IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 16 April 2024, accepted 30 June 2024, date of publication 15 July 2024, date of current version 29 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3427698

PERSPECTIVE

Laying the Foundation for Augmented Reality in Music Education

LUC NIJS AND BAHAREH BEHZADAVAL
University of Luxembourg, Belval, 4365 Esch-sur-Alzette, Luxembourg
Corresponding author: Luc Nijs (luc.nijs@uni.lu)

ABSTRACT Since the beginning of humankind, the use and development of technologies have played an essential role in enhancing human abilities and creating new possibilities for action and expression. As such, new technologies have captured the imaginations of educational scholars and practitioners. One of the latest developments in the evolution of digital technologies is the ability to enrich or ''augment'' reality with digital content, thereby establishing an ''augmented reality'' (AR). This perspective article aims to promote and support advancements in the domain of music educational AR by proposing an interdisciplinary knowledge base and its concrete translation into design ideas for further developments in this emerging domain. We explore the integration of Augmented Reality (AR) into Music Education (MusEdAR), highlighting its nascent state and potential for innovation. First, we provide an overview of the current state-of-the-art, focusing on what types of AR are used, what visual content is digitally added to the learning experience, and how movement is integrated. Next, we discuss how to expand the MusEdAR domain. We argue for a solid interdisciplinary knowledge base to advance the design and use of MusEdAR applications and present themes such as embodied learning or developing creativity and associated theories. Based on these theories, we presented practical design ideas.

INDEX TERMS Augmented reality, creativity, embodied learning, music education, participatory sense-making.

I. INTRODUCTION

Since the beginning of humankind, the development and use of technology have played an important role in enhancing human abilities and creating new possibilities for action and expression [1]. The driving force behind today's technological innovation is the development of computers. Since its inception, its capabilities have considerably increased. Computer-based technologies have become more powerful, faster, smaller, and more mobile. In addition, interactions with computers have become increasingly natural and social [2]. The power of these developments lies in making human-computer interaction more natural by matching it to how we function in everyday life without a computer. Consequently, digital technology has become pervasive in human life. Having gone through a digital revolution (Industry 3.0), we are now entering the next stage (Industry 4.0), characterized by a fusion of technologies that increasingly merge the physical, digital, and biological spheres [3].

FIGURE 1. The Virtuality Continuum: from left to right: 1) physical reality, 2) AR: adding digital content to the physical reality, 3) AV: adding real objects to a virtual reality, 4) a completely digital world. Figure created by Elias Nijs.

A rapidly evolving development in the domain of digital technologies is the creation of alternative immersive realities, whether completely digital (Virtual Reality (VR)) or mixing physical and digital reality (eXtended Reality (XR), such as Augmented Reality (AR) or Augmented Virtuality (AV)). Mixed Reality (MR) is an overarching denotation for environments that combine real and virtual objects (Fig. 1).

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

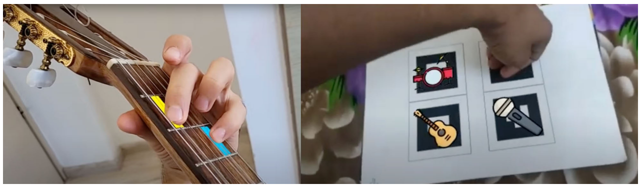

FIGURE 2. Left: GeetAR: using AR to show where to press on the fretboard and with which fingers. See: https://www.youtube.com/watch?v=pxEW3Jyy1TI. Right: Augmented Reality Music Player: using barcode markers that trigger 3D objects or images, allowing a user to play or pause any instrument. See: https://www.youtube.com/watch?v=LBwODrH83bs.

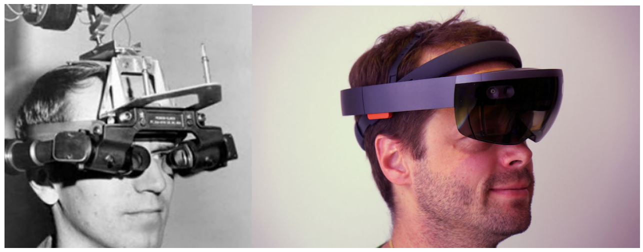

FIGURE 3. Left: Sutherland's Sword of Damocles. Right: contemporary AR Head Mounted Display (HDM).

The possibilities of such alternative reality creating technologies capture the imagination of many educators and educational scholars, who envision new avenues for training or education (e.g., [4]) and research (e.g., [5], [6]) based on the idea that extended realities (XR) such as Virtual Reality (VR) and Augmented Reality (AR) provide a constructive and effective learning environment.

Here, we focus on Augmented Reality, which superimposes digital (visual) content to the actual or "real" environment, whether looking through a screen (e.g., of a smartphone or tablet) or a Head-Mounted Display (HMD) such as AR Smart Glasses, thereby blending the physical and virtual worlds. In recent years, AR has shown a steep development curve driven by ongoing technological advancements [7]. Note that two strategies exist for superimposing digital content onto the real environment: marker-based and markerless. The former uses predefined visual cues (markers) such as QR codes or specific images recognized by the AR system, which are used as anchors to overlay digital information onto the real world (Fig. 2, right). The latter does not require any predefined markers but uses advanced algorithms to understand and interact with the environment, allowing for more spontaneous and versatile AR experiences (e.g., [8]) (Fig.2, left). Augmented Reality (AR) has come a long way since the 1960s, when Ivan Sutherland developed the first head-mounted display, known as the "Sword of Damocles," a device that allows viewing images in 3D (see Fig. 3, left). In the 1990s, Boeing Computer Services created the first augmented reality (AR) system for aircraft construction workers. The 2000s brought AR games and applications (think of Pokémon Go!), while in the 2010s AR revolutionized, for example, the retail and social media industries. Today, AR is used in everything from education to healthcare and has become more accessible than ever thanks to smartphones and AR devices such as Hololens and other AR Smart Glasses (Fig. 3, right). According to Leong ([9], p. 234), it "creates a fresh and new feel that at the same time extends beyond existing expectations and practices." According to Billinghurst and Dünser [10], AR enables the creation of a powerful learning environment that combines complex problem-solving with teamwork. Moreover, AR environments have been shown to be efficient in promoting collaborative learning at different ages and for various subjects (e.g., [11], [12]).

In recent years, AR has gradually entered the music education domain. Scholars and educators are increasingly exploring its possibilities for enhancing and innovating music learning. Although the domain of music educational AR (MusEdAR) emerged more than two decades ago, it is still in its infancy. This perspective article aims to promote and support advancements in the domain of MusEdAR by proposing an interdisciplinary knowledge base and its concrete translation into design ideas for further developments in this emerging domain. To do so, we reviewed the existing literature on MusEdAR to examine the current state of the art, focusing on using visuals and movement in MusEdAR applications. Based on the review, we elaborate on how, in our view, to advance the MusEdAR domain. We argue that there is a need for a solid interdisciplinary knowledge framework to drive advancements in this domain regarding content, visuals, and the integration of body movement. Therefore, we present theories that may constitute such a framework. In addition, we propose concrete design ideas for future developments. We acknowledge that implementing these ideas could be challenging. However, we believe that next to proposing theoretical frameworks, elaborating on possible concrete design ideas may benefit the (Mus)EdAR community.

II. MUSIC EDUCATIONAL AR: CURRENT STATE-OF-THE-ART

As MusEdAR is an emerging domain for research and practice, few overview articles have been published thus far. Apaydinli [13] published a content analysis of 35 music education studies on AR. To reveal existing trends, the author examined their descriptive features, methodological features, and outcomes. Findings show that most studies have focused on piano and guitar to investigate the user experience. The analysis of existing studies shows that AR applications facilitate and promote learning to perform, rendering learning exciting, fun, and motivating. According to the author, existing challenges concern technical issues and learning content.

