Human Behavior and Emerging Technologies

WILEY

RESEARCH ARTICLE OPEN ACCESS

Improvisational Music Therapy in the Metaverse: Assessing the Feasibility of an Immersive VR Support System

Christian Corrò[1] | Luca Chittaro[1] | Eva Pupo[2] | Donatella Bellomo[3] | Stefano Navone[2]

[1]Human-Computer Interaction Lab, Department of Mathematics, Computer Science and Physics, University of Udine, Via delle Scienze 206, Udine 33100, Italy | [2]Music Therapy School InArteSalus, Via Pasini 67/C, Schio 36015, Italy | [3]Music Therapy School Stratos, Viale Kennedy 53, Bari 70124, Italy

Correspondence: Christian Corrò (christian.corro@uniud.it)

Received: 10 March 2025 | **Revised:** 10 February 2026 | **Accepted:** 12 February 2026

Academic Editor: Ana Junça Silva

Keywords: immersive virtual reality | metaverse | music therapy

ABSTRACT

Improvisational music therapy, a form of art therapy that relies on using musical instruments in a shared space with a music therapist, currently lacks suitable digital tools for conducting remote sessions. This limitation hinders the possibility of conducting the therapy in conditions where in-person sessions are impractical. For example, during the COVID-19 pandemic, due to social distancing measures, many improvisational music therapy programs were suddenly halted, compromising users' progress and severely limiting the work of music therapists. This study contributes to the literature by introducing the first system to support remote improvisational music therapy with immersive virtual reality in a metaverse context. The system has been designed and validated by a multidisciplinary team of computer scientists, music therapists, and clinical psychologists. The results of the user study indicate that the system provides the conditions necessary for effective nonverbal communication and musical interplay within an immersive virtual environment. We also identify further improvements to enhance the fidelity of nonverbal communication, as it is the fundamental relational component of improvisational music therapy. This work shows the potential of immersive virtual reality and the metaverse to enable remote improvisational music therapy sessions, providing a foundation for future clinical validation studies.

1 | Introduction

Music therapy (MT) is defined by the World Federation of Music Therapy as the systematic use of music and its elements, for example, sound, rhythm, melody, and harmony, by a certified music therapist working with individuals to facilitate and promote communication, relationships, expression, and other relevant therapeutic objectives in order to address physical, emotional, mental, social, and cognitive needs [1]. MT is administered across diverse contexts, such as mental health, medical, community, developmental, and educational contexts [2, 3]. MT methods are organized in two categories: *active* and *receptive*. Active methods, including improvisation, involve users in creating music. In contrast, receptive MT involves users in listening to music and includes verbal processing of feelings and experiences [4]. Essential elements of an MT intervention, whether active or receptive, comprise the presence of a certified music therapist, involvement of users, structured musical interaction, and a structured therapeutic framework, which includes the setting, that is, the environment, in which the MT session occurs. These elements are necessary for the practice to be referred to as MT [2]. For example, the practice of medical professionals providing listening to pre-recorded music, for example, for helping a patient relax before a surgical procedure, is referred to as "music medicine" rather than MT [5]. Improvisation is the main active MT method used in adult mental health

[2, 6]. Improvisational methods include any experience in which the user actively participates in spontaneous music-making with the music therapist through sonorous–musical interactions, that is, playing musical instruments and vocalizing [2]. The use of verbal language is not a primary component: The focus is on nonverbal communication and expression through music creation and interaction. The purpose of improvisation is not artistic or aesthetic but rather relational and communicative; for this reason, it is not necessary for the user to have any practical or theoretical musical skills [7]. In improvisational MT (IMT), especially with a psychodynamic–relational approach [8], sharing a space, that is, the setting, between the user and the music therapist plays a crucial role [7]. The setting facilitates the exchange of nonverbal cues, such as posture, facial expressions, gaze, and proxemics, and supports interaction with musical instruments that can be shared. Thanks to the interaction with musical instruments in the setting, it is possible to establish the expressive dimension of *interplay*, which refers to a musical and expressive interaction based on an exchange of actions, where each performer influences the other [7, 9]. IMT aims to foster a dyadic relationship between the music therapist and the user, shaped through interplay. This relationship develops gradually, and its intensity varies within moments and with varying degrees of *affect attunement,* as the dyad engages in co-creation and musical exchange [10–12]. Affect attunement refers to the process by which two individuals share and respond to each other's affective states [13]. The exchange of nonverbal cues through sonorous–musical interaction and embodiment is the key element for nonverbal communication that facilitates emotional coregulation and affect attunement [13, p. 66]. Sharing affective states is crucial for establishing an *intersubjective dimension* between the music therapist and the user [13, 14]. The intersubjective dimension refers to the experience of "mental contact" that interacting subjects can establish during any form of interpersonal communication [15]. In IMT, the user is free to create sonorous–musical production, opening a nonverbal communicative channel with the music therapist. In response, the music therapist can start a sonorous–musical production in an attempt to initiate a nonverbal dialog with the user [12]. Throughout the sonorous–musical interaction, the dyad engages in a process of co-creation and shared expression, unfolding within an implicit relationship that emerges spontaneously and bypasses higher order cognitive processes [12, 16]. In this way, IMT can also be effective in therapeutic contexts where the user has difficulty or is unable to use verbal communication [16]. In summary, IMT is based on processes of affect attunement and aims to develop an intersubjective dimension between the music therapist and the user through a nonverbal communication channel to pursue therapeutic objectives.

However, co-presence in a shared space with manipulation of musical instruments that can be shared, which are essential requirements to conduct an IMT session [7], cannot be replicated in a video call when physical in-person sessions are not feasible. This makes it impossible to conduct remote IMT, preventing the initiation or continuation of treatments in scenarios where in-person therapy is not viable. For example, the COVID-19 pandemic had significantly disrupted IMT. During the COVID-19 pandemic, people needed to adapt to social distancing and lockdown measures. In this context, psychological interventions that rely on dialog, such as psychotherapy, had temporarily transitioned to digital platforms to ensure continuity of care, primarily using video calls as the means of interaction between therapists and clients [17–20]. However, the inherent limitations of conducting IMT through video calls led to an abrupt interruption of treatments, thus compromising users' progress and also severely limiting employment opportunities for music therapists. The challenges for IMT are not limited to pandemic-related restrictions. For example, in rural areas, where geographical remoteness often restricts access to MT programs [21], remote IMT would help mitigate access barriers. Additionally, in infectious disease wards, where isolation rules limit physical contact, or with severely immunocompromised patients, remote IMT would allow therapy without risking safety. These are just a few scenarios that underscore the importance of coming up with innovative solutions that could make IMT possible even if physical in-person sessions are not feasible.

In a metaverse context, the music therapist could interact in real time through avatars in the same virtual environment (VE), experienced through immersive virtual reality (VR), and interact with virtual musical instruments (VRMIs) together. VR could help to expand the accessibility of IMT, reaching people who would otherwise be excluded from traditional IMT. People with conditions such as hikikomori, severe social anxiety, infectious diseases, or motor impairments that prevent them from attending in-person therapy sessions can benefit from remote IMT. By removing physical and geographical barriers, VR could provide a safe environment where users can interact with music therapists and actively participate in IMT without the limitations imposed by health issues or mobility difficulties. The two theoretical concepts that represent a potential intersection between MT and the metaverse are effective attunement and sense of presence. Sense of presence refers to the psychophysiological experience of "being there" in a VE, where users feel immersed as if they were physically present in that space [22]. Sense of presence could play a crucial role in facilitating affect attunement, even in the absence of physical co-presence. Research suggests that VR can evoke emotions and psychophysiological responses comparable to those experienced in real-world settings [23]. Therefore, an immersive VE that effectively induces a sense of presence could provide the necessary conditions for affect attunement. Unfortunately, the literature on using the metaverse to conduct MT sessions is very limited, especially for IMT. Most studies focus on the use of single-user VR to expose users to images or videos with selected music [24–26], a receptive MT method. Only two studies address active MT [27, 28], allowing users to generate sounds by interacting with virtual objects. Both studies are seriously limited by the fact that the proposed systems do not support the co-presence of the music therapist in the VE, and thus do not meet an essential requirement of active MT. To the best of our knowledge, no research study has yet focused on using a metaverse approach to support remote IMT sessions where the music therapist and the user can interact in real-time with shared VRMIs in a virtual setting designed by music therapists.

This paper aims to address this gap by proposing the first system that aims to support IMT sessions remotely, providing music therapists with a tool for ensuring the continuity of users' therapeutic journeys in conditions where in-person sessions are not feasible. The system has been designed and evaluated by a multidisciplinary team of computer scientists, music therapists, and clinical psychologists. Figure 1 shows examples of the interactions that users and music therapists can engage in using the proposed system. We first present the system's design and implementation. Then, we report on a preliminary user study that evaluates whether the system can provide the technical and interactional requirements necessary to support IMT sessions. Specifically, we assess whether the system enables co-presence, interaction with VRMIs, and perceived quality of interpersonal relationships between the music therapist and participants. We evaluate whether participants reported experiences are consistent with the relational constructs that underlie therapeutic engagement in IMT, such as motivation, self-efficacy, and perceived reciprocity. Our work establishes technological foundations for remote IMT and identifies specific areas requiring further development before clinical deployment. This work represents a first step in enabling remote IMT by establishing the technological foundations. The insights gathered from both users and music therapists can guide future improvements to the immersive support system.

2 | Related Work

VR is increasingly explored as a tool for psychological interventions thanks to its capability to provide controlled, immersive, and interactive environments, allowing users to engage in effective adaptations of face-to-face interventions [29]. Existing VR interventions provide a range of interactive systems that leverage similar or identical hypothesized mechanisms of action as traditional face-to-face interventions [30], while providing greater flexibility in intervention timing, greater cost-effectiveness, and an increased ability to tailor interventions to individual preferences [29]. Research has consistently demonstrated VR's effectiveness across various psychological interventions, including exposure therapy [31], anxiety disorders [32], stress reduction [33], rehabilitation for cognitive and emotional disorders [34], eating disorders [35], phobias [31], and post-traumatic stress disorder [36]. For a comprehensive review of VR for psychological interventions, see Wiebe et al. [37].

