerful tool to support children in the recognition of emotions from traditional social cues such as face and gestures, opening new possibilities of training with music.

5.3. Colored animations

Not only music, but also colors can enhance memory abilities (cf. Boyd, 2019): a colored design supports learners to process materials cognitively, and influences readers' emotions by motivating them. The process to induce learners to look at particular colored content (Plass et al., 2014) may be even more important in autism, as is it amply documented that autistic people memorize information mainly through visual thinking.

We have introduced color changes in "What does Tina/o feel?" and Animated Stories activities that are most suitable to this. Indeed, some children verbalize this explicitly: for instance, CH10 stated that he did not like blue as Tino is sad when it becomes blue. Our results are consistent also with the studies in literature in the domain of robotic agents on how colored LEDs can enhance emotion recognition (Casiddu et al., 2021), in particular with studies on the relation between colors and the emotions, e.g., the participants to the study (Johnson et al., 2013) recognized the colored patterns of red-anger, yellow-happiness and blue-sadness.

No child was found insensitive to both colors and music. Only one child did not associated colors with the perception of emotions, but he was not even bothered by the presence of the colored elements on the screen, he simply barely noticed them. Nevertheless, this child was engaged by music pieces. A second child was annoyed by music, but he appreciated the colored animations. Indeed, as shown by (Cermak et al., 2010), children with ASC had variable sensory preferences: some children may not be engaged by colors or by music. Indeed, providing both color and music feedback we can match the different preferences of ASC children and thus provide multi-modal feedback to a larger part of ASC children. Following their different sensorial preferences, only colored animations can be selected by children who dislike music and only music feed-back by children who do not notice or dislike colors. These preliminary observations should be deepened in future studies involving more children to access if sensory preferences of all ASC children are respected, and to deepen on how we can improve the application to maximize efficacy and to personalize feedback to fit preferences of all the children.

5.4. Animated stories

Music soundtrack in films indeed enhances the emotion perception as well as the positive/negative role of actors (Marshall & Cohen, 1988). In the same way different studies demonstrated that the selection of colors in movies influence audience's emotions, attitudes, reactions, and interpretations of the plot (Bellantoni, 2012). We introduce this inside Animated Stories to help children feeling emotions of the virtual character. This feedback allowed also to improve the child's understanding of the short plot, that was built with three main sections: (i) avatar's initial state (calm or happy depending on the story), (ii) event and (iii) avatar's emotional reaction. In fact, most of children not only recognized the emotions portrayed but they empathized with Tino/a and instinctively expressed with their own face of the same emotion expressed by Tino/a. This reinforcement by music and colors, was expressed even more vividly by verbal children as most of them started to talk with the therapist about their private experiences when they felt the same emotion portrayed on the screen (Section 4.4). The confidences on negative emotions were intimate and deep.

We can assume that the animations with Tino/a deeply involved the children that identified with the protagonist, helping them to remember and tell the therapist when they feel a certain emotion in daily life, this allowed the therapist to explain the child how to manage these emotions and how to react properly.

Finally, music soundtrack has also the potentially to improve socialization: a child not participating in this study (that was doing gymnastic) got close to CH8, attracted by music, and watched animated stories together. We can assume that it happened due to the engagement of the music pieces, as expressed in literature by (Millet et al., 2021), (Klimmt et al., 2019). Since this child was engaged by the music listened from another area of the gym. Then the animated stories captured the attention of both children that seated together nearby enchanted by the animations. This is an interesting result since it shows that our tool could also help improving socialization for those children with a more severe level of autistic condition.

Further studies should be performed focusing on the potentials of social interaction of animated stories on emotions: as first step more stories should be developed on both negative and positive primary emotions, in this way the child would have access to a wider set of animations to watch together with peers during training; second a study should be held to observe children's behavior and engagement in watching animated stories with peers, exploring if effectively these animations can enhance also socialization, and if adaptations to MiEmo are needed to support socialization.

6. Conclusion

MiEmo has shown to be highly useable and, in line with the current research, it reinforces the view that digital platforms have a large potentiality to support ASC intervention. Children with high-functioning ASC, were easily engaged in the activities, while children with medium-functioning ASC required occasional therapist support, particularly with most challenging tasks like Feature Unveil.