Turchet et al. [14] adopted a broader perspective by examining music in Extended Realities (XR). To clearly define Musical XR, the authors also included an email interview with eight experts who contributed significantly to the field in both academia and industry. Interestingly, the authors looked deeper into the different sensory modalities (visual, auditory, haptic, and proprioceptive) implemented by music-related XR. Their review analyzed 199 works within the last decade. It focused on primary functions, including composition, performance, entertainment, and education, using a series of conceptual dimensions pertaining to technical, artistic, perceptual, and methodological domains. Their findings regarding AR in education show that the piano is the most prominently investigated. AR applications mainly target adult novice learners, although an increasing number of studies focus on children. According to the authors, challenges concern the pedagogical foundation, accessibility of tools, and focus on creativity.

An interesting approach to reviewing existing studies on augmented piano prototypes is the work of Deja et al. [15]. Similar to Turchet et al. [14], the authors interviewed ten piano teachers and teachers of piano didactics to verify whether existing prototypes address issues exposed by these experts. They found that augmented piano prototypes contribute to the development of synchronized movement and body posture, sight-reading, motivation, and encouraging improvisation. The authors introduce a novel system that contributes to piano improvisation and offer an investigation of spatiotemporal models built from different movements of users' fingers and hands to predict user errors during improvisation.

Apaydinli [13] and Turchet et al. [14] pleaded for broadening the target audience of AR/XR applications, involving not only beginners but also intermediate- and advanced-level students. Apayd1nl1 included students with disabilities. As the most recent research addresses AR in instrumental music education, especially piano and guitar, Apayd1nl1 suggests future research in other fields, such as vocal training and music theory. The author also noted that few studies in this domain have addressed abilities such as improvisation and music comprehension. Turchet et al. [14] argued that there is a need for practices that contribute to new learning experiences (e.g., multi-user for social experiences, engaging more artists and audiences). They also plead for more evaluations to assess their effectiveness.

Furthermore, Apayd1nl1 states that most research is conducted by researchers in the domain of computer science. Like Turchet et al., he pleaded for more collaboration between computer scientists and music educators. Through interdisciplinary research, collaboration between computer scientists and music educators can be made more effective in combining AR and music education.

Interestingly, Turchet et al. acknowledge the need to establish pedagogical practices and define standards for musical AR. In this perspective article, we complement both review studies by focusing on two specific aspects of AR: the use of visuals and movement. First, we provide an overview of the existing work and how it addresses visuals and movement. We searched for articles using Google Scholar and Web of Science. Articles were selected using the following inclusion criteria: i) being published in peer-reviewed journals, conferences, and book chapters; ii) addressing music learning and not music performance or artistic projects; iii) focusing on AR, not VR; and iv) only English articles. Then, we discuss a possible direction for future developments based on knowledge frameworks related to, for example, musicology (e.g., embodied music interaction) and pedagogy (e.g., constraint-led approach). We connect concrete ideas for further MusEdAR development to the proposed knowledge frameworks. Thus, we wish to fuel the discussion on the future of MusEdAR and contribute to its further development. We believe that our work may contribute to the design and development of novel MusEdAR systems and pedagogical practices by providing ideas for setting standards for MusEdAR embedded within a solid theoretical foundation.

A. VISUALS IN MUSEDAR

Most AR applications for music education aim to assist learning through visual elements that provide different types of information (e.g., keys to press or the length of notes). As Turchet et al. [14] stated, most applications have significant visual components.

While visuals come in different forms (static, dynamic, or color-based), they often support instrumental technique, such as correct playing or performance accuracy involving, for example, correct notes, accurate rhythms, or dynamics (e.g., [16], [17], [18]). In addition, visual information is sometimes presented as feedback on user performance [19], [20], [21].

Most AR applications in music education focus on instrumental music, particularly piano and keyboard. They use colored shapes (primarily rectangular) to demonstrate the keys that users should press [17], [18], [16], [19], [22], [23], [24], [25], [26], [27]. The colors were used for different purposes. For example, Trujano et al. [25] and Lu et al. [28] used different colors for chord notes, and Cai et al. [29] used green and red rectangles for the right and left hand, respectively. In addition, some applications use the colors as feedback, showing, for example, notes played correctly, incorrectly, or simply missing [18], [19], [21], [26], [30]. Most of the time, the height of the rectangles (piano roll) indicates the duration of the notes, but in some studies, such as Cai et al. [29], it is related to the amount of pressure to give learners information about the volume of the sound. However, rectangles sometimes contain more information. For instance, in Weing et al. [26], rounded corner rectangles indicated legato notes.

Other research implementing this representing approach towards AR technologies (visual information as guidance for fingering and playing) aimed at learning to play the guitar and electric bass guitar [31], [32], [33], [34].

Visualization in MusEdAR sometimes involves figurative representations for different purposes. For example, Fernandez et al. [23] introduced a virtual character that provides feedback on the user's performance: the character dances at higher scores and stops dancing at lower scores. Das et al. [22] and Marky et al. [17] used 3D models to show the user the correct posture and fingering. The system proposed by Xiao et al. [35], [36] visualizes music as an animated character walking on a physical keyboard.

Another function of the visual components in AR music applications is to disclose musical expression. For instance, Santini [37] presented different interfaces in which virtual elements, such as a sphere and water's undulatory motion, represented a visual expression of the music, both for the player and the audience.

Visualization also plays a specific role in MusEdAR applications as an aid for synchronization. The AR Drum Kit system developed by Yamabe et al. [38] uses visually animated elements, such as lines and ripples, to support correct timing. Furthermore, they used different colors to provide feedback to the user regarding the proper timing of hitting the drum pads. Gali et al. [39] also created an MR system to promote motor synchronization in children by following visual elements, such as geometric shapes, projected onto a floor.

Hopkins et al. [40] explored AR-based ways to deal with the shortcomings (e.g., delay, lacking the richness of interpersonal interaction and collaboration) of platforms for remote collaboration such as Zoom®. The authors developed AR Drum Circle, an AR platform that aims to facilitate collaborative remote drumming by using Nreal Light AR glasses to provide a 3-dimensional visualization or avatar of the drummers: player 1 sees the avatar of player two and vice versa. In addition, Gerry et al. [41] explored using an avatar, focusing on learning how to play the piano. The authors propose the ADEPT system, a Mixed Reality (MR) application in which students share a first-person, embodied perspective with a piano teacher (avatar) to facilitate learning the proper finger, hand, wrist, and torso configurations to produce various sounds on the piano. In this way, the user's perspective is not entirely dominated by that of another virtual avatar or a real person but rather enhanced by a virtual projection of another person's embodied perspective overlaid on their own perspective. The user perceives the virtual hands of the teacher as positioned on a tangible keyboard. A blue highlight visually represented the currently pressed key. The user's view encompasses the pianist's face and the upper body displayed on the upright grand piano.

Regarding the use of visualizations, it is important to consider the devices used. AR "augments" a real-world setting by adding digital content (e.g., virtual objects) that can be seen through the screen of a hand-held device (e.g., smartphone, tablet) or Head-Mounted Displays (HMD) such as AR Glasses (Fig. 4). AR Smart Glasses (binocular HMD) can be video see-through (i.e., seeing reality through a video that is first captured by one or two cameras mounted onto the display) or optical see-through (i.e., a graphical overlay on the real world, preserving the direct view of the world without perspective conversion in viewpoint and field of view). In the domain of music education, screens have been adopted in different ways, using mobile phones, tablets, and monitor screens. For instance, Farinazzo Martins et al. [42] introduced four small screen-based applications called Music-AR to support the development of music perception, including timbre, pitch, and sound intensity. The system involved a computer, monitor screen, and camera. Lu et al. [28] designed an AR game, ChordAR, to teach children about harmony. Their system operated in Windows, Android, and iOS. Guclu et al. [43] developed an application installed on tablets to teach recorder notation.

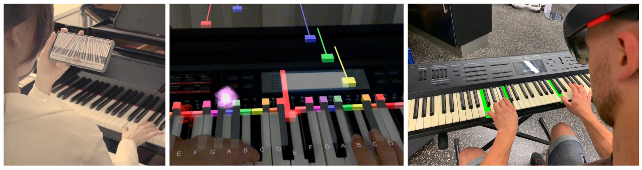

FIGURE 4. FIGURE 4. Left: Playano ®, a virtual piano assistant for smartphones, visualizing notes directly onto the piano. Middle: Magic Keys ®, an AR piano learning app for the Meta Quest VR headset. Right: a system developed by Marky et al. (2021) [17], using an HMD to augment the piano keys with color.