Unfortunately, despite the growing use of VR in psychological interventions, its application in MT remains very limited. Most related studies have focused on employing VR to enhance receptive MT sessions, where users are exposed to images or videos accompanied by selected music. For instance, Perez et al. [26] conducted a study where they evaluated a receptive VR-based MT (VR-MT) system with 16 elderly nursing home patients who had various neurological disorders, including mild mental health issues, moderate cognitive disorders, or depressive symptoms. They exposed the patients to various 360° music videos. Results showed notable spatial presence in patients, regardless of their cognitive or depression levels, with minimal motion sickness, validating the usage of immersive media for MT in elderly patients. Kanehira et al. [25] employed a similar approach, involving 8 undergraduate students, who were immersed in a VE while they listened to different types of music, including their favorite ones. They reported a decrease in stress levels. However, this type of approach does not follow proper MT methodology as it lacks session supervision by a music therapist. Brungardt et al. [24] conducted a study on a 2-day receptive VR-MT intervention for hospitalized patients in palliative care. This intervention involved patients working with a music therapist to create customized soundtracks tailored to their individual preferences and therapeutic needs. After creating the personalized soundtracks, the patients were immersed in nature-based 360° videos where they could listen to their custom soundtracks alongside the nature sounds of the videos. Seventeen participants completed the intervention, reporting high usability and satisfaction. Most described positive emotional and physical responses to VR-MT, with immediate benefits reported like respite from medical circumstances, for example, experiencing escape and peace from diagnostic stress, relaxation, and reduced pain. The study intentionally avoided interactive elements within the VE to simplify use. However, patients' feedback highlighted a desire for a more interactive VE. The receptive VR-MT intervention was feasible, usable, and acceptable for this user group. Lin et al. [28] present a VR-MT system aimed at reducing stress and anxiety. The system was evaluated with 19 participants, showing a reduction in stress. The paper states that the system includes a receptive mode that provides a calming environment with music, and an active mode that allows users to play VRMIs. Unfortunately, the study is significantly limited by the fact that there is no co-presence of a music therapist with the user in the VE. This cannot qualify as MT because, as previously mentioned, MT is not merely the act of listening to or producing music; it is a therapeutic process conducted by a music therapist. It is important to note a significant difference between prior receptive VR-MT work and our approach, which are complementary. Indeed, previous studies, for example, [24], evaluated receptive MT, which differs fundamentally from IMT. While receptive approaches involve listening to preselected music in immersive environments, IMT requires real-time co-presence with a music therapist, active music-making through shared instruments, and bidirectional nonverbal communication through musical interplay. The co-presence of the music therapist is essential for establishing an intersubjective dimension with the user. These requirements present distinct technical and methodological challenges that are not addressed by receptive VR-based MT systems. Our work focuses on a VR system that can deal with those different challenges. Furthermore, none of the previous studies has focused on constructing a setting in a VE. The design of the setting is a crucial component of the IMT methodology, aiming to create an environment that facilitates and enhances users' ability to express themselves.

3 | The Proposed Support System

This section presents the technical design of the system, detailing the components and implementation choices that support the core requirements of remote IMT. The design and implementation of the system aimed at enabling nonverbal communication through interaction with VRMIs and vocalization, utilizing readily available and user-affordable technology, to facilitate the establishment of an intersubjective dimension between the user and the music therapist. The proposed system consists of three key components: two VR applications and a web-based dashboard (Figure 2). The first application is meant for the music therapist's users and allows them to select the gender of their avatar, offers a tutorial to familiarize them with how to interact with VRMIs, and then enables access to the virtual setting for the remote IMT session with the music therapist. The second application, meant for music therapists, incorporates features to help them in session management. It

(a) (b) (c)

FIGURE 1 | Screenshots of an improvisational music therapy session from the user's perspective. The music therapist and the user are as follows: (a) in their individual spaces, both playing their own tongue drums; (b) in the shared space, playing the marimba and cymbal together; (c) sharing the same area of the individual space, playing the glockenspiel together.

includes a wristwatch, visible only to the music therapist, with a countdown timer to track the duration of sessions, and a system for automatically gathering data on the session. The collected data include quantitative relational indicators that facilitate the assessment of the therapeutic process and sessions from a relational perspective by the music therapist. These indicators include features of sonorous–musical productions, such as intensity and rhythmic pulse, as well as interpersonal distance and the direction of gaze, which facilitate the identification of attuned and disattuned behaviors within an IMT session by the music therapist [12]. The collected data can be accessed and visualized through a dashboard after each session. The system was developed in Unity Version 2021.3.14f1. Avatar models (Figure 3) were created with Ready Player Me [38]. For avatars' body movements, we used the inverse kinematic solver VRIK Version 1.9 [39]. To enable real-time multiplayer interaction between the music therapist and the user, we used the Normcore framework Version 2.3.1 [40]. Our two VR applications exchange data through incremental delta updates, for example, if an object's position needs to be shared and the object moves upward, only the y-coordinate is sent to the server rather than the entire position vector. This approach provides efficient data synchronization across the applications and supports the creation of persistent VEs, as the data sent to the server remain stored between sessions [40]. The system was run on two Meta Quest 2 headsets with their controllers. For audio input and output, we used over-ear headphones with a microphone (Sennheiser Urbanite XL).

3.1 | Avatars

As mentioned before, the exchange of nonverbal cues through embodiment is one of the key elements of nonverbal communication. Furthermore, research suggests that employing embodiments in VEs significantly increases both the sense of presence, that is, the feeling of being "there" [41], and social presence, that is, the sense of being with another [42, 43]. Therefore, using an embodiment in the system can aid in creating optimal conditions to foster the development of an intersubjective dimension. An ideal system for remote IMT sessions should include accurate body tracking capable of providing a faithful reconstruction of embodiment, that is, the avatar, including facial expressions, gaze, and body movements. Such comprehensive body tracking could be obtained with costly technologies like face tracking, eye tracking, and external trackers, but this would not be affordable for users. Considering the capabilities of the Meta Quest 2 headset, we focused on three aspects of embodiment: facial expressions, gaze, and body movements. Furthermore, we addressed the locomotion system, focusing on its impact on proxemics.

3.1.1 | Facial Expressions

The first aspect considered is the reconstruction of the avatar's facial expressions. Facial expressions are a crucial source of nonverbal information that contributes to communication [44]. Moreover, in immersive VEs, facial expressions enhance social presence [45]. However, since Meta Quest 2 headset does not support face tracking, we had to rely on the audio captured by the microphone to infer facial expressions when a sound is made, such as vocalization or laughter. To reconstruct the mouth movements of avatars, we employed an audio-based lip-sync system. Specifically, we used Meta's lip-sync [46], which uses a deep neural network to analyze audio from the microphone and synthesize visemes in real-time. These visemes are then used to modify the blend shapes of the avatar face. The lip-sync system also detects laughter allowing to modify the avatar's face accordingly.

3.1.2 | Gaze

The second aspect we considered is gaze. Gaze plays a pivotal role in modulating social interactions and enhancing communication

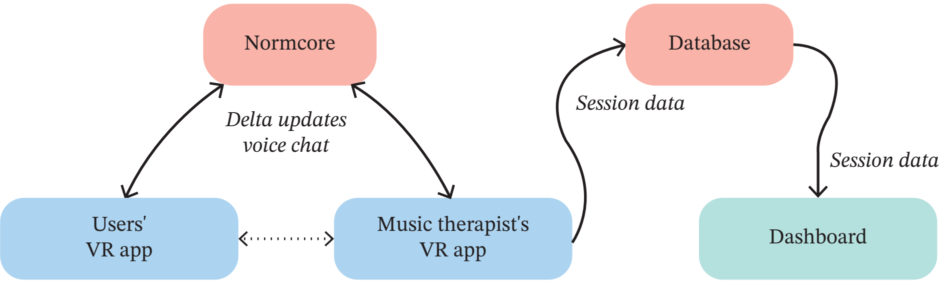


FIGURE 2 | Diagram of the system components.

(a) (b) (c)

FIGURE 3 | Avatars used in the VE. (a) Music therapist; (b) female user; (c) male user.

between individuals [47]. In fact, gaze is a relational indicator that the music therapist uses to assess the session's progress [12]. However, the lack of eye tracking in the Meta Quest 2 limits the accurate measurement of gaze. Given that the absence of gaze can result in less effective social interactions [48], we opted to manipulate gaze rather than maintain it fixed. We thus manipulated the avatar's gaze by using a manually predefined attention map, which gives a weight to each object in the VE. For example, every VRMI is assigned a 0.5 weight, whereas the avatar heads are assigned a weight of 1. Therefore, the system will prioritize directing the gaze toward the other avatar in the VE. The gaze manipulation system selects a gaze target based on object weight, the distance of the object from the avatar, and current head direction of the avatar. Avatar's eyes are then rotated to follow the selected gaze target, within pre-set rotation limits, $\pm 45°$ horizontally and $\pm 30°$ vertically, to ensure the animation remains natural. Finally, to enhance the realism of the avatar's gaze, we added blinking [49].

3.1.3 | Body Movements

The third aspect considered involves the reconstruction of the avatars' body movements. We have examined both upper body movements, including hand gestures, and lower body movements, including leg movements. Body movements, including posture, are relational indicators used by the music therapist to gather information about the user's affective state. Without access to additional trackers beyond the controllers and the headset, it was necessary to reconstruct the avatar's body movement through inverse kinematics. By tracking three points, specifically the head and both hands, it was possible to approximate the movement of the avatar's upper body. The choice between using controllers or hand tracking for hand movements was carefully considered. We assessed both hand tracking and controllers for hand movements in early user testing sessions in terms of precision and expressiveness. Hand tracking allows for a more accurate reconstruction of hand movements, enhancing gesture expressiveness, which could be a source of nonverbal cues in an IMT session. However, hand tracking issues resulting from occlusion or excessively rapid hand movements would compromise the interaction with VRMIs, thus affecting interplay, which is one of the fundamental components for conducting an IMT session. Therefore, we chose to use controllers over hand tracking. This choice avoids the bare hands tracking issues and additionally provides haptic feedback, which can increase the sense of presence. Haptic feedback is modulated based on the collision speed with objects in the VE. Reconstructing the avatar body movements using inverse kinematics works well if tracked points (end effectors) are available [50]. Without end effectors for the feet, the avatar's leg movements cannot be synchronized with the user's movements. For example, if the user lifts a leg, the headset will not detect this movement, making it impossible to reproduce it on the avatar. To prevent breaks-in-presence due to discrepancies between actual movement and its virtual representation, we decided to make the avatar's legs visible only in third-person view, while hiding them in first-person view.