The innovative combination of classical emotion training with multimodal feedback (music, colors and gestures) was well-received: Animated Stories, in particular, revealed to be a potentially powerful tool to enhance emotion perception for most children; moreover, it can be an instrument to help children express their personal emotional experiences and to foster socialization, with children interacting together while watching animated stories.

While this study provides valuable insights into the potential of the MiEmo platform for improving socio-emotional training, in children with Autism Spectrum Condition (ASC), several limitations must be acknowledged. First, the relatively small sample size of 19 participants limits the generalizability of the findings. On the other side, this preliminary study focuses on the usability of the platform, exploring its fit with the sensory preferences of children and the training needs of the therapists. Future studies with larger, more diverse participant groups are necessary to assess long-term effects and determine whether improvement is sustained over time. The absence of a control group also limits causal conclusions about the platform's effectiveness at this stage. Finally, while the study offers promising short-term outcomes, longitudinal studies are essential to evaluate the MiEmo platform's impact on emotion recognition and social skills in everyday life.

CRediT authorship contribution statement

Eleonora Chitti: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rossana Actis-Grosso:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Formal analysis, Data

curation, Conceptualization. **Paola Ricciardelli:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Benedetta Olivari:** Resources, Methodology, Investigation, Conceptualization. **Cecilia Carenzi:** Resources, Methodology, Investigation, Conceptualization. **Mariantonia Tedoldi:** Resources, Methodology, Investigation, Conceptualization. **N. Alberto Borghese:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: This work was partially supported by University Grants: H2020 EU project ESSENCE (grant n. 101,016,112) and by PRIN2022 MIUR Italian project AIRCA (grant n. G53D23002860006). All authors declare that they have no conflicts of interest. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was partially supported by H2020 project ESSENCE (grant n. 101016112) and by PRIN2022 MIUR Italian project AIRCA (grant n. G53D23002860006).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbr.2024.100549.

Data availability

Data will be made available on request.

References

Actis-Grosso, R., Bossi, F., & Ricciardelli, P. (2015). Emotion recognition through static faces and moving bodies: A comparison between typically developed adults and individuals with high level of autistic traits. *Frontiers in Psychology, 6*(OCT), 1–14. https://doi.org/10.3389/fpsyg.2015.01570

Actis-Grosso, R., et al. (2017). Can music be figurative? Exploring the possibility of crossmodal similarities between music and visual arts. *Psihologija, 50*(3), 285–306. https://doi.org/10.2298/PSI1703285A

Altinpulluk, H., Kilinc, H., Firat, M., & Yumurtaci, O. (2020). The influence of segmented and complete educational videos on the cognitive load, satisfaction, engagement, and academic achievement levels of learners. *Journal of Computers in Education, 7*(2), 155–182. https://doi.org/10.1007/s40692-019-00151-7

Baron-Cohen, S. (2006). The hyper-systemizing, assortative mating theory of autism. *Progress In Neuro-Psychopharmacology & Biological Psychiatry, 30*(5), 865–872. https://doi.org/10.1016/j.pnpbp.2006.01.010

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). *Does the autistic child have a 'theory of mind'?*.

Baron-cohen, S., Spitz, A., & Cross, P. (1993). Do children with autism recognise surprise? A research note. *Cognition & Emotion, 7*(6), 507–516. https://doi.org/10.1080/02699939308409202

Bauminger, N., & Kasari, C. (2000). Loneliness and friendship in high-functioning children with autism. *Child Development, 71*(2), 447–456. https://doi.org/10.1111/1467-8624.00156

Bellantoni, P. (2012). *If it's purple, someone's gonna die: the power of color in visual storytelling*. Routledge.

Belmonte, M. K., & Yurgelun-Todd, D. A. (2003). Functional anatomy of impaired selective attention and compensatory processing in autism. *Cognitive Brain Research, 17*(3), 651–664.

Boyd, L. (2019). Designing sensory-inclusive virtual play spaces for Children. *Proceedings of the 18th ACM international conference on interaction design and children, IDC 2019*.