In different studies, users did not look through or at the device to see virtual content. Instead, visual elements were projected onto various surfaces using projectors. Most of the time, the surfaces are instruments. For example, Weing et al. [26] and Xiao et al. [36] used projectors in AR systems for piano. Yamabe et al. [38] and Zhang et al. [44] applied AR projectors to drum kit and Guqin, respectively. Gali et al. [39] adopted an interesting approach, using a projection system to foster motor synchronization in children. In this study, the visual content (colorful geometric shapes as well as images of glitter clouds) accompanied by music was projected onto a large floor (6 × 6 m), and a group of four children moved between different specified spots.

HMDs have mostly been used in piano and keyboard learning (e.g., [21], [24], [30]). Microsoft's HoloLens prevails among these HMD devices. Other HMDs include Trivisio ARvision-3D [19], Nreal Light [40], Samsung Odyssey head-mounted display [45], HTC Vive [18], [21], and Epson Moverio BT300 [27].

B. MOVEMENT IN MUSEDAR

Although music performance has started incorporating gesture-based AR, its implementation in music education remains limited. After reviewing the scarce number of MusEdAR apps and platforms, it is evident that existing AR applications mainly aim to monitor postures and gestures while playing an instrument, most often involving the piano. For example, Huang et al. [46] used virtual hands in a marker-less AR-based piano teaching application for beginners. The learner could practice piano by following the virtual fingers on the keyboard. Kerdvibulvech [47] used Microsoft

Kinect® to track the hands while piano playing, allowing them to interact with a virtual piano. Marky et al. [17] used motion tracking to inform piano students of their hand posture and finger movement to help them during their practice sessions. Cai et al. [29] also proposed an AR-supported system for the student and the teacher in a piano class that created a real-time model of the teacher's hands and enabled the student to mimic the teacher's fingering. Gerry et al. [41] created the ADEPT system that focused on training muscle memory in beginner piano students. Its primary objective is to teach students how to execute precise piano movements to produce desired sounds. ADEPT aims to enhance students' understanding of their bodies and movements, enabling them to build a fine-grained awareness of their physical abilities, similar to that of a skilled pianist.

In certain cases, AR systems for hand posture and fingering have been used for the guitar. Del Rio-Guerra et al. [32] and Martin-Gutierrez et al. [34] introduced AR systems for learning to play chords on a guitar, focusing on fingering. In addition, Zhang et al. [44] introduced ChinAR to help students use appropriate fingering when learning how to play the Guqin, a traditional Chinese instrument. By using a 3D camera, the system also allowed the learner to control different options (e.g., switch between play and pause) by using hand gestures (e.g., thumb up). Johnson et al. [45] provided guidance to the user by displaying virtual hands while playing the theremin (an instrument played without physical contact) in their mixed reality system called MR:emin.

Interestingly, Xiao et al. [35], [36], did not use the piano player's movement directly but, inspired by the Dalcroze approach to music learning [48], augmented the piano with visuals that suggest movement, namely walking in different ways. In the Dalcroze approach, walking to music in various expressive ways is considered fundamental to promoting musical understanding [48].

In some cases, a correct posture is not the main aim of MusEdAR, but the objective is to engage with music through the body. For instance, in a simple AR/VR simulation system, Chung et al. [20] engaged the user in interactive rhythm games through movement and foot stepping. Users were expected to match their movements (stepping on a pair of foot stepping boards) to the rhythm patterns, guided by falling virtual objects. Hopkins et al. [40] developed AR Drum Circle, an AR platform intended to create a sense of presence and collaboration for users in remote drumming experiences using avatars. Although this research discussed the importance of body gestures and physical cues (e.g., head nods and facial expressions) in musical collaboration, it was nevertheless unable to represent these aspects of body language in their avatars. Gali et al. [39] developed Mandala, an MR pre-interactive (non-interactive visual/auditory cues) system for children focusing on full-body and interpersonal synchronization through rhythmic activities. This study aimed to promote children's interpersonal entrainment by

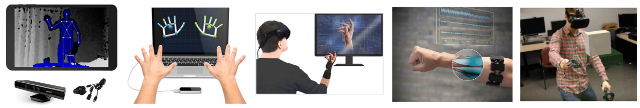

FIGURE 5. Different ways of tracking motion in AR. From left to right: Kinect® camera, Leap Motion®, inertial 7DOF motion sensor, Myo® and Hololens® controller.

following visual elements (temporal patterns and precision) in different tasks accompanied by music.

Furthermore, Santini [49] discussed GesturAR, an application that introduces a new concept of musical notation, that is, embodied interactive notation, and allows users to create music through gestures and to store, recall, and organize tracked movement trajectories in 4D (space and time). GesturAR uses an HTC Vive Pro headset and a Leapmotion®. Avanzini et al. [50] developed a mobile-based AR application called AREmbody to improve children's perception of tonal harmony. This system allows children to enhance their awareness of chords by bodily engaging them in music games. For this purpose, AREmbody applied visual symbols instead of notations for each chord of different scale degrees.

AR can integrate motion tracking in different ways, such as using cameras (e.g., Microsoft Kinect®), tracking devices (e.g., Leap Motion®), inertial motion sensors (e.g., gloves), muscle sensors (e.g., MYO armband), or controllers that accompany the HMD (Fig. 5). MusEdAR implements different types of motion-tracking devices. The use of cameras is twofold. Sometimes, the camera is a separate device integrated into the entire setup. For instance, Molloy et al. [18] and Rigby et al. [21] used the ZED Mini stereo, a lightweight depth and motion sensing camera, in combination with the HTC Vive (HMD). Similar research used other types of cameras, such as the Creative Senz3D camera used by Zhang et al. [44]. Sometimes, the camera is integrated into a Head-Mounted Display (HDM) or in a handheld device, such as the mobile AR application developed by Martin-Gutierrez et al. [34] to facilitate learning the acoustic guitar. The HoloLens (e.g., [17], [24], [30]), and NReal Light [40] are see-through HMDs that use cameras to track the users' movements.

MusEdAR systems also use specific devices to track motion, particularly when addressing postures and gestures. Some motion trackers used in MusEdAR are Vive trackers (e.g., [18], [49]) and Leap Motion (e.g., [29], [51]). Gali et al. [39] used a handheld luminous pointer that helps the system detect the user (the child) using lights of different colors.

Furthermore, some MusEdAR applications use other types of sensors to input data. For example, Chung et al. [20] applied a pair of foot-stepping boards equipped with pressure sensors in a rhythm-based AR/VR simulation system. AREmbody [50] used an accelerometer and gyroscope sensors to track the positions of the cue points captured by the camera.

In the previous sections, we discussed MusEdAR in relation to visuals and movement. Before discussing future work, it is necessary to address sound and the engines used to develop the systems.

Sound is an essential element in MusEdAR. Accordingly, sound detection is implemented in MusEdAR systems to, for example, track users' accuracy when playing music [34], [49]. Consequently, MusEdAR systems may involve additional equipment such as microphones [37], [40].

The Music Instrument Digital Interface (MIDI) is another vital technology used in music-related AR systems [52]. Many AR systems as well as other digital applications in music education have implemented this protocol. For example, in AR systems designed for learning piano or other instruments, the MIDI input and output are used to transfer data between devices (e.g., [21], [24], [45]). AR applications generally use a variety of engines. In the domain of MusEdAR, some applications operate with Unity and Vuforia Engine [29], [32]. Sometimes, Unity is combined with other engines. For example, Johnson et al. [45] and Wallevik [53] developed their MR applications using Unity and the Mixed Reality Toolkit (MRTK). Zeng et al. [27] used ARToolkitX for the FunPianoAR to identify and track a marker.