3.1.4 | Locomotion

A locomotion system is essential for enabling system use even in scenarios where the physical space is more constrained than the VE. While teleportation is a locomotion system characterized by usability and motion sickness reduction [51], it provides only instantaneous movement and lacks the capability for continuous, subtle adjustments in the interpersonal distances between the user and the music therapist, which is crucial for signaling trust and engagement in an IMT context. Therefore, we adopted a continuous locomotion system based on joystick translation and real-world rotation [51].

3.2 | Setting and VRMIs

As previously mentioned, in IMT methodology, the setting is a fundamental component of the therapeutic framework. The setting must be constructed aiming to create an environment that facilitates and enhances users' ability to express themselves. IMT setting is constructed by carefully positioning musical instruments within the available space. Musical instruments used in IMT are selected based on the psycho-physical characteristics of the users and the context of the intervention [7]. Percussive musical instruments are typically used to facilitate interaction with musical instruments for patients who have never played a musical instrument or those with physical or cognitive impairments.

3.2.1 | Advantages of VR Setting

Using VR offers considerable advantages in constructing the setting, including customization, stimuli control, isolation from disruptive events, and selecting musical instruments that would be too difficult to use in a real setting. VR enables the creation of customized virtual settings tailored to the specific needs of different users by adjusting the selection and properties of VRMIs, such as their sound characteristics (e.g., tonality and scale), and customizing environmental aspects like materials, colors, and furnishings. This level of control over the setting construction is not possible in in-person sessions, as one must adapt to the available musical instruments and physical spaces. Conducting an IMT session in a VR setting also allows for environmental stimuli control. It is possible to eliminate stimuli that may disturb patients and thus potentially compromise the session. For example, it is possible to create a VR setting that remains identical in every detail for the initial IMT sessions. This is particularly crucial with patients who have symmetry–obsessive–compulsive disorder, as even small changes in the setting can be perceived as a disturbance [52], making it difficult to continue with the session. Additionally, by using VR, it is possible to achieve greater isolation compared to an in-person session from disruptive external events that could compromise the user's state of flow, interrupt potential moments of affect attunement, and therefore hinder the establishment of the interpersonal dimension between the user and the music therapist. VR also allows the selection of musical instruments that would be impractical to use in a real setting. For example, it is possible to use expensive musical instruments, like marimbas, without the costs and logistical challenges of the real musical instrument. These advantages of VR can be used to construct an ideal IMT setting.

3.2.2 | The Constructed Setting and VRMIs

For the system, we selected VRMIs aimed at neurotypical users who could interact while standing and moving around the setting. We built four sets of boomwhackers (Figure 4(a)), two tongue drums (Figure 4(b)), two glockenspiels (Figure 4(c)), two tambourines (Figure 4(d)), one marimba (Figure 4(e)), and a cymbal (Figure 4(f)). This set of VRMIs includes six percussion instruments, four of which are harmonic, meaning they can produce multiple notes simultaneously: the glockenspiel, boomwhackers, marimba, and tongue drum. In MT, clients do not need to have any level of musical knowledge. For this reason, it is necessary to facilitate the users' interaction with musical instruments to avoid the fear of producing dissonant sounds. The selected scales for the harmonic VRMIs allow minimizing the production of dissonant sounds. The glockenspiel and the boomwhackers provide the musical notes within the C major scale (C, D, E, F, G, A, B). The tongue drum features the C major pentatonic scale, consisting of five tones (C, D, E, G, A). The ability to change the sound of VRMIs by using bare hands or mallets plays a crucial role in the exploratory process of the users. They are also important in MT proposals, which are sonorous–musical activities proposed by the music therapist to facilitate engagement. These proposals leverage sound parameters such as intensity, timbre, pitch, and duration. To enable further variations in sound parameters, it is possible to change the sound intensity of instruments by applying varying force with both bare hands and mallets. Through these proposals, the music therapist can attempt to establish a relationship with the user by observing their behavior in response to the presented proposal, creating attunements. The instrument sounds were sourced from high-quality .wav sample packs, each recorded at 16-bit depth and 44.1-kHz sampling rate [53].

The setting was designed to include both an individual space (Figure 5(a)) and a shared space (Figure 5(b)) through the arrangement of the VRMIs. This layout was designed to allow users to stay in the space they deemed most comfortable for themselves at any given time. Figure 1 shows examples of how the users and the music therapist can move in the constructed spaces. We placed a table on the individual space, arranging a pair of tongue drums and glockenspiels in a mirror layout. This arrangement allowed the music therapist and the user to face each other while playing their own VRMIs (Figure 6(a)). The two sets of boomwhackers were positioned on the ground adjacent to the table. The marimba and the cymbal were located next to the table, creating the shared space where it was possible to share the same VRMI (Figure 6(b)). Providing users the option to choose whether to share a musical instrument by moving through space allows for the identification of relational indicators, such as proxemics. In IMT, voice is considered a musical instrument. Vocalizations are used as a communicative–relational tool since their musical aspects, such as inflection, intensity, and timbre, facilitate emotional exchange between two individuals [54]. Therefore, we integrated the voice chat provided by Normcore into the system.

3.2.3 | Latency of VRMIs

To build the real-time shared interaction with VRMIs, we used Normcore. When a VRMI is struck by either the user or the music therapist, an event is generated with information about the produced musical note, the originating VRMI, the sound intensity, which is mapped to the collision velocity, and whether the note was hit with mallets or hands. This allows the playback of the musical note within the user's or music therapist's application. However, event propagation is inevitably affected by network transmission delays. We assessed the latency using a ping test, measuring the round-trip time. The latency was observed to be between 40 and 80 ms, while the maximum latency for playing a musical piece together remotely while maintaining cohesion and synchrony is around 20–30 ms. Latencies above this can disrupt the timing and coherence necessary for a seamless musical performance [55]. However, as previously mentioned, the main goal of IMT is not musical performance but to leverage sonorous–musical production as a means of nonverbal communication. Therefore, latency issues, which are known to impact remote musical performances, become acceptable for IMT as long as the latency remains below 100 ms. Beyond the threshold of 100 ms, the music experience becomes noninteractive [55] and potentially disruptive to the therapeutic session.

3.3 | Dashboard

At the end of each traditional IMT session, the music therapist must fill out a detailed observational report to monitor the user's progress in the therapeutic journey. This document outlines the relational indicators observed during the session and includes observations on the session and on the therapy's progress. Filling out the observational report is time-consuming for music therapists because, currently, the only way they have to obtain relational indicators is by manually reviewing video recordings of the sessions and taking notes. Therefore, to support this process, we built a dashboard (Figure 7) to display the relational indicators collected automatically by the system, saving the time needed to review the recorded video of the entire session. The dashboard details the features of the user's sonorous–musical productions, including intensity and the beats per minute of each rhythm cell played. Rhythm cells are identified by pauses between one played note and the subsequent one that exceed a threshold of 3 s. The dashboard also shows the number of eye contacts, which are identified by the direction of the user's and music therapist's heads, the interpersonal distance over time, visualized by a line chart, and the music therapists' and the users' positions over time through a heatmap. Furthermore, the dashboard provides session duration and the history of all musical notes played with their intensity, and the originating VRMI.

(a) (b) (c)

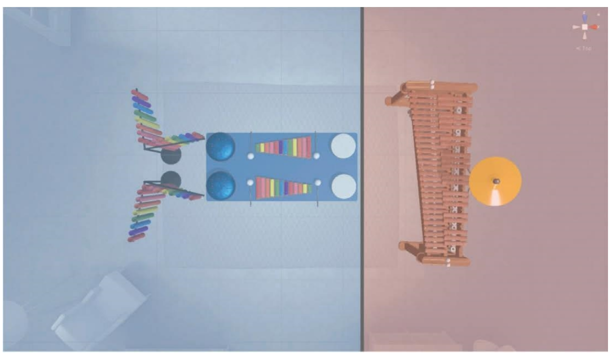
(d) (e) (f)

FIGURE 4 | VRMIs built and used in the virtual setting. (a) Boomwhackers. (b) Tongue drum. (c) Sound bars. (d) Tambourine. (e) Marimba. (f) Cymbal.

The core idea behind the dashboard is that data traditionally collected manually by music therapists through video session reviews can be collected automatically and objectively by the system. This automation could relieve therapists from the tedious task of data collection, allowing them to focus more on analyzing relational dynamics rather than spending time on data collection. It is important to note that the dashboard should be understood as a supportive tool for session analysis rather than a complete replacement for therapeutic evaluation. Many relational aspects of IMT sessions are inherently complex and difficult to codify into quantitative data, requiring the therapist's professional interpretation and analysis.

4 | Materials and Methods

The purpose of this study is to assess if the proposed immersive VE for IMT could produce the desired qualities of interpersonal relationship, which can allow the music therapist to effectively conduct IMT sessions.



(a) (b)

FIGURE 5 | Virtual music therapy setting divided into (a) individual space and (b) shared space.

(a)

(b)

FIGURE 6 | Screenshots of a session. The music therapist and the user are as follows: (a) in their individual space and (b) in the shared space.

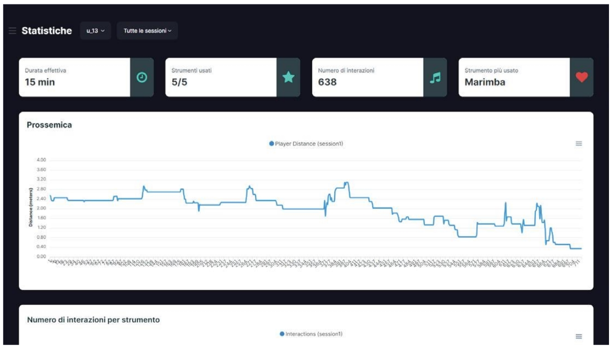


FIGURE 7 | A screenshot of part of the dashboard that shows the session duration, the number of VRMIs used, the total number of musical notes played, the most used VRMI, and a line chart that shows the distance between the user and the music therapist over time.