Brooke, J. (1996). SUS: A quick and dirty usability scale system usability scale view project useable systems view project. *Usability evaluation in industry, 189*(194), 4–7.

Cabibihan, J.-J., Javed, H., Ang, M., & Aljunied, S. M. (2013). Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. *Int J Soc Robot, 5*(4), 593–618.

Campbell, A. (2018). Face morph. UC Berkeley. [Online]. Available: https://inst.eecs.berkeley.edu/~cs194-26/fa18/upload/files/proj4/cs194-26-adf/. (Accessed 4 March 2024).

Casiddu, N., et al. (2021). Designing synthetic emotions of a robotic system. In *Advances in intelligent systems and computing* (pp. 148–155). https://doi.org/10.1007/978-3-030-58282-1_24

Cermak, S. A., Curtin, C., & Bandini, L. G. (2010). Food selectivity and sensory sensitivity in children with autism spectrum disorders. *Journal of the American Dietetic Association, 110*(2), 238–246.

Chang, B., Xu, R., & Watt, T. (2018). *The impact of colors on learning*.

Chevalier, P., et al. (2017). Impact of sensory preferences of individuals with autism on the recognition of emotions expressed by two robots, an avatar, and a human. *Autonomous Robots, 41*(3), 613–635.

Cowen, A. S., Fang, X., Sauter, D., & Keltner, D. (2020). *What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures, 117*(4).

De Capitani, G. (2019). *Prototype of app user interface as support to rehabilitation training for autistic individuals: recognition of emotions and facial expressions*. Supervisor: R. Actis-Grosso, MSc Thesis, Department of Psychology, University of Milan-Bicocca.

Della-Torre, M. E., Zavagno, D., & Actis-Grosso, R. (2021). The interpretation of E-motions in faces and bodies derived from static artworks by individuals with high functioning autistic spectrum. *Vision, 5*(2).

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). Faces - a database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods, 42*(1), 351–362. https://doi.org/10.3758/BRM.42.1.351

Ekman, P. (2009). Darwin's contributions to our understanding of emotional expressions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1535), 3449–3451.

Ekman, P., & V Friesen, W. (1978). *Facial action coding system*.

Franklin, A., Sowden, P., Burley, R., Notman, L., & Alder, E. (2008). Color perception in children with autism. *Journal of Autism and Developmental Disorders, 38*(10), 1837–1847. https://doi.org/10.1007/s10803-008-0574-6

Fridenson-Hayo, S., et al. (2017). Emotiplay': A serious game for learning about emotions in children with autism: Results of a cross-cultural evaluation. *European Child & Adolescent Psychiatry, 26*(8), 979–992.

Gebauer, L., et al. (2014). Intact brain processing of musical emotions in autism spectrum disorder, but more cognitive load and arousal in happy vs. sad music. *Frontiers in Neuroscience*, (8).

Grandin, T. (2018). Chapter 1: Autism and visual thought. In *Thinking in pictures* (Expanded Edition).

Green, L. (2017). *Music, informal learning and the school: A new classroom pedagogy*. Routledge.

Holden, R. J., & Karsh, B. T. (2010). *The Technology Acceptance Model: Its past and its future in health care*.

Holmes, S. (2017). *The impact of participation in music on learning mathematics*. UCL. PhD Thesis.

Huggins, C. F., Donnan, G., Cameron, I. M., & Williams, J. H. G. (2020). Emotional self-awareness in autism: A meta-analysis of group differences and developmental effects. *Autism, 25*(2), 307–321.

Ismail, U. U., Ramli, R., & Rozzani, N. (2021). User experience/user interface (ux/UI) design for autistic spectrum disorder (asd) color based emotion detection system: A review. In *2021 IEEE international conference on automatic control & intelligent systems (I2CACIS)* (pp. 114–119).

Janzen, T. B., & Thaut, M. H. (2018). Rethinking the role of music in the neurodevelopment of autism spectrum disorder. *Musicae Scientiae, 1*. https://doi.org/10.1177/2059204318769639

Jaswal, V. K., & Akhtar, N. (2019). Being versus appearing socially uninterested: Challenging assumptions about social motivation in autism. *Behavioral and Brain Sciences, 42*.