III. DISCUSSION: LOOKING AHEAD

Although the use of movement in the domain of MusEdAR is still scarce, visualizations play an essential role [14]. Existing visualizations appear to reflect a common use in the domain of interactive music educational technologies, often driven by the belief that computer-based monitoring (using quantification of sound and movement) bypasses teachers' shortcomings in providing feedback, such as proneness to ambiguous interpretation and delayed feedback (e.g., [54]). Arguably, this focus on monitoring (posture, accuracy) shows that AR-based educational activities may still adhere to a master-apprentice approach, which has been criticized for being too teacher-centered, with a focus on technique for reproductive imitation [55]. For example, the typical piano roll type of added visualizations (Fig. 6) promotes perhaps more the exact rhythmic reproduction of the music, whereby developing the technical proficiency to "play in time" with the piano roll is more important than interpretation and expressive timing. In this way, important aspects of learning, such as learner autonomy, self-efficacy, self-regulation, individual artistic voice, and creativity, are often neglected (see [55], [56]).

In addition, when addressing playing technique, visuals are often concurrent and focus on promoting an internal focus (e.g., [17], [29]). Research has indicated that visuals can have a degrading effect when promoting an internal focus (on one's movements, body parts, or the feel of the movement) [57]. By contrast, feedback that promotes an external focus is more effective than a reduced feedback frequency [58].

In our view, to deal with the above critiques, the design and implementation of MusEdAR require a solid

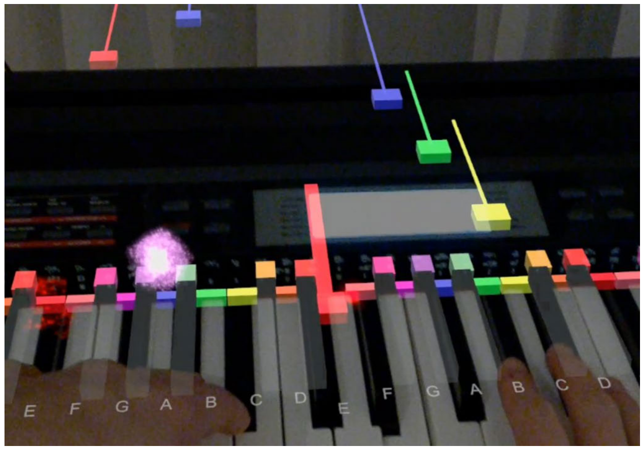

FIGURE 6. Typical visualization of a piano roll, superimposed on the keyboard or piano using an HMD.

TABLE 1. Possible themes and related theories supporting the design and development of MusEdAR applications.

Developing an embodied musical understanding	Embodied Music Cognition (e.g., [64]) Embodied Music Pedagogy (e.g., [65], [66]) Conceptual Metaphor theory (e.g., [67]
Develop creativity	Embodied Musical Creativity (e.g., [68]) Theories on affordances (e.g., [69]
Promote participatory sense-making	Participatory sense-making (e.g., [70]
Embrace non-linearity	Constraints-led approach (e.g., [71] Embodied Music Pedagogy (e.g., [65], [66])

pedago-musicological framework to achieve specific objectives. In this section, we elaborate on the possible objectives in relation to specific theoretical frameworks, implementing a multimodal approach using visuals and movement (see Table 1 for an overview). According to Turchet et al. [14], it is particularly the combination of sensory modalities that promotes immersion but also introduces a "unique set of affordances" that may contribute to a meaningful and powerful learning experience (e.g., [59], [60]). This is related to the modality principle of the cognitive theory of multimedia learning, according to which presenting learning content in different modalities (e.g., auditory and visual) helps learners process such content more effectively [61], [62]. Visual feedback helps build referential connections, that is, connections that integrate stimuli in different sensory modalities (e.g., visual and auditory) with one another and with relevant prior knowledge. According to Mayer et al. [63], this is essential to constructivist learning.

Furthermore, we believe that the current focus on performance may be complemented by MusEdAR that promotes the development of musical perception and understanding and associated meaningful responses to music (e.g., movement, drawing), collaborative learning, and musical creativity. These are important when learning objectives target the learner's comprehensive growth in the music field [65]. Musical learners interact with music by perceiving, responding, performing, and creating music. Accordingly, it is crucial to design and implement technological tools, here MusEdAR, to approach this comprehensive growth.

A. DEVELOPING AN EMBODIED MUSICAL UNDERSTANDING

While AR applications for music learning often use visuals to represent aspects of music (e.g., note duration, pitch), and the integration of movement most often involves monitoring specific playing gestures, the combination of motion tracking and visualization can be used to promote the development of musical understanding instead of playing accuracy.

An embodied understanding of music implies the ability to feel music from within (e.g., [72]), that is, through the bodily experience of different aspects of the music, such as structure, melodic contour, or rhythm [73]. Such understanding is rooted in an embodied interaction with music that, according to Leman [64], is based on three basic mechanisms that allow and support the transformation of a stream of sounds into a meaningful musical experience, also called the process of enactment (see also [74]). The first basic mechanism is entrainment or the process of being pulled towards synchronization [75]. Entrainment does not necessarily occur automatically or smoothly. Indeed, some conditions must be met. For example, one must be able to detect salient moments in the music (e.g., beat), perform rhythm or movement patterns, and adapt the performance of such patterns to fit the overall timing framework [76]. According to Leman ([64], p. 114), entrainment enables three sensorimotor mechanisms: finding, keeping, and even being the beat. Visuals can support the process of entrainment by providing representations of how the music unfolds over time, thereby helping to find and keep the beat. They can address different aspects of timing, such as the beat or the time between beats. In particular, when using macro beats, the time between these beats can support entrainment, and accordingly, also synchronization (see Fig. 7). This can also be related to a second mechanism, namely alignment, and in particular, the distinction between phase and interphase alignment ([64], p. 156). The former concerns the correspondence of one's actions (e.g., stepping, tapping the foot or hand) to salient elements in the music, most often the beat. As shown in Fig. 7, these actions do not necessarily coincide with each beat. In a triple meter, such an action can coincide with the first beat. In a quadruple meter it can, for example, be on the 1st and 3rd beat, or on the 2nd and the 4th beat. The latter, interphase alignment, concerns what happens between salient markers (Fig. 7, right). The basic mechanisms of alignment and entrainment are closely connected to a third basic mechanism, namely prediction. Prediction concerns the ability to sense what comes next in the music and anticipate the outcome of a movement, such as hitting a drum or reaching a point in space to the beat. Leman [64] distinguished three interaction situations that contribute to prediction. First, attenuation occurs when, owing to successful prediction, the self-generated sensory information resulting from playing or moving to music no longer requires conscious monitoring, freeing up attention for other elements in the musical interaction. Second, facilitation occurs when the interaction with music is made easier by facilitating the prediction of one channel in music, such as timing, leaving aside other channels, such as melody or harmony. Finally, disambiguation occurs when uncertainties in the music in terms of perceptual or affective-expressive content (e.g., varied meters, emotions) that might impede prediction and interfere with pattern recognition and emergence, are reduced. Visuals can contribute to attenuating self-generated sensory information by promoting an external focus through the visualization of action outcomes (performing, moving), to facilitation by showing how different aspects (e.g., meter, harmony, melody) of the music unfold, and to disambiguation by providing visual prompts that disentangle the complexity of the music. Similarly, movements, possibly invited by the visualization, may contribute to each interaction situation [74].

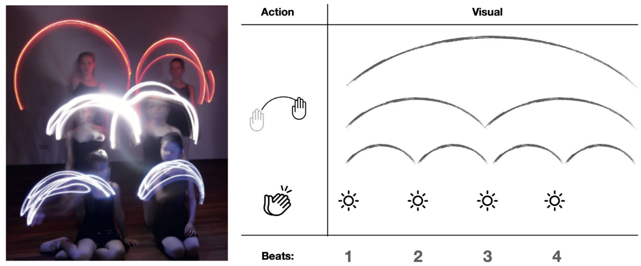

FIGURE 7. Left: Actions and visuals showing different modes of alignment. Right: possible realization with light bulbs; see [77].