4.1 | Measures

4.1.1 | Demographic Questionnaire

We recorded participants' age and gender, and interviewed them about how many hours they had previously used VR and whether they played a musical instrument.

4.1.2 | Quality of IMT Questionnaire

To measure the perceived quality of IMT, we administered the 21-item questionnaire provided in Table 1 and referred to as QIMT in the following. The 10 factors assessed by the QIMT concern the perceived quality of interpersonal relationships and users' readiness and predisposition toward adherence to an IMT setting. Table 1 highlights which items correspond to each factor. Participants were asked to indicate how much they agreed with each sentence based on the IMT experience they had just tried.

Essential constructs for ensuring users' effective participation and benefit from a relational therapeutic process include motivation (an internal state that activates, directs, and sustains behavior over time) [56], self-efficacy (awareness of being able to manage and implement a task) [57], emotional intelligence (the ability to feel, recognize, and manage one's emotions) [58], perceived well-being (a state of pleasure or discomfort in relation to a lived experience) [59], flow (a state of engagement and immersion in an activity over time) [60], and satisfaction. The combination of these factors characterizes compliance with all treatments based on relationships [61], particularly nonverbal ones, such as IMT [62]. Within an IMT context, these factors and the perceived quality of the interpersonal relationship serve as indicators of the establishment of an intersubjective dimension that develops during the sessions [62]. The items for the following constructs were adapted from prior work, incorporating references to the IMT session: Motivation items were taken from [63], self-efficacy from [57], emotional intelligence from [58], perceived well-being from [59], and flow from [64]. The items that measure self, social, and spatial presence factors were taken from prior work on multimodal and nonverbal communication in collaborative VEs [64–66]. Items are rated on an 11-point scale (0 = "not at all," 10 = "very"). Since Item 7 contains a negative statement, its value was reversed. To calculate the score of each subscale, the corresponding answers were averaged. The reliability analysis conducted for each subscale indicated acceptable or higher internal consistency, with all Cronbach's alpha coefficients greater than 0.69. Specifically, the Cronbach's alpha coefficients for the subscales were as follows: Motivation ($\alpha = 0.73$), Self-efficacy ($\alpha = 0.76$), Emotional Intelligence ($\alpha = 0.69$), Perceived Well-being ($\alpha = 0.77$), Self-Presence ($\alpha = 0.75$), Presence in Relationship ($\alpha = 0.85$), Presence in Space ($\alpha = 0.70$), and Perceived Reciprocity ($\alpha = 0.81$).

4.1.3 | Customizable Interactions Questionnaire (CIQ)

The CIQ measures five factors that are important in understanding how useful the system was for completing specific tasks from multiple viewpoints: quality of interactions, subjective assessment of ability to perform the specified task using the specified mode of interaction, comfort in performing the tasks, quality of the sensory enhancements (sound, touch) provided by the mode of interaction, and the consistency of the interaction with user expectations [67]. CIQ was used to assess the users' perceived quality of interaction with VRMIs. If the interaction with VRMIs were unsatisfactory, it could negatively impact the interplay, which is crucial for enabling the intersubjective dimension between the user and the music therapist. The specified task was to play VRMIs, and the specified mode of interaction was controllers. The reliability analysis conducted for each subscale indicated acceptable or higher internal consistency, with all Cronbach's alpha coefficients greater than 0.74. Specifically, the Cronbach's alpha coefficients for the subscales were as follows: Quality of Interactions ($\alpha = 0.77$), Comfort ($\alpha = 0.74$), Assessment of Task Performance ($\alpha = 0.79$), Consistency with Expectation ($\alpha = 0.84$), and Quality of the Sensory Enhancements ($\alpha = 0.80$).

4.1.4 | Participants' Feedback

Participants were asked to give feedback on the system's strengths and weaknesses, and to provide suggestions for its improvement.

4.2 | Participants

The study was approved by the Institutional Review Board of our university and involved 23 participants: 17 males (74%) and 6 females (26%) with an age range of 20–57 years old ($M = 24$, $SD = 7.39$). The participants were recruited by reaching out directly to graduate and undergraduate students at our university, and they did not receive any compensation for their participation. Most participants (92%) had limited or no experience with VR. Specifically, 5 participants (22%) were first-time VR users and 16 (70%) had used VR for less than 10 h in their lives. Only 4 participants (17%) reported having experience in playing a musical instrument. A power analysis was conducted using G∗Power Version 3.1.9.7 [68] to determine the minimum sample size required. Results indicated the required sample size to achieve 95% power for detecting a large effect, at a significance criterion of $\alpha = 0.05$, was $N = 23$ for a two-sided one-sample t-test. Thus, the recruited sample size is adequate.

4.3 | Procedure

Participants were individually involved in a 50-min session that included a briefing, a tutorial to familiarize themselves with how to interact with VRMIs, an IMT session, and the completion of

TABLE 1 | QIMT questionnaire used in the study.

	Factor	Item
1	Motivation	I feel motivated to continue the experience
2	Self-efficacy	I felt adequate
3	Emotional intelligence	I felt engaged
4	Self-efficacy	I felt satisfied with my actions/behaviors
5	Motivation	I explored and used all the musical instruments
6	Self-efficacy	I feel I delivered a good musical performance
7	Perceived well-being	I felt discomfort
8	Emotional intelligence	I experienced and identified one or more emotions
9	Emotional intelligence	I was able to express my emotional state
10	Self-presence	I felt like my avatar's body was my own body
11	Self-presence	When something happened to my avatar, I felt like it was happening to me
12	Social presence	It felt like I was in the same room as my experience partner
13	Perceived reciprocity	I felt like my experience partner was in sync with my actions
14	Spatial presence	I felt like I was really there inside the virtual environment
15	Spatial presence	I felt as if I could reach out and touch the objects or people in the virtual environment
16	Social presence	I felt like my experience partner was aware of my presence
17	Perceived reciprocity	I felt like my experience partner understood my intentions
18	Perceived well-being	I enjoyed interacting in the virtual environment
19	Perceived well-being	I found the experience fun
20	Flow	I felt that time passed without noticing it
21	Satisfaction	I am satisfied with the overall experience

questionnaires. Participants were welcomed and received an initial briefing designed with music therapists to mirror briefings that are typically conducted in an IMT session. The briefing clarified that the VR experience involved interacting with VRMIs, emphasizing that no prior musical skills were needed, and musical performance was not going to be evaluated in any way. Following the typical communication provided during the briefing with neurotypical users in traditional IMT sessions, participants were informed that there was no specific goal to achieve. Instead, they could freely explore the environment and engage with VRMIs. As in a traditional IMT session, participants were asked to refrain from speaking during the session but were allowed to produce other sounds, such as vocalizations. Finally, they were clearly informed that the system automatically collected anonymous data that were going to be analyzed for research purposes. All participants consented to participate in the study. They were informed that the IMT session was going to last up to 15 min, but they could end it anytime. All participants completed the 15-min session without ending it prematurely. After wearing the VR headset and over-ear headphones, participants could choose the gender of their avatar and begin the tutorial. During the tutorial, participants were instructed to use the controllers to grab objects. Specifically, users were required to pick up cubes of different colors and place them on a table. Furthermore, the tutorial detailed how participants could move within the VE by walking or using the controller's joystick. Once the tutorial was completed, participants automatically entered a virtual room where the remotely connected music therapist began the IMT session. The same certified music therapist conducted all sessions. The IMT session was structured into four phases: introduction, initial exploration, sonorous–musical improvisation, and conclusion. In the introduction, the music therapist welcomed the participant and informed him/her that the session would last up to 15 min and that they were free to use the VRMIs as they wished. Then, the initial exploration phase lasted 1 minute, during which participants freely explored the setting and the VRMIs. The music therapist observed how participants engaged with VRMIs, their confidence level, and propension to explore. In this phase, the music therapist waited for the participant to take the initiative to play a VRMI. If participants did not take the initiative, the music therapist introduced simple rhythmic or melodic cells for easy imitation through a sonorous–musical proposal. All participants spontaneously interacted with at least one VRMIs within the first minute. After the initial exploration phase, the sonorous–musical improvisation phase started: The music therapist sought to establish interplay by responding to the participants' sonorous–musical production, either by mimicking it or by introducing variations. If no recognizable sonorous–musical cells were played by participants, the music therapist introduced simple rhythmic and melodic cells to foster and enhance interplay. Halfway through the session, the music therapist increased the level of interaction with the participant by introducing vocalizations and new sonorous–musical cells. The conclusion phase involved the final minute of the session, during which the music therapist guided the improvisation to an end, reducing the pulse of rhythmic and melodic cells and decreasing the sound intensity. Finally, the music therapist announced that the session had ended and thanked the participant for his/her involvement. Participants were then helped to remove the headset and invited to fill out the demographic, QIMT, and CIQs in order. A unique code was provided to each participant so that (s)he could write it in each questionnaire to ensure anonymity. After completing the questionnaires, participants were asked to provide comments on the strengths, weaknesses, and possible improvements to the system. Participants were then thanked for their participation.

5 | Results

All statistical analyses were conducted using R Version 4.3.1. For each subscale of the QIMT questionnaire, we verified that the scores obtained were statistically different from the hypothesized population mean of 6. This threshold was selected because items of the questionnaire were rated on an 11-point scale, where a score of 6 denotes a level of construct perception greater than the neutral midpoint. Data from the questionnaire were tested for normality using the Shapiro–Wilk test. The resulting p values for each subscale were less than 0.05, indicating that the data did not follow a normal distribution. As recommended in [69], we utilized the one-sample Wilcoxon signed-rank test as a nonparametric alternative to the one-sample t-test, suitable when the data do not follow a normal distribution. Following [70], the effect size was calculated using the r statistic, which is a standardized measure of effect size for the one-sample Wilcoxon signed-rank test. According to Cohen's guidelines, an r effect size of 0.1 is considered small, 0.3 is medium, and 0.5 is large. The results of the test are reported in Table 2 and visualized in Figure 8.