Jayashankar, A., & Aziz-Zadeh, L. (2023). Disgust processing and potential relationships with behaviors in autism. *Current Psychiatry Reports, 25*(10), 465–478. https://doi.org/10.1007/s11920-023-01445-5

Johnson, C. I., Bailey, S. K. T., & Van Buskirk, W. L. (2017). Designing effective feedback messages in serious games and simulations: A research review. In *Instructional techniques to facilitate learning and motivation of serious games* (pp. 119–140). Springer International Publishing.

Johnson, D. O., Cuijpers, R. H., & van der Pol, D. (2013). Imitating human emotions with artificial facial expressions. *Int J Soc Robot, 5*(4), 503–513. https://doi.org/10.1007/s12369-013-0211-1

Junaidi, A. R., et al. (2022). Usability testing of emotion recognition assistance application for children with autism spectrum disorder. In *8th int. Conf. On Education and technology IEEE* (pp. 309–314). ICET.

Juslin, P. N. (2019). *Musical emotions explained: Unlocking the secrets of musical affect*. OUP Oxford.

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research, 33*(3), 217–238.

Karkhaneh, M., Clark, B., Ospina, M. B., Seida, J. C., Smith, V., & Hartling, L. (2010). *Social StoriesTM to improve social skills in children with autism spectrum disorder: A systematic review*.

Katagiri, J. (2009). The effect of background music and song texts on the emotional understanding of children with autism. *Journal of Music Therapy, 46*(1), 15–31. https://doi.org/10.1093/jmt/46.1.15

Ke, X., Song, W., Yang, M., Li, J., & Liu, W. (2022). *Effectiveness of music therapy in children with autism spectrum disorder: A systematic review and meta-analysis*. Frontiers Media S.A.

Kiili, K., De Freitas, S., Arnab, S., & Lainema, T. (2012). The design principles for flow experience in educational games. *Procedia Computer Science, 15*, 78–91. https://doi.org/10.1016/j.procs.2012.10.060

Klimmt, C., Possler, D., May, N., Auge, H., Wanjek, L., & Wolf, A.-L. (2019). Effects of soundtrack music on the video game experience. *Media Psychology, 22*(5), 689–713.

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience, 15*(3), 170–180.

Lewis, J., & Sauro, J. (2018). *Item benchmarks for the system usability scale* (Vol. 13, pp. 158–167).

Limbert, W. M., & Polzella, D. J. (1998). Effects of music on the perception of paintings. *Empirical Studies of the Arts, 16*(1), 33–39. https://doi.org/10.2190/V8BL-GBJK-TLFP-R321

Ludlow, A. K., Wilkins, A. J., & Heaton, P. (2008). Colored overlays enhance visual perceptual performance in children with autism spectrum disorders. *Res Autism Spectr Disord, 2*(3), 498–515.

Maenner, M. J., et al. (2020). Prevalence of autism spectrum disorder among children aged 8 Years — autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *MMWR Surveillance Summaries 2020, 69*(4), 16. https://doi.org/10.15585/mmwr.mm6916a4

Marshall, S. K., & Cohen, A. J. (1988). Effects of musical soundtracks on attitudes toward animated geometric figures. *Music Perception: An Interdisciplinary Journal, 6*(1), 95–112. https://doi.org/10.2307/40285417

Mazzoni, N., et al. (2020). Motion or emotion? Recognition of emotional bodily expressions in children with autism spectrum disorder with and without intellectual disability. *Frontiers in Psychology, 11*(Mar).

Mehlenbacher, B. (2010). *Instruction and technology: Designs for everyday learning*. MIT Press.

Millet, B., Chattah, J., & Ahn, S. (2021). Soundtrack design: The impact of music on visual attention and affective responses. *Applied Ergonomics, 93*, Article 103301. https://doi.org/10.1016/j.apergo.2020.103301

Nakajima, K., Minami, T., & Nakauchi, S. (2017). Interaction between facial expression and color. *Scientific Reports, 7*.