Interestingly, Gallagher and Lindgren [78] pointed out the benefits of using Mixed Reality to learn through movement based on enactive metaphors, that is, a metaphor that can turn into action or that we bring into existence through our action. This viewpoint aligns with the embodied nature of music cognition, according to which the body's engagement and movement are integral and fundamental to the understanding and creation of music [64]. Moreover, music is often experienced and described metaphorically (e.g., [79]). The use of gesture-based AR may appeal to the metaphorical nature of musical experience, creating an embodied foundation for learning by prompting (''body cueing'') learners to move to and engage with the music in novel ways [80]. Different types of metaphors [67], such as metaphors based on the CONTAINER image scheme (inside and outside, e.g., ''the melody is IN the key of C''), the SOURCE-PATH-GOAL image scheme (the relation of movement from start to goal; e.g., ''Melody is movement ALONG A PATH),'' or the VERTICALITY image scheme (''up'' and ''down'' relations; e.g., ''The music is GOING UP''). Image schemes refer to

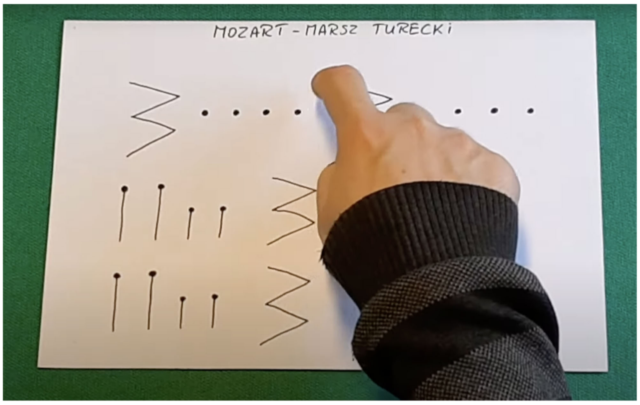

FIGURE 8. A child following rhythm-informed shapes (musicogram) representing the unfolding of the music it is listening to.

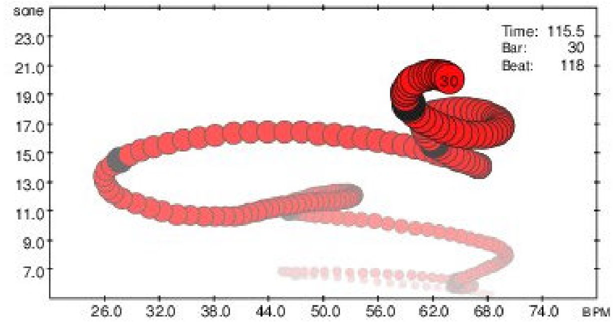

FIGURE 9. The Performance Worm is an animation that depicts expressive aspects of a musical performance. The position of the head of the worm represents the tempo (horizontal axis) and loudness (vertical axis). Dixon and Goebl [86].

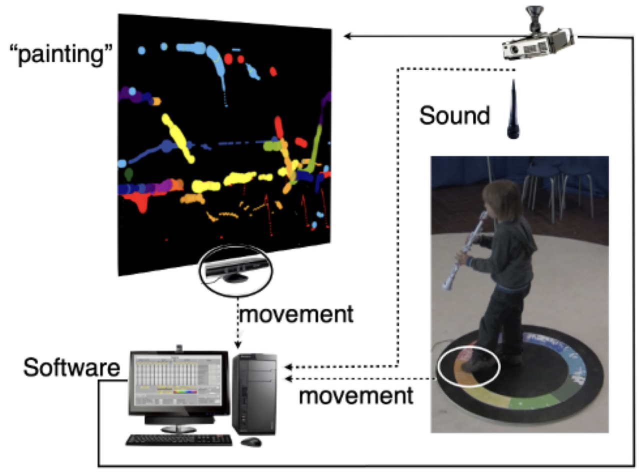

FIGURE 10. The Music Paint Machine, an interactive music educational application that allows musicians to create a digital painting by combining music and movement.

"recurring patterns of our sensory-motor experience" and are fundamental to our understanding of the world (e.g., music) [81]. The use of such metaphors in the implementation of gesture-based AR (see also: [82]) is, in our view, an avenue to explore as it may not only be beneficial to the development of music understanding (e.g., on harmonic progression and modulation; e.g., [83]) but also to the development of musical creativity (see next section; see also: [84]).

Furthermore, existing offline, possibly technology-mediated, musical activities may inspire the design of visualizations and motor involvement. Musicograms or musicomovigrams are one source of inspiration [85]. These are static visual representations of music (incl., rhythm and melody) that invite bodily engagement with the visuals while listening by following the shapes with a finger (Fig. 8). Arguably, interacting with such a musicogram supports the above-mentioned basic mechanisms: movements need to be performed smoothly within the timing of the music, the shapes help align with elements in the music, and seeing the musicogram facilitates predicting what comes next in the music. Interestingly, AR can transform such visualizations into dynamic visualizations. Indeed, while visualizations in AR can be static and represent shapes to be followed, shapes can also unfold during the interaction, thereby appealing more to predictive mechanisms and contributing to the basic mechanism of prediction. The unfolding of such shapes can also be steered by a student and display aspects of the student's actions. Such actions can include moving, singing, or playing an instrument. A source of inspiration can be the "Air Worm" (Fig. 9), as presented in the work of Dixon et al. [86]. Another example is the Music Paint Machine (Fig. 10), an interactive system that invites students to create a digital painting by combining playing an instrument with movement (torso and feet) [87], [88]. While the system monitors movement and sound to create the painting, the visualizations do not aim solely to monitor performance but also to provide creative visualizations that invite students to play with musical parameters and, as such, to develop musical and instrumental skills. Arguably, the combination of music and sound supports alignment by, for example, visually showing how a melodic contour moves up and down. Furthermore, the system promotes the development of audiation skills and, thereby, prediction. Indeed, defined initially as "hinged mosaic relationships linked to networks of comparative pattern structures," ([89], p. 6) audiation or "thinking music in the mind with understanding" is based on the assimilation and comprehension of the sounds one is hearing. However, according to Gordon [89], audiation not only involves listening to and understanding music but also predicting what will come next, based on one's familiarity with tonal and rhythmic conventions. An additional source of inspiration is the Dalcroze approach. For example, to learn about meter and timing, the Dalcroze approach may pair students. One student has a hand drum and moves the drum in alignment with the music. Another student is invited to hit the drum on the beat. This is a powerful combination of doing (moving the drum vs. hitting the drum) and seeing (gestures to the music vs. beats). Both actions rely on different timing mechanisms: while hitting the drum uses event-based timing, moving to the music involves emergent timing. The former operates as an internal clock (think: metronome) based on internally sensed time intervals; the latter involves sensing time through the bodily experience of a movement [90]. However, the timing mechanism of one student's action is always complemented by that of the perceived action (executed by the other student). Arguably, although further research is needed, the timing mechanism of the perceived action is simultaneously triggered based on the activity of the mirror neurons, a specific type of neurons that promote the simulation (including timing) of an observed individual's actions ([91], see also [41], [92]).

Translating such activities into AR, whereby the other person or object is replaced by digital visualization or whereby several learners share the same visual content, may not only foster new approaches to AR but also facilitate research on, for example, timing and synchronization (see [93]). Integrating movements and visualization combines enactive and iconic representations of music (see [94]), rather than focusing on symbolic representations (e.g., notation). Using movement and visualizations to promote musical sense-making supports the interactive dialectics of expressive interaction with music, a process that evokes emergent patterns in music, from which new configurations may emerge [64]. As such, it contributes to understanding and making sense of music. It also paves the way for a creative interaction with music.

B. DEVELOPING CREATIVITY

According to Nijs et al. [68], movement is a powerful way of developing musical creativity. Their basic idea is that inviting people to move to the music is a way to provoke flexible and creative navigation of the musical affordance landscape. A musical affordance landscape refers to an individual's perception of multiple affordances in music [95]. Musical affordances are considered to be those aspects of music that define what we can do with it [96]. In other words, they express the action possibilities in the musical environment that are specified by (1) specific elements in the structure of the music and (2) the sensorimotor abilities that support detecting and responding to these elements [96], [97]. According to Glăveanu ([98], p. 196), creativity is "a process of perceiving, exploiting, and generating novel affordances during socially and materially situated activities," whereby affordances become apparent only when one is engaged with the environment.