Similarly, we checked for each subscale that the scores obtained from the CIQ were different from the hypothesized population mean of 4, which represents the neutral midpoint of the scale. Data did not follow a normal distribution since the p values of the Shapiro–Wilk test were less than 0.05. Therefore, we used the one-sample Wilcoxon signed-rank test as a nonparametric alternative to the one-sample t-test. The results of the one-sample Wilcoxon signed-rank test are reported in Table 3 and visualized in Figure 9.

5.1 | Participants' Feedback

Participants' feedback focused on interactions with VRMIs, the perception of the setting, and the interpersonal relationship established with the music therapist, frequently eliciting positive emotional responses ($n = 20$) but, in a few cases, negative ones ($n = 3$). Most participants appreciated the VRMIs, finding them easy to use and realistic ($n = 12$). For example, "*One of the strengths of the experience was the musical instruments I interacted with. They were varied, the interaction was satisfying, and the produced sounds were realistic.*" [P1], "*The musical instruments, with their colors and sizes, ensured simple and precise use*." [P10], "*Interacting with the musical instruments was very pleasant. The sound seemed like that of a real musical instrument.*" [P6]. Participants' feedback also revealed that the setting was well constructed ($n = 6$). Some participants shared their individual VRMIs with the music therapist ($n = 12$). For example, "*The surrounding space completely enveloped me, the sounds, the spaces immersed me in that dimension. […] at the beginning, we were separated, and each played their own musical instrument, but when we went together on the marimba, it was nice to play together*." [P3]. Most participants found the navigation within the setting easy and explored the setting and the VRMIs autonomously ($n = 17$). For example, "*Moving around the room was very easy*." [P8], and "*The tutorial and the simplicity of the controls allowed for quick learning. […] the mallets were used exactly as in the real world, so exploring the musical instruments was simple and fun*." [P23]. None of the participants reported motion sickness. Regarding the relational aspect, participants reported positive feedback ($n = 6$), such as "*Even though I couldn't speak, the experience was positive. Even without speaking, we established a friendly relationship. […] The other person understood what I wanted to do*." [P17] and "*It was possible to interact directly with the tutor as if we were simultaneously in the same room. We could play together, and this made me feel involved.*" [P20]. Negative feedback related to interpersonal relationships with the music therapist also emerged. The presence of a human partner with whom one could not verbally interact caused discomfort in some users ($n = 3$): "*I felt a strong sense of embarrassment throughout the experience, not knowing the person in the room and feeling embarrassed to play together*." [P4], "*Spending a lot of time with someone without talking made me upset.*" [P11], and "*I felt increasing deep embarrassment. I didn't know the other person.*" [P5]. These comments highlight how the virtual setting can elicit real emotional responses in users, both positive and negative, just as it happens during in-person IMT sessions. The negative responses that emerged related to the sense of inadequacy about playing a musical instrument or the shyness of having to interact with a stranger also occur during in-person IMT sessions. The music therapist, noticing these problems, implements strategies to mitigate the negative emotional response in subsequent sessions. Some users provided feedback on potential system improvements ($n = 4$). These include offering more precise gestures, which were not achievable using the controllers ($n = 2$). For example, "*I would have preferred to use the musical instruments without the controllers, allowing for greater freedom to use my hands,* for example*, to attempt to communicate through gestures*." [P16]. Some users suggested that another potential improvement could include enhancing the capability to generate sounds using VRMIs and through environmental interactions ($n = 2$). For example, "*The sounds produced by the musical instruments were very realistic. However,* for example*, dragging the mallet across the table did not produce any sound.*" [P2].

5.2 | Music Therapist's Feedback

The music therapist involved in the evaluation of the system provided key insights into its potential and limitations. Despite no prior VR experience, the music therapist adapted quickly to the VE, describing the controller-based interactions for instrument manipulation and environment navigation as intuitive and reliable. From a logistical standpoint, the therapist highlighted how VEs give IMT sessions significant flexibility. She highlighted how in-person IMT sessions often face logistical difficulties, including space availability, the transportation and cost of large or delicate instruments, and unexpected acoustic issues. In some physical settings, for example, reverberation can make high-intensity instruments unusable, as they may produce unpleasant auditory effects for certain users. In VR, these concerns diminish significantly. Large or expensive instruments (e.g., marimbas) can be used without logistical complications. The locomotion system lets users participate in the session from real-world small spaces in ways that approximate the freedom of

TABLE 2 | Mean, standard deviation, p value, and effect size of the IMT subscales, one-sample Wilcoxon signed-rank test with H_0: $\mu = 6$.

Scale	M	SD	p value	r
Motivation	8.57	1.39	< 0.001	0.76
Self-efficacy	6.71	1.73	< 0.05	0.08
Emotional intelligence	7.29	1.61	< 0.001	0.35
Perceived well-being	7.23	1.01	< 0.001	0.20
Self-presence	8.20	0.62	< 0.001	0.37
Presence in relationship	9.17	0.43	< 0.001	0.79
Spatial presence	9.11	0.58	< 0.001	0.79
Perceived reciprocity	8.70	0.37	< 0.001	0.79
Flow	7.78	1.67	< 0.001	0.32
Satisfaction	8.83	0.98	< 0.001	0.72

a much larger room, which is often unavailable for in-person IMT sessions. Furthermore, setting preparation is drastically simplified in VR because its original layout remains instantly available between sessions, ensuring that instruments retain the same original position. This feature is particularly beneficial, for example, for users with obsessive–compulsive disorders, who often rely on stable, consistent spatial configurations to reduce anxiety and facilitate engagement.

Beyond logistical convenience, the therapist noted that VR also helps reduce disruptive stimuli, such as ambient noise or people talking nearby, that often negatively affect in-person sessions. Furthermore, the therapist noted that the system could be particularly beneficial for individuals with limited fine motor skills or reduced hand–eye coordination, such as older adults with dementia. In in-person sessions, these users might unintentionally produce discordant sounds or become frustrated when playing physical instruments. By contrast, the VR environment ensures that even imprecise gestures yield pleasant, consistent sounds, potentially reducing performance anxiety and encouraging exploration. Additionally, the avatar-based interaction seemed to lower inhibitions, allowing users to vocalize freely and more often than they might in an in-person session. Moreover, she reported that many essential traditional IMT techniques remained feasible in VR, including mirroring (i.e., replicating a user's melodic or rhythmic lines), asynchronous imitation (i.e., following the user's line with timing shifts), variation-based imitation (i.e., changing tempo or intensity of the user's melodic line), antiphonal exchange (i.e., responding to or concluding a user's musical phrase), dialog (i.e., introducing new melodic lines during a pause or continuing the user's line if the pause is short), and rhythmic or melodic anchoring (i.e., using the user's rhythmic pulse as a backbone or shifting the tonal center of the user's melody). She regarded this as a strong indication of the system's ability to replicate core aspects of in-person sessions. These observations are consistent with other findings on VR-based psychological interventions, which highlight greater flexibility in intervention timing, improved cost-effectiveness, and the capacity to tailor sessions to individual needs [29].

Despite these advantages, however, the therapist stressed that improvements in nonverbal communication are key to achieving the full benefit of IMT. Accurate gaze tracking, face tracking, and higher fidelity body-motion capture would help her interpret subtle emotional cues that are crucial for in-person IMT. She could infer certain emotions only through vocal intonation or avatar movement. For instance, although the therapist occasionally perceived smiling through subtle vocal shifts, the actual facial expression was missing. This absence of nuanced visual feedback (e.g., silent smiles, eye contact) limits the therapist's ability to respond in real-time to nonverbal indicators of a user's emotional state. Another limitation she highlighted was the limited range of sounds produced by the VRMIs. In a real setting, a single instrument does not produce one fixed sound but can be modulated based on numerous factors, such as the angle of the mallet strike. Without these subtle variations, the virtual instruments offer fewer expressive possibilities than their physical

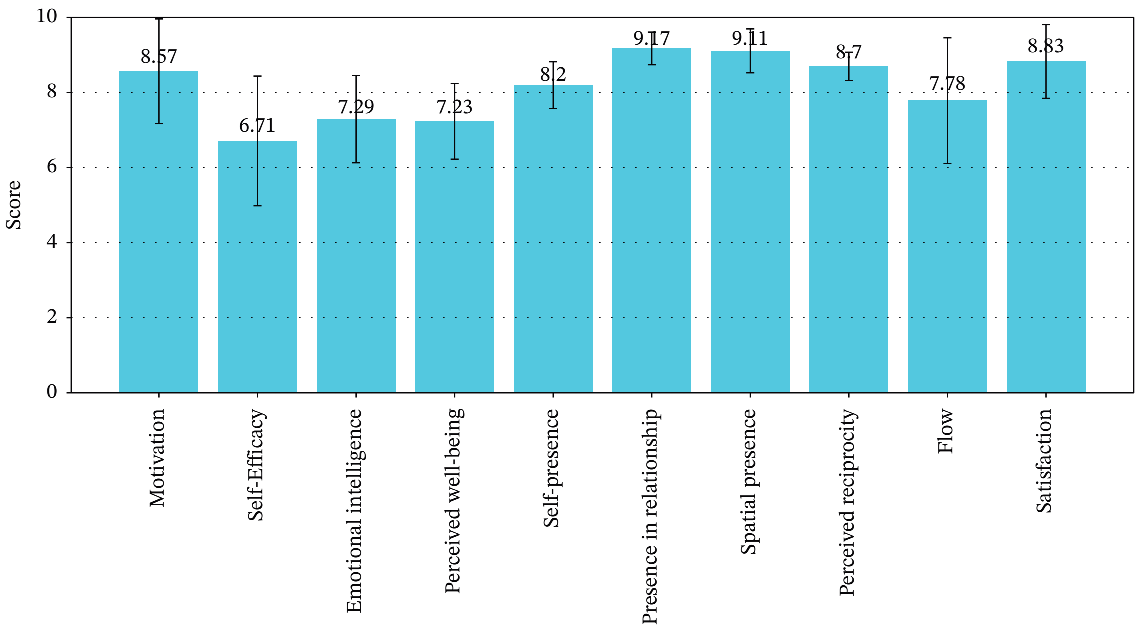


FIGURE 8 | Mean scores of each subscale of the QIMTR questionnaire.