Nielsen, J. (1993). Iterative user-interface design. *Computer (Long Beach Calif), 26*(11), 32–41.

Panda, R., Malheiro, R., & Paiva, R. P. (2023). Audio features for music emotion recognition: A survey. *IEEE Trans Affect Comput, 14*(1), 68–88. https://doi.org/10.1109/TAFFC.2020.3032373

Pirovano, M., et al. (2016b). Exergaming and rehabilitation: A methodology for the design of effective and safe therapeutic exergames. *Entertain Comput, 14*, 55–65.

Pirovano, M., Mainetti, R., Baud-Bovy, G., Lanzi, P. L., & Borghese, N. A. (2016). Intelligent game engine for rehabilitation (IGER). *IEEE Trans Comput Intell AI Games, 8*(1), 43–55.

Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction, 29*, 128–140.

Quintin, E.-M., et al. (2011). Emotion perception in music in high-functioning adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 41*(9), 1240–1255. https://doi.org/10.1007/s10803-010-1146-0

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178.

Sánchez, J. L. G., Vela, F. L. G., Simarro, F. M., & Padilla-Zea, N. (2012). Playability: Analysing user experience in video games. *Behaviour & Information Technology, 31*(10), 1033–1054.

Schell, J. (2019). *The art of game design: A book of lenses* (3rd ed.). CRC Press.

Scherer, K. R. (2005). *What are emotions? And how can they be measured?*.

Serret, S., et al. (2014). Facing the challenge of teaching emotions to individuals with low- and high-functioning autism using a new serious game: A pilot study. *Molecular Autism, 5*(1), 1–17.

Smagorinsky, P. (2016). An autistic life, animated through the world of Disney: A loving autoethnography. In P. Smagorinsky (Ed.), *Creativity and community among autism-spectrum youth: Creating positive social updrafts through play and performance* (pp. 269–291). New York: Palgrave Macmillan US.

Takahashi, F., & Kawabata, Y. (2018). The association between colors and emotions for emotional words and facial expressions. *Color Research & Application, 43*(2), 247–257. https://doi.org/10.1002/col.22186

Tang, J. S. Y., et al. (2019). A systematic review and meta-analysis of social emotional computer based interventions for autistic individuals using the serious game framework. *Research in Autism Spectrum Disorders, 66*.

Uljarevic, M., & Hamilton, A. (2012). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders, 43*. https://doi.org/10.1007/s10803-012-1695-5

Völlm, B. A., et al. (2006). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *NeuroImage, 29*(1), 90–98.

Wardani, N. K., & Mustikasari, B. F. (2023). Colour preference on picture therapy cards in children with ASD. *KnE Social Sciences, 8*(15). https://doi.org/10.18502/kss.v8i15.13938

Windl, H., & Constantine, L. (2003). Usage-centered design: Scalability and integration with software engineering. *Human-Computer Interaction: Theory and Practice, 1*, 449–543.

Xie, F., Pascual, E., & Oakley, T. (2023). Functional echolalia in autism speech: Verbal formulae and repeated prior utterances as communicative and cognitive strategies. *Frontiers in Psychology, 14*.

Zhang, Q., et al. (2021). Facial emotion training as an intervention in autism spectrum disorder: A meta-analysis of randomized controlled trials. *Autism Research, 14*(10), 2169–2182.

Zoerner, D., Schütze, J., Lucke, U., Kirst, S., & Dziobek, I. (2016). Zirkus Empathico mobile training of socio-emotional competences for children with autism, mobile app, socio-emotional competence, autism. In *Proc. - IEEE 16th international conf. On advanced learning technologies, ICALT* (pp. 448–452).

Rego, P., Moreira, P. M., & Reis, L. P. (2010). *Serious games for rehabilitation: A survey and a classification towards a taxonomy. Proc. of the 5th* (pp. 1–6). Spain: Technologies CISTI, Santiago de Compostela.

Diagnostic and statistical manual of mental disorders (DSM-5®). (2013). American Psychiatric Publishing.