Through the integration of visualizations and movement, AR can support the navigation of the musical affordance landscape, emphasizing both obvious and hidden elements in music. Think of the many dynamic visualizations of music that show, for example, the different voices in the music of Bach, Beethoven and so many others (Fig. 11). Shapes, colors, and sizes can be used to prompt responses. For example, lines horizontally connecting objects might support focusing on related elements in the music (e.g., melody); vertical grouping and colors may prompt disambiguation

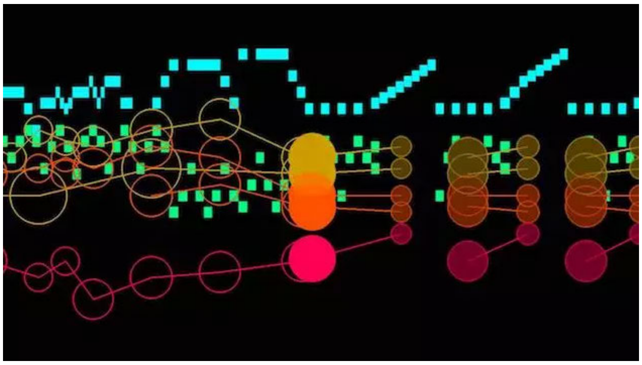

FIGURE 11. Stephen Malinowski's visualization of the Rondo from Mozart's Piano Quartet in G Minor, K.478. See: https://www.youtube.com/watch?v=OW6lEEWO1AU.

and focus on a certain aspect (e.g., melody vs. harmony). Moreover, when learners share the same visual content and their physical actions are monitored through sensors, they might respond to different elements and, for example, upon a certain prompt, switch between musical elements and associated responses. The combination of music, visuals, and motor responses is of interest in relation to the theory of attention dynamics as proposed by Large and Jones ([99]; see also [64], p. 106). According to this theory, listeners' attention fluctuates over time in response to rhythmic patterns they encounter. In other words, attention is dynamic. Visualizations can help capture and maintain learners' attention by providing an immersive and dynamic complement to the auditory experience. They can map the temporal progression of music, aiding in the understanding of its structure; highlight rhythmic patterns, making it easier for listeners to follow and predict the timing of musical events; represent upcoming musical phrases or highlight repetitive motifs, enhancing the listener's predictive capacities; and reduce cognitive load on the auditory system, allowing for a more balanced distribution of attentional resources.

Moreover, presenting learners with visualizations that make sense of the complexity of music as exemplified above (e.g., showing different voices), may manipulate the specification (e.g., what action is possible; e.g., moving one hand or both hands) and selection (determining which action to perform; e.g., moving one hand) of possible actions, for example in relation to specific educational goals (e.g., focus on the bass vs. melody, on major vs. minor chords). Both are constitutive of purposeful affordance navigation, that is, the process by which individuals interact with their environment by perceiving and utilizing affordances (e.g., [69], [95]). This relates to metastability, which not only allows flexible switching between actions but also context-sensitive selective openness. Indeed, it enables the behavioral flexibility necessary for intentional affordance navigation. Visualizations that support making sense of the complexity of the music may foster a metastable state, which entails the ability to possess different co-existent pattern-forming tendencies

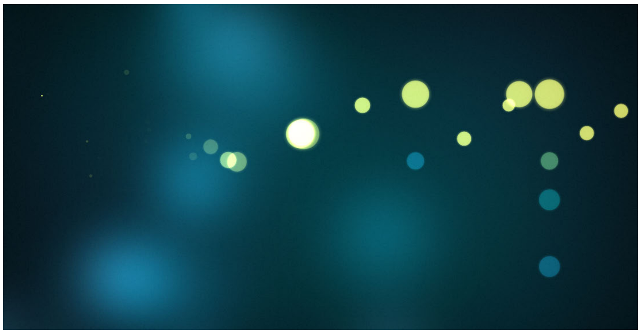

FIGURE 12. Screenshot of Touch Pianist. Try it out at: https://touchpianist.com/.

[100]. By supporting context-sensitive selective openness and flexible switching between activities, metastability plays an important role in creative affordance navigation, whereby new affordances are discovered or created [98]. Note that, above, the underlying idea is that learners respond, whether individually or in a group, to music but do not create or play music. Creating or playing music using AR is a step further that can be implemented using gesture-based MusEdAR.

An inspirational example is the application Touch Pianist (Fig. 12), which allows users to perform a specific composition by tapping the space bar or screen. Each tap triggers the next event in the music. This application not only allows users to perform the composition as they know it but also to explore expressive timing. Touch Pianist is, however, to be played individually, but could be developed into a collaborative MusEdAR application whereby students play together and, as such, need to coordinate their playing to make the composition sound meaningful. In the next section, we discuss collaborative work and the co-creation of meaning. A next step is that, rather than collaboratively re-creating existing music through the time-based triggering of musical events, learners can create their own music and visuals. Music can be used to invent visual creations or the visuals can inspire to compose or improvise music (see also the Music Paint Machine, e.g., [88]). An envisaged gesture-based MusEdAR application could allow students to select or create their own sounds, visual objects, and actions to trigger the sounds and objects. This could promote artistic creation and creativity, support meaningful interactions, and foster social skills. In the next section, the co-creation of meaning, as an important aspect of joint music-making, is addressed.

C. PROMOTING PARTICIPATORY SENSE MAKING

Participatory sense-making involves participating in each other's sense-making of music through shared, active involvement in music ([70], p. 31). It allows for the co-creation and joint understanding of music, as it develops over time. To foster participatory sense-making, it is important to create a close coupling and coordination between individual learners and, as such, to create a shared context based on intersubjective intentionality [101].

Interacting with music lends itself perfectly to participatory sense-making. Making sense of music is, in essence, an active, social, and culturally constructed process whereby "patterns of coordination arise, evolve, break down, and re-occur during social encounter," ([102], p. 5) thereby contributing to the transformation of a stream of sounds into a meaningful experience [64], [66]. As Keller and Repp [103] states, referring to rhythmic joint activity as a specific class of joint action, this involves the coordination of each other's thoughts and movements in space and time in view of communicating or changing something in the environment (e.g., music, a joint response to music). Successful interaction between musicians involves the continuous bidirectional exchange of information that "allows effective coupling into an organic whole with characteristic traits."([93], p.4) In this way, jointly responding to music (e.g., through movement) and playing or creating music together (e.g., improvising) promote the transformation of me-agency into we-agency. That is, an individual's sense of agency ("I did it!") becomes valued within the collective agency of the group ("We did it!") ([104]; see also [74]).

While the use of AR in interaction with music is not a condition sine qua non for the emergence of participatory sense-making, it may contribute to fostering the joint creation and ownership of musical meaning in different ways. Elaborating on effective interpersonal coupling and coordination, Liebermann-Jordanidis et al. [105] pointed out the need for (1) information coupling between one's actions and the actions of others, and (2) integration of joint action outcomes while maintaining a distinction between self and others. The former (integration) allows monitoring and evaluating joint action outcomes against shared performance goals, whereas the latter (segregation) allows a sense of agency and autonomy. By adding virtual content to the learning environment, MusEdAR can promote both integration and segregation. Note that, while participatory sense-making does not require the integration of movement-based activities, we suggest that future work in MusEdAR could benefit from integrating gesture-based interactions, especially regarding participatory sense-making. After all, movement-based activities not only promote individual explorations of the music but also lend themselves perfectly to engaging in participatory sense-making through joint movement (e.g., [97], [106]).

To illustrate this, the work of Van Kerreboeck et al. [93] is useful. Their study involved an augmented reality visual stimulus in the shape of a drum circle, based on which participants engaged in tapping a 2:3 polyrhythmic pattern at a fixed tempo. Rotating virtual spheres on the drum circle supported individual tapping of isochronous sequences, showing when to tap. A successful performance, that is, tapping within a certain margin (62.5ms before or after the visual tap instruction), the stimulus gradually faded. Unsuccessful tapping resulted in the return of the visual stimulus. Thus, individual and joint action outcomes affected the visual outcomes. Here, visual and auditory coupling are combined, whereby the visuals play an important role in increasing synchrony (individually with the tap). The results showed improved performance (decreased prediction error, increased movement energy, and increased sense of agency). Interestingly, visual contact positively influenced experiential qualities and prosocial effects, and the level of partner realism (virtual vs. real) affected the sense of shared agency and self-other merging [93].