TABLE 3 | Mean, standard deviation, p value, and effect size for the CIQ subscales, one-sample Wilcoxon signed-rank test with H_0: $\mu = 4$.

Scale	M	SD	p value	r
Quality of interactions	6.30	0.47	< 0.001	0.87
Comfort	6.43	0.43	< 0.001	0.88
Assessment of task performance	6.38	0.75	< 0.001	0.81
Consistency with expectation	5.99	0.78	< 0.001	0.63
Quality of the sensory enhancements	6.43	0.66	< 0.001	0.87

counterparts, limiting the expressive richness compared to in-person practice. Finally, she found the visualization of the relational indicators through the dashboard very useful. She was not accustomed to having immediate, data-driven insights, such as the recorded evolution of interpersonal distance, and appreciated how the graphical display provided objective confirmation of her observations. For example, when she noticed the user's decreasing interpersonal distance during a session, she found the same shift reflected in the dashboard's metrics, making it easier to track relational indicators like proxemics compared to video reviews. In addition, she suggested that a transcript of the notes played by users could help music therapists assess melodic and rhythmic variations at a glance, giving further insights on the evolution of relational dynamics through sonorous–musical productions.

6 | Discussion

The results indicate that the proposed system can support key requirements for remote IMT, including co-presence with a music therapist, shared interaction with musical instruments, and nonverbal musical interplay within an immersive VE. High scores for motivation, self-efficacy, emotional intelligence, and perceived well-being are relevant because these constructs are foundational to effective engagement in IMT. Motivation facilitates sustained involvement in activities, self-efficacy reflects confidence in one's ability to participate, emotional intelligence enables recognition and expression of emotional states, and perceived well-being indicates positive affective responses to the experience, all of which are considered important prerequisites for IMT practice [61, 62]. The high scores reported for self-presence, social presence, and spatial presence further underscore the system ability in creating a convincing and immersive setting, facilitating a sense of being together in a shared space, which is vital for the IMT process [62]. Overall, these results suggest that participants perceived high quality in interpersonal relationships through nonverbal communication and musical interplay between the music therapist and the user in a shared space, indicating that the system successfully provided the essential interactional elements for IMT. These elements may support relational processes associated with the emergence of an intersubjective dimension, which is a central goal of IMT [62]. The CIQ results and the feedback provide an understanding of the participants' experiences with the system. The quantitative results from the CIQ indicate a high level of user satisfaction across various dimensions of the experience. Notably, the factor of quality of the sensory enhancements received high scores, reflecting users' appreciation for the realistic representation of the VRMIs. This aspect was also highlighted in the positive feedback from participants, who praised the VRMIs for their realism. The quality of interactions and the assessment of task performance also received high scores, suggesting that the system facilitated interactions with VRMIs. This aligns with user feedback that emphasized the ease of use of VRMIs. Overall, the results of the QIMT and CIQ show the proposed system's effectiveness in meeting the essential requirements to conduct an IMT session. The system indeed effectively facilitated the co-presence between the user and the music therapist within a shared virtual space, where musical instruments could be manipulated and shared. The positive relational experiences reported by participants, including the spontaneous choice to share individual instruments with the therapist, suggest that the system might support the spatial and interactive dynamics that characterize in-person IMT. However, the discomfort expressed by some participants when interacting nonverbally with an unfamiliar person reflects a realistic aspect of IMT practice: initial sessions often involve self-consciousness, which therapists address through gradual rapport-building. The fact that the virtual setting elicited these authentic emotional responses, both positive and negative, indicates that the system could be used to support meaningful interpersonal dynamics. This is consistent with recent reviews of VR in mental health interventions, which show that VEs can elicit authentic emotional and behavioral responses comparable to real-world settings [37]. The music therapist's feedback provides critical insight into the system's viability for IMT practice. The ability to adopt traditional IMT techniques within the VE indicates that the system supports core

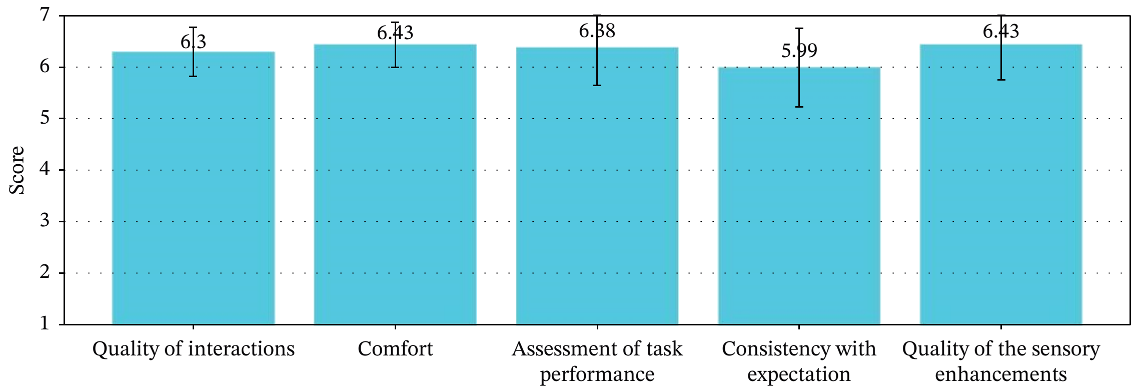


FIGURE 9 | Mean scores of each subscale of the CIQ questionnaire.

elements of the IMT practice. The logistical advantages the music therapist identified, that is, instrument availability without physical constraints and consistent setting configuration, address practical barriers that often limit in-person IMT. The VR system offers potential applications beyond addressing logistical or geographical barriers. In particular, it may make IMT accessible to people who would face substantial challenges participating in traditional in-person IMT regardless of logistics. For example, even for people who have no logistical and geographical barriers that make it difficult to meet a music therapist, conditions such as severe social anxiety and/or prolonged social withdrawal (e.g., hikikomori) may be another barrier. In these cases, users may find the avatar-mediated interaction less intimidating than face-to-face sessions, making the meeting possible in VR. Moreover, people with motor impairments could benefit from the possibilities offered by VR that are not available in physical settings. For example, the system could be configured to amplify the force applied to virtual instruments, allowing users with reduced motor strength to perceive themselves as striking instruments with greater intensity than they could physically achieve in in-person IMT. These observations suggest that VR-based IMT may not only replicate in-person sessions when they are logistically unfeasible, but also expand access to populations for whom traditional IMT presents inherent challenges.

The proposed system has significant potential for improvement in two primary aspects: improving the precision of nonverbal cues through better tracking and enhancing the sound modulation capabilities of VRMIs. Integrating face and eye tracking to accurately reproduce facial expressions and gaze directions would enable the acquisition of key elements of interpersonal communication. This would allow the music therapist to decode more easily the user's emotional state. With more precise information on the user's emotional state, the music therapist could employ various strategies to adapt the therapeutic approach accordingly, enhancing the therapeutic process. Moreover, using more robust hand tracking algorithms that minimize tracking issues to allow for more natural gesturing, as opposed to using controllers or tactile gloves that also provide haptic feedback during interaction with VRMIs, would increase the flow of nonverbal information that can be exchanged between the user and the music therapist. Further development could also involve the implementation of accurate posture reconstruction through additional trackers or robust pose estimation algorithms to avoid relying solely on inverse kinematics, ensuring the most faithful transmission of body language information possible. Additionally, expanding the range of possible sound modifications with VRMIs, not limited to collision speed but including other parameters, like material composition and impact angle, is essential to facilitate user expression and improve users' consistency with expectations. Finally, developing and applying algorithms for detecting moments of affect attunement through the analysis of sonorous–musical interactions and body language might help to automatically create the observational report. This could be beneficial because it would avoid the manual investigation by the music therapist of the collected relational indicators, which can be a time-consuming and error-prone process. This would offer music therapists a tool capable of gathering valuable session data to quantify the progress of IMT sessions, addressing the current lack of quantifiable data in the evaluation of therapeutic progress. Taken together, the results suggest that the metaverse, experienced through immersive VR, could be considered as a promising solution for building digital tools useful for music therapists in situations where in-person sessions are not feasible. This aligns with the literature that portrays the metaverse as a tool with the potential to facilitate mental health care and treatment [71–75].

7 | Limitations and Future Work

To the best of our knowledge, our paper presents the first proposal and feasibility study of a system for conducting remote IMT sessions through immersive VR in a metaverse context. Our study opens a new line of research and paves the way for extensive future work. This paper should be interpreted as a technology demonstration with preliminary user experience evaluation rather than a clinical validation study. While the study provides valuable insights into system feasibility and user experience, several limitations should be acknowledged.

First, participants were university students without identified therapeutic needs who engaged in single 15-min sessions designed to include IMT elements but not structured toward specific therapeutic goals. This design allowed us to assess whether the system provides the fundamental interactional elements required to conduct an IMT session, including copresence with a music therapist, manipulation of shared virtual instruments, and nonverbal communication through musical interplay. Although participants reported high levels of motivation, presence, and perceived reciprocity, these subjective responses collected after a single short session are insufficient to evaluate therapeutic outcomes. Establishing clinical validity requires studies with patient populations across multiple sessions and measurable therapeutic objectives. The distinction between system feasibility and therapeutic efficacy is critical, as positive user experiences in controlled laboratory settings do not necessarily translate to benefits in clinical contexts. At this stage, it is not possible to determine whether the proposed system supports the establishment of an intersubjective dimension between the music therapist and users through nonverbal communication and musical interplay, which is a prerequisite for using the system in clinical practice. Future work should therefore evaluate the system with actual patients, track outcomes across treatment programs, and compare virtual IMT to traditional in-person sessions.

Second, the sample presents a marked gender imbalance, with 74% male and 26% female participants. This limitation is particularly significant given documented gender differences in emotional processing [76, 77]. The male predominance may have influenced the magnitude of reported emotional constructs, potentially leading to lower scores for emotional intelligence and related factors compared to what might be observed in a gender-balanced or a female-majority sample. This imbalance limits the generalizability of the quantitative findings across genders. Future studies should employ gender-balanced sampling and investigate whether gender moderates responses to virtual IMT settings.

Third, all sessions were conducted by a single music therapist. While this ensured consistency in therapeutic approach, it limits generalizability regarding how different therapists might use the system and whether therapeutic style interacts with the virtual medium. Future studies should involve multiple therapists with varied backgrounds and approaches to IMT.

Finally, the study did not include a direct comparison with in-person IMT sessions. Such comparisons would be valuable for identifying how the virtual experience differs from traditional practice and for understanding which aspects of in-person IMT are successfully replicated versus which require further technological development. The mediated nature of the metaverse may also influence initial experiences of shyness or embarrassment, an aspect that warrants systematic investigation.

Despite these limitations, this work contributes to understanding how immersive virtual reality and the metaverse can be designed to support remote IMT. The system architecture, design decisions, and preliminary findings provide a foundation for future research that should address these limitations.

8 | Conclusions

MT requires innovative solutions to develop effective tools for addressing barriers that hinder the execution of in-person sessions. In this paper, we proposed and evaluated the first system that aims to support remote IMT sessions by leveraging VR and a metaverse. The proposed system was designed and created by integrating the expertise of computer scientists, music therapists, and clinical psychologists to address the specific requirements of IMT, namely, co-presence with a music therapist in a shared space, manipulation of shared VRMIs, and support for nonverbal communication through musical interplay. We showed that our VR metaverse experience supports essential requirements for an IMT session, as reflected by the high evaluations on constructs related to motivation, self-efficacy, presence, perceived reciprocity, and the perceived quality of the interpersonal relationship between the user and the music therapist. This paper opens a new line of research and paves the way for extensive future work. Future studies should focus on expanding system capabilities to convey more precise nonverbal information for facial expressions, gestures, and body movement, and to generate a more diverse variety of sounds from VRMIs and from environmental interactions, to enhance the expressive possibilities of users. Additionally, future studies should evaluate the system across multiple sessions with clinical populations and compare outcomes to traditional in-person IMT.

Funding

No funding was received for this manuscript. Open access publishing facilitated by Universita degli Studi di Udine, as part of the Wiley - CRUI-CARE agreement.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. World Federation of Music Therapy, *What is Music Therapy?* (2011), https://www.wfmt.info/about.

2. S. Aalbers, L. Fusar-Poli, R. E. Freeman, et al., "Music Therapy for Depression," *Cochrane Database of Systematic Reviews* 2017, no. 11 (2017): https://doi.org/10.1002/14651858.CD004517.pub3.

3. J. Edwards, ed., *The Oxford Handbook of Music Therapy (Vol. 1)* (Oxford University Press, 2015), https://doi.org/10.1093/oxfordhb/9780199639755.001.0001.

4. K. E. Bruscia, *Defining Music Therapy*, 3rd ed. (Barcelona Publishers, 2014).

5. J. Bradt, C. Dileo, and M. Shim, "Music Interventions for Preoperative Anxiety," *Cochrane Database of Systematic Reviews* 2013, no. 6 (2013): CD006908, https://doi.org/10.1002/14651858.CD006908.pub2.

6. C. Gold, H. P. Solli, V. Krüger, and S. A. Lie, "Dose–Response Relationship in Music Therapy for People With Serious Mental Disorders: Systematic Review and Meta-Analysis," *Clinical Psychology Review* 29, no. 3 (2009): 193–207, https://doi.org/10.1016/j.cpr.2009.01.001.

7. K. E. Bruscia, *Improvisational Models of Music Therapy (pp. 105, 590)* (Charles C Thomas, Publisher, 1987).

8. P. L. Postacchini, A. Ricciotti, and M. Borghesi, *Musicoterapia (2nd)* (Carocci, 2014).

9. J. B. Nielsen and U. Holck, "Synchronicity in Improvisational Music Therapy–Developing an Intersubjective Field With a Child With Autism Spectrum Disorder," *Nordic Journal of Music Therapy* 29, no. 2 (2020): 112–131, https://doi.org/10.1080/08098131.2019.1680571.

10. S. Malloch and C. Trevarthen, *Communicative Musicality: Exploring the Basis of Human Companionship* (Oxford University Press, 2009).

11. M. Pavlicevic, *Music Therapy in Context: Music, Meaning and Relationship* (Jessica Kingsley Publishers, 1997).

12. A. Raglio, M. Gnesi, M. Monti, et al., "The Music Therapy Session Assessment Scale (MT-SAS): Validation of a New Tool for Music Therapy Process Evaluation," *Clinical Psychology & Psychotherapy* 24, no. 6 (2017): https://doi.org/10.1002/cpp.2115.

13. D. N. Stern, *The Interpersonal World of the Infant: A View From Psychoanalysis and Developmental Psychology* (Basic Books, 1985).

14. D. N. Stern, *Forms of Vitality: Exploring Dynamic Experience in Psychology, the Arts, Psychotherapy, and Development (First Published)* (Oxford University Press, 2010).

15. C. Trevarthen, "The Concept and Foundations of Infant Intersubjectivity," in *Intersubjective Communication and Emotion in Early Ontogeny* (Cambridge University Press, 1998), 15–46.

16. T. Wigram and J. De Backer, eds., *Clinical Applications of Music Therapy in Psychiatry* (Kingsley, 1999).

17. G. G. Barker and E. E. Barker, "Online Therapy: Lessons Learned From the COVID-19 Health Crisis," *British Journal of Guidance and Counselling* 50, no. 1 (2022): 66–81, https://doi.org/10.1080/03069885.2021.1889462.

18. V. Békés and K. Aafjes-van Doorn, "Psychotherapists' Attitudes Toward Online Therapy During the COVID-19 Pandemic," *Journal of Psychotherapy Integration* 30, no. 2 (2020): 238–247, https://doi.org/10.1037/int0000214.

19. M. Nazemi, S. Kiani, and S. Zakerabasali, "Tele-Mental Health During the COVID-19 Pandemic: A Systematic Review of the Literature Focused on Technical Aspects and Challenges," *Health Science Reports* 6, no. 10 (2023): e1637, https://doi.org/10.1002/hsr2.1637.

20. L. S. Rains, C. Dalton-Locke, S. Landau, J. J. Needle, and S. Johnson, "Variations in the Uptake of Telemental Health Technologies in Community and Crisis Mental Health Services During the Early Pandemic: A Survey of Mental Health Professionals in the UK," *BMC Psychiatry* 22, no. 1 (2022): 776, https://doi.org/10.1186/s12888-022-04385-1.

21. A. Zubala, N. Kennell, and S. Hackett, "Art Therapy in the Digital World: An Integrative Review of Current Practice and Future

Directions," *Frontiers in Psychology* 12 (2021): https://www.frontiersin.org/articles/10.3389/fpsyg.2021.600070.

22. M. Slater and S. Wilbur, "A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments," *Presence: Teleoperators and Virtual Environments* 6, no. 6 (1997): 603–616, https://doi.org/10.1162/pres.1997.6.6.603.

23. M. V. Sanchez-Vives and M. Slater, "From Presence to Consciousness Through Virtual Reality," *Nature Reviews Neuroscience* 6, no. 4 (2005): 332–339, https://doi.org/10.1038/nrn1651.

24. A. Brungardt, A. Wibben, A. F. Tompkins, et al., "Virtual Reality-Based Music Therapy in Palliative Care: A Pilot Implementation Trial," *Journal of Palliative Medicine* 24, no. 5 (2021): 736–742, https://doi.org/10.1089/jpm.2020.0403.

25. R. Kanehira, Y. Ito, M. Suzuki, and F. Hideo, "Enhanced Relaxation Effect of Music Therapy With VR," in *2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (2018), 1374–1378, https://doi.org/10.1109/FSKD.2018.8686951.

26. P. Perez, E. Vallejo, M. Revuelta, M. V. Redondo Vega, E. Guervós Sánchez, and J. Ruiz, "Immersive Music Therapy for Elderly Patients," in *Proceedings of the 2022 ACM International Conference on Interactive Media Experiences, IMX'22* (2022), 47–52, https://doi.org/10.1145/3505284.3529961.

27. E. Baka, M. Kentros, G. Papagiannakis, and N. Magnenat-Thalmann, "Virtual Reality Rehabilitation Based on Neurologic Music Therapy: A Qualitative Preliminary Clinical Study," *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10925 LNCS* (2018): 113–127, https://doi.org/10.1007/978-3-319-91152-6_9.

28. X. Lin, S. Mahmud, E. Jones, et al., *Virtual Reality-Based Musical Therapy for Mental Health Management* (2020).

29. W. A. Turner and L. M. Casey, "Outcomes Associated With Virtual Reality in Psychological Interventions: Where are We Now?" *Clinical Psychology Review* 34, no. 8 (2014): 634–644, https://doi.org/10.1016/j.cpr.2014.10.003.

30. B. A. Clough and L. M. Casey, "Technological Adjuncts to Increase Adherence to Therapy: A Review," *Clinical Psychology Review* 31, no. 5 (2011): 697–710, https://doi.org/10.1016/j.cpr.2011.03.006.

31. C. Botella, J. Fernández-Álvarez, V. Guillén, A. García-Palacios, and R. Baños, "Recent Progress in Virtual Reality Exposure Therapy for Phobias: A Systematic Review," *Current Psychiatry Reports* 19, no. 7 (2017): 42, https://doi.org/10.1007/s11920-017-0788-4.

32. D. Opriş, S. Pintea, A. García-Palacios, C. Botella, Ş. Szamosközi, and D. David, "Virtual Reality Exposure Therapy in Anxiety Disorders: A Quantitative Meta-Analysis," *Depression and Anxiety* 29, no. 2 (2012): 85–93, https://doi.org/10.1002/da.20910.

33. S. Riches, P. Jeyarajaguru, L. Taylor, et al., "Virtual Reality Relaxation for People With Mental Health Conditions: A Systematic Review," *Social Psychiatry and Psychiatric Epidemiology* 58, no. 7 (2023): 989–1007, https://doi.org/10.1007/s00127-022-02417-5.