In our view, van Kerrebroek's work shows the potential of using MusEdAR in an educational context, in view of learning specific musical skills (e.g., rhythmic) and developing prosocial abilities through participatory sense-making. This is particularly the case when using devices such as AR smart glasses that allow seeing the real environment in combination with the virtual content.

D. EMBRACE NON-LINEARITY

Music learning is often organized in a linear manner, leading students step by step through a predetermined learning path. Presenting an Embodied Music Pedagogical framework, Bremmer and Nijs [65], [66] [107] introduced Non-Linear Pedagogy in the domain of music education. The authors adopt a dynamical systems theory perspective to music learning and teaching, seeing the learning processes as emerging from the learner's goal-oriented, situated, adaptive actions in a dynamic learning environment. At the center of this process is the interaction between the learner(s), teacher, and the learning content. Importantly, the teacher becomes part of the process, not as an external regulator but as a constitutive part of the whole.

One of the core ideas of the Embodied Music Pedagogy as presented by Bremmer and Nijs [65], [107] is that the dynamics of a music lesson can be shaped through a set of constraints, that is, individual, task, and environmental constraints (see also: [108]). Individual constraints refer to the characteristics of an individual, such as their level of perceptual, emotional, and cognitive functioning, or motor abilities [109]. Task constraints refer to the goal of a specific task, feedback on the task, asking questions, or materials used during a learning experience [110], [111]. Environmental constraints refer to the physical factors surrounding learners that shape or limit their behavior [110].

Arguably, augmenting the real environment with virtual objects enables the manipulation of the different constraints and, as such, supports shaping the dynamics of the learning process. First, AR visuals may introduce environmental constraints by enriching the learning environment through prompts that guide and steer behavior. For example, the previously described work of Van Kerrebroeck et al. [93] uses visual prompts that steer tapping in a binary or ternary meter. In addition, the often-used piano rolls (e.g., [25]) may steer users towards exact timing rather than expressive timing. However, the latter might be stimulated using Touch Pianist or an AR version of the application (see section on Developing Creativity). Interestingly, augmenting the learning environment with digital content enables sensory manipulations that are impossible in the real world. For example, based on learners' performances or on specific learning goals, colors or their characteristics (e.g., grayscale, brightness, transparency), shapes, or sounds may change [112].

Second, gesture-based MusEdAR may introduce task constraints, inviting users to, for example, focus on a certain voice (e.g., melody versus bass), or to respond with continuous versus discrete movements to elements in the music. As Fortuna and Nijs [113], [114] showed, using verbal vs. movement and discrete vs. continuous movement, not only movement but also the way one moves to music affects musical sense-making (as shown in their drawing on music).

Importantly, the visuals in MusEdAR can function as augmented feedback, that is, information on an action provided by an external source, providing both knowledge of performance (i.e., information relating to an action parameter) and knowledge of results (i.e., information relating to an action outcome) [115]. Both may not only complement intrinsic information, that is, information available to the senses when performing an action, but may also provide new information with the aim of supporting the learning process. While the former may help to reinforce an existing action repertoire (e.g., following the melody with the movement of one hand and refine synchronization), the latter may support broadening learners' spectrum of available responses by inviting them to explore different action possibilities to find a solution (e.g., aligning the movement of two hands to melody and bass) (see also [112] on motor learning in neurological rehabilitation).

Finally, MusEdAR may help manipulate individual constraints, that is, intentionally alter the individual constraints of an individual, to encourage exploration, adaptation, and problem-solving (see [116], p. 65). One way to do so is to tailor an activity to the attributes of the learners, such as their physical, psychological, and emotional characteristics. For example, the visualizations may be adapted according to age or level of expertise.

Another way to do so is to manipulate what learners bring to a learning task, thus allowing them to explore new ways to address the task. For example, when moving to music, manipulations may involve changing which body parts to use (e.g., only hands vs. whole body) or freezing vs. freeing the degrees of freedom in the joints, thereby limiting the movement possibilities in response to music.

Different techniques can be used to manipulate individual constraints. One technique is differentiation, which augments the environment in accordance with learners' individual abilities. For example, the visuals provided may differ among the learners. In this case, smart technology may play an important role. When a system learns about a user's abilities, it may adapt to how it augments the environment. For example, a learner who has difficulties synchronizing movements to the music may be "granted" a larger error margin that gradually narrows, while a learner who synchronizes easily might be presented with a much smaller error margin or with a more complex visualization. Such an approach may impact skill-challenge balance and, as such, affect motivation, avoiding frustration or boredom (e.g., [117]). Here, the Miror-impro system is a possible source of inspiration [118]. This application allows children to improvise with a virtual copy of themselves as a kind of partner, discovering what elements in the replies of MIROR-Impro stay the same or what changes. The system learns about the style of the children and, based on a Markovian generation scheme, recombines their musical input to ''answer'' the child in the same style and, as such, engage in a musical dialogue [119].

Another technique involves a gradual increase in the degree of difficulty or complexity. This is an important element in learners' development. According to Bronfenbrenner [120], development requires the repeated occurrence of increasingly complex activities (proximal processes) over time. For example, augmenting the real environment with visualizations of music (see Fig. 11) may be simple in the beginning (e.g., showing simplified rhythm patterns) and gradually become more complex (e.g., gradually adding rhythmic complexity).

Task variability is also a useful technique. According to Schmidt's [121] variability of practice hypothesis, variability in practice conditions is important for learning. Varying a task may address different individual characteristics. For example, changing the type of movement or the element in the music to respond to can be imposed using different visual prompts (e.g., dots vs. lines, connecting different elements).

Shaping the different constraints in view of a well-defined learning goal and adopting techniques, such as those described above have a unique and specific impact on learners' actions by creating different learning opportunities. In our view, AR can contribute to a constraints-led approach to music teaching and learning by implementing the use of visuals and movement.

Finally, an important note regarding the constraints-led approach is the interconnectedness between the different constraints. According to the constraint-led approach, the interaction between the different constraints allows a learner to self-organize when attempting to deal with a given learning opportunity (e.g., [109]). As the interaction between the constraints shapes emergent behavior, it is important to understand how this works and to build on that understanding to design learning environments that support learners. Consequently, designing visual and movement-based elements in gesture-based MusEdAR requires careful consideration of the constraints and their interactions.

E. DEVICES

Currently, many AR-based applications use handheld devices, such as smartphones or tablets, whereas head-mounted AR devices (HMD-AR-devices) are quite scarcely used. Considering the above ideas for the advancement of MusEdAR, we favor using AR Smart Glasses (ARSG), even though some of the ideas could be realized with a screen-based application or a hand-held device. ARSG, such as Microsoft HoloLens, Everysight Raptor, or Google Glass, are devices that are ''worn like regular glasses and merge virtual information with physical information in a user's view field.''([122], p. 172) They give users ''sustained, hands-free access to data and can transmit and receive information wirelessly.'' ([123], p. 47) In our view, AR Smart Glasses offer promising potential for using AR in music education. First, they provide an alternative to sedentary screen time [124]. Excessive sedentary screen time may lead to attentional, cognitive, and physical problems [125]. Second, they allow real-world group interaction, which is an important element in music learning (e.g., [126]). Indeed, in AR, interaction entails a mixture of virtual and physical objects, people, and environments, and consequently still allows for onsite face-to-face communication, rather than being mediated by, for example, avatars [127]. Third, they allow a hands-free AR experience. This enables playing an instrument (e.g., percussion instruments) and freely moving in response to music. In addition, it allows the combination of AR Smart Glasses with inertial motion sensors, thus allowing movement tracking of the hands (e.g., [128]).