34. D. P. Rowland, L. M. Casey, A. Ganapathy, M. Cassimatis, and B. A. Clough, "A Decade in Review: A Systematic Review of Virtual Reality Interventions for Emotional Disorders," *Psychosocial Intervention* 31, no. 1 (2022): 1–20, https://doi.org/10.5093/pi2021a8.

35. G. Riva, D. Di Lernia, E. Sajno, et al., "Virtual Reality Therapy in the Metaverse: Merging VR for the Outside With VR for the Inside," *Annual Review of Cybertherapy & Telemedicine* 19 (2021): 3–8.

36. A. 'S. Rizzo and R. Shilling, "Clinical Virtual Reality Tools to Advance the Prevention, Assessment, and Treatment of PTSD," *European Journal of Psychotraumatology* 8, no. 5 (2017): 1414560, https://doi.org/10.1080/20008198.2017.1414560.

37. A. Wiebe, K. Kannen, B. Selaskowski, et al., "Virtual Reality in the Diagnostic and Therapy for Mental Disorders: A Systematic Review," *Clinical Psychology Review* 98 (2022): 102213, https://doi.org/10.1016/j.cpr.2022.102213.

38. Ready Player Me, (2024), https://docs.readyplayer.me/ready-player-me.

39. VRIK, (2024), http://www.root-motion.com/finalikdox/html/page16.html.

40. Normcore, (2023), https://normcore.io/documentation/.

41. B. E. Mennecke, J. L. Triplett, L. M. Hassall, Z. J. Conde, and R. Heer, "An Examination of a Theory of Embodied Social Presence in Virtual Worlds," *Decision Sciences* 42, no. 2 (2011): 413–450, https://doi.org/10.1111/j.1540-5915.2011.00317.x.

42. F. Biocca, C. Harms, and J. K. Burgoon, "Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria," *Presence: Teleoperators and Virtual Environments* 12, no. 5 (2003): 456–480, https://doi.org/10.1162/105474603322761270.

43. C. S. Oh, J. N. Bailenson, and G. F. Welch, "A Systematic Review of Social Presence: Definition, Antecedents, and Implications," *Frontiers in Robotics and AI* 5 (2018): 114, https://doi.org/10.3389/frobt.2018.00114.

44. R. J. R. Blair, "Facial Expressions, Their Communicatory Functions and Neuro-Cognitive Substrates," *Philosophical Transactions of the Royal Society B: Biological Sciences* 358, no. 1431 (2003): 561–572, https://doi.org/10.1098/rstb.2002.1220.

45. S. Kimmel, F. Jung, A. Matviienko, W. Heuten, and S. Boll, "Let's Face It: Influence of Facial Expressions on Social Presence in Collaborative Virtual Reality," *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023): 1–16, https://doi.org/10.1145/3544548.3580707.

46. Meta, "Oculus Lipsync for Unity Development," (2021), https://developer.oculus.com/documentation/unity/audio-ovrlipsync-unity/.

47. N. J. Emery, "The Eyes Have It: The Neuroethology, Function and Evolution of Social Gaze," *Neuroscience & Biobehavioral Reviews* 24, no. 6 (2000): 581–604, https://doi.org/10.1016/S0149-7634(00)00025-7.

48. M. Argyle, M. Cook, and D. Cramer, "Gaze and Mutual Gaze," *The British Journal of Psychiatry* 165, no. 6 (1994): 848–850, https://doi.org/10.1017/S0007125000073980.

49. A. R. Bentivoglio, S. B. Bressman, E. Cassetta, D. Carretta, P. Tonali, and A. Albanese, "Analysis of Blink Rate Patterns in Normal Subjects," *Movement Disorders* 12, no. 6 (1997): 1028–1034, https://doi.org/10.1002/mds.870120629.

50. P. Caserman, P. Achenbach, and S. Göbel, "Analysis of Inverse Kinematics Solutions for Full-Body Reconstruction in Virtual Reality," in *2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH)* (2019), 1–8, https://doi.org/10.1109/SeGAH.2019.8882429.

51. F. Buttussi and L. Chittaro, "Locomotion in Place in Virtual Reality: A Comparative Evaluation of Joystick, Teleport, and Leaning," *IEEE Transactions on Visualization and Computer Graphics* 27, no. 1 (2021): 125–136, https://doi.org/10.1109/TVCG.2019.2928304.

52. A. S. Radomsky and S. Rachman, "Symmetry, Ordering and Arranging Compulsive Behaviour," *Behaviour Research and Therapy* 42, no. 8 (2004): 893–913, https://doi.org/10.1016/j.brat.2003.07.001.

53. C. Henson, *Pianobook* (2022), https://www.pianobook.co.uk/packs/.

54. E. Rayner, "Matching, Attunement and the Psychoanalytic Dialogue," *International Journal of Psycho-Analysis* 73, no. 1 (1992): 39–54.

55. M. Rofe and F. Reuben, "Telematic Performance and the Challenge of Latency," *Journal of Music, Technology & Education* 10, no. 2 (2017): 167–183, https://doi.org/10.1386/jmte.10.2-3.167_1.

56. R. M. Ryan and E. L. Deci, "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions," *Contemporary Educational*

Psychology 25, no. 1 (2000): 54–67, https://doi.org/10.1006/ceps.1999.1020.

57. G. Chen, S. M. Gully, and D. Eden, "Validation of a New General Self-Efficacy Scale," *Organizational Research Methods* 4, no. 1 (2001): 62–83, https://doi.org/10.1177/109442810141004.

58. N. S. Schutte, J. M. Malouff, L. E. Hall, et al., "Development and Validation of a Measure of Emotional Intelligence," *Personality and Individual Differences* 25, no. 2 (1998): 167–177, https://doi.org/10.1016/S0191-8869(98)00001-4.

59. R. Tennant, L. Hiller, R. Fishwick, et al., "The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): Development and UK Validation," *Health and Quality of Life Outcomes* 5, no. 1 (2007): 63, https://doi.org/10.1186/1477-7525-5-63.

60. J. Nakamura and M. Csikszentmihalyi, "The Concept of Flow," in *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*, ed. M. Csikszentmihalyi (Springer, 2014), 239–263, https://doi.org/10.1007/978-94-017-9088-8_16.

61. J. K. Aronson, "Compliance, Concordance, Adherence," *British Journal of Clinical Pharmacology* 63, no. 4 (2007): 383–384, https://doi.org/10.1111/j.1365-2125.2007.02893.x.

62. G. Manarolo, *Manuale di Musicoterapia: Teoria, Metodi e Strumenti per la Formazione* (Carocci Faber, 2020).

63. R. M. Ryan and E. L. Deci, "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being," *American Psychologist* 55, no. 1 (2000): 68–78, https://doi.org/10.1037/0003-066X.55.1.68.

64. F. Herrera, S. Y. Oh, and J. N. Bailenson, "Effect of Behavioral Realism on Social Interactions Inside Collaborative Virtual Environments," *PRESENCE: Virtual and Augmented Reality* 27, no. 2 (2018): 163–182, https://doi.org/10.1162/pres_a_00324.

65. E. Han, M. R. Miller, C. DeVeaux, et al., "People, Places, and Time: A Large-Scale, Longitudinal Study of Transformed Avatars and Environmental Context in Group Interaction in the Metaverse," *Journal of Computer-Mediated Communication* 28, no. 2 (2023): zmac031, https://doi.org/10.1093/jcmc/zmac031.

66. C. Oh, F. Herrera, and J. Bailenson, "The Effects of Immersion and Real-World Distractions on Virtual Social Interactions," *Cyberpsychology, Behavior, and Social Networking* 22, no. 6 (2019): 365–372, https://doi.org/10.1089/cyber.2018.0404.

67. M. Gao and D. A. Boehm-Davis, "Development of a Customizable Interactions Questionnaire (CIQ) for Evaluating Interactions With Objects in Augmented/Virtual Reality," *Virtual Reality* 27, no. 2 (2023): 699–716, https://doi.org/10.1007/s10055-022-00678-8.

68. F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, "G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences," *Behavior Research Methods* 39, no. 2 (2007): 175–191, https://doi.org/10.3758/BF03193146.

69. H. Coolican, *Research Methods and Statistics in Psychology*, 7th ed. (Routledge, 2019).

70. J. Cohen, "A Power Primer," *Psychological Bulletin* 112, no. 1 (1992): 155–159, https://doi.org/10.1037/0033-2909.112.1.155.

71. D. Benrimoh, F. D. Chheda, and H. C. Margolese, "The Best Predictor of the Future—The Metaverse, Mental Health, and Lessons Learned From Current Technologies," *JMIR Mental Health* 9, no. 10 (2022): e40410, https://doi.org/10.2196/40410.

72. A. Cerasa, A. Gaggioli, F. Marino, G. Riva, and G. Pioggia, "The Promise of the Metaverse in Mental Health: The New Era of MEDverse," *Heliyon* (2022).

73. Y. K. Dwivedi, L. Hughes, A. M. Baabdullah, et al., "Metaverse Beyond the Hype: Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy," *International Journal of Information Management* 66 (2022): 102542, https://doi.org/10.1016/j.ijinfomgt.2022.102542.

74. G. Riva, C. Malighetti, and S. Serino, "Virtual Reality in the Treatment of Eating Disorders," *Clinical Psychology & Psychotherapy* 28, no. 3 (2021): 477–488, https://doi.org/10.1002/cpp.2622.

75. G. Riva and B. K. Wiederhold, "What the Metaverse is (Really) and Why We Need to Know About It," *Cyberpsychology, Behavior, and Social Networking* 25, no. 6 (2022): 355–359, https://doi.org/10.1089/cyber.2022.0124.

76. M. Grossman and W. Wood, "Sex Differences in Intensity of Emotional Experience: A Social Role Interpretation," *Journal of Personality and Social Psychology* 65, no. 5 (1993): 1010–1022, https://doi.org/10.1037//0022-3514.65.5.1010.

77. U. Hess, S. Senécal, G. Kirouac, P. Herrera, P. Philippot, and R. E. Kleck, "Emotional Expressivity in Men and Women: Stereotypes and Self-Perceptions," *Cognition & Emotion* 14, no. 5 (2000): 609–642, https://doi.org/10.1080/02699930050117648.