We also favor the integration of motion tracking using sensors. This enables not only interaction with digital content but also the quantification of movement in the interaction in view of scientifically investigating bodily engagement. Note that some AR Smart Glasses enable interactions using mid-air gestures (e.g., pointing, swiping, waving hands), but these are rather limited and fatiguing [129]. Combining AR glasses with motion sensors allows richer interaction with digital content (e.g., [130]). This is an important asset, particularly for music learning. First, movement is an essential part of music learning [66] and, as such, ideally is part of the interaction possibilities of a music AR application. In addition, the complexity of music requires a rich bodily interaction (e.g., playing an instrument). Furthermore, quantifying movement allows learners' progress to be tracked in terms of, for example, rhythmic accuracy or synchronization. In other words, using data analytics, it is possible to integrate and connect fine-grained data (e.g., movement and sound) about different learners' engagement in the AR environment and different aspects (e.g., synchronization, quantity, and quality of motion) of bodily engagement with music. Moreover, as this type of data is immediately accessible, it can be fed into the interaction loop within the AR environment, enabling feedback on the joint interaction (e.g., changing colors or displays based on the level of the learner's synchronization; [93].

Importantly, while AR smart glasses may seem to be a viable avenue for future developments, their use poses several challenges when implementing the above ideas. A first challenge concerns the computational power of the device, which is required to process data and generate adequate and fine-grained visuals. On the one hand, owing to a limited battery capacity, the power efficiency is critical. However, to deliver a smooth AR experience, processing speed of, for example, graphical data and sensor input is essential. As such, the challenge is to maximize power efficiency without jeopardizing the quality of the presented content (see also: [131]). In our view, using a solid theoretical framework to design visuals may support the development of simple but effective visualizations. In addition, as Danielsson et al. ([132], p. 2) argue, "processing power and batteries will improve over time to a point where all the necessary performance for most uses can fit in a device that can be worn like regular glasses, at which point there would be no added value in increasing the size of a pair of ARSG."

Another challenge concerns age. Unfortunately, HDM-ARs are mostly considered for use above the age of 12 years. For example, on their webpage "Product safety warnings and instructions," Microsoft states that HoloLens is not intended for use by children under the age of 13. Consequently, many AR applications focus on adult learners. Nevertheless, some studies have used HDM-ARs at younger ages. For example, Lauer et al. [133] used Microsoft's HoloLens 2 with elementary school children to investigate the device usability. They found that its use positively affects the children's activity-related achievement emotions. They also concluded that usability and efficiency are related to technical aspects, but also to different interaction modes such as tapping, voice command, and air-tapping, and children's preferences for them. Similarly, Munsinger et al. [134] tested the HoloLens 2 with 5th-grade elementary school children and investigated the differences between interaction modes (i.e., voice, gesture, and clicker) based on the measurement of their input errors and elapsed time to complete a tutorial and game.

In the domain of music education, only a few studies have applied HDM to children. For instance, Molero et al. [30] used Microsoft HoloLens in the HoloMusic XP system, a gamified XR system to learn how to play the piano. The system was not designed specifically for children, but the experimental research was conducted with participants aged between 6 and 24 years. Wallevik [53] designed and developed a prototype called TappyBeatsXR, an MR application using the HoloLens 2, for all skill levels to create music. Although the researcher tested the prototype with five adults, the system was firmly suggested for children and teenagers because it did not require prior knowledge of music and the interface was easy to use.

In addition, Lauer et al. [133] discussed the fact that most devices are designed for adults and their body dimensions (e.g., interpupillary distance), which may lead to distorted interaction with AR-objects (e.g., inability to reach).

An important concern regarding the age of HDM users is the possible risks or disadvantages for their visual development. Young children, who are in a critical period of visual development, may strain their eyes when viewing inferior-quality 3D content for long stretches at a time. However, according to the American Optometric Association and many eye specialists, there is no evidence that viewing 3D images harms a child's eyes. Evidently, a limited time spent (e.g., 30 min) and a careful introduction are beneficial (see also [133]). It is important to note that MusEdAR should not aim to replace existing practices without technology but rather to enrich current practices. In addition, a viable method is to generate rich but simple visualizations.

F. VISUALS AND GESTURES

In addition to elements such as computational power, battery, and visual development, the use of simple visualizations is also beneficial to the learning process itself, more specifically in relation to cognitive load. Visuals can have a degrading effect on learning (e.g., [135], [136]) following the possible introduction of extra cognitive load or by the stimulation of an internal focus of attention. This may occur when the information that reaches the learner is too complex (e.g., [137], [138]). This might increase the intrinsic (load placed on the learner by the nature of the materials being learned) and even extraneous (load that is unnecessary) cognitive load of the learning content. Consequently, the germane cognitive load (required by the methods used to present new knowledge to a learner) is reduced at the expense of genuine learning [139]. For these reasons, it is important that the visual feedback provided by MusEdAR satisfies Bruner's instructional principle of economy, which states that the amount of information that must be held in mind and processed to achieve comprehension must be reduced [140], [141]. Visual feedback should be easy to interpret [138], [142]. Additionally, visual feedback must also satisfy Bruner's instructional principle of power, stating that learners must be stimulated to make connections between topics that seem separate [142]. For example, while a novice might assume that pitch and loudness are separate qualities of a sound, visual feedback may clarify that playing louder might have an effect on pitch. While the principle of economy might reduce extraneous cognitive load by leaving out information that is not immediately relevant, the principle of power might reduce the intrinsic cognitive load by clarifying the relationships between the different elements and thereby reducing the perceived complexity of the information. This allows for the integration of these elements into a unified schema.

Buchner et al. [143], who systematically analyzed existing research on AR and cognitive load, state that "compared to other technologies, AR seems to be less cognitively demanding and also leads to higher performance." However, the authors acknowledge that the results are based on media comparison studies that have been criticized for years, and conclude that AR glasses can unnecessarily increase cognitive load.

In gesture-based MusEdAR, visualizations are often connected to the bodily actions of the learner. As already mentioned, this may lead to an internal focus (i.e., focusing on how an action is performed rather than on its result) and, as such, to degraded learning. In addition to carefully considering what to visualize to reduce cognitive load and avoid an internal focus, it is also important to consider motor load when using gesture-based MusEdAR. Motor load refers to the level of difficulty or complexity of the motor tasks involved. It can be manipulated, for example, by defining the number of possible movements (e.g., only one hand versus two hands)(e.g., [144]), by changing the speed of involved movements (e.g., moving to every beat versus moving every four beats; see Fig. 7) or asking for more difficult coordination (doing the same movement with the two hands vs. doing something different with each hand, switching between hands) (e.g., [145]).

Importantly, motor and cognitive load were found to be related. Reiser et al. [146] showed that an increase in the complexity of movements causes an increased workload in the cognitive system, thereby reducing the availability of cognitive resources for another - cognitive - task. Similarly, when playing, for example, a scale on one's instrument, introducing a difficult stepping pattern may cause extra cognitive load (remembering the steps) and interfere with the correct playing of the scale (e.g., leaving out a flat or sharp and as such no longer playing in the right key).

However, when interacting with music, movement may decrease cognitive load. In this context, Leman ([64], p. 140) refers to facilitation as an interactive situation during which a cognitive task is outsourced to a sensorimotor loop to facilitate prediction. A typical example is timekeeping or keeping count of the beats. One option is mentally counting the beats (''one, two, three, four, one, two, three, four'') to keep a steady beat. However, this might reduce one's capacity to deal with other musical aspects (e.g., interpretation) and, as such, lead to less expressive performance (see also [147], p. 14]). An often-used alternative is to tap the feet. Another option is to perform a stepping pattern that clearly makes one feel the difference between the beats (see [55]).

Given the above, the successful design and use of MusEdAR requires understanding and careful consideration of motor and cognitive load, as it can affect both cognitive processes and motor performance, especially in combination with visualizations.

IV. CONCLUSION

In this article, we provide a brief overview of the existing AR applications for music education. While existing systematic overviews [13], [14], [15] consider work in the domain of MusEdAR until 2020, we have also integrated some more recent work. In addition, we took a different stance, focusing mainly on visualization and movement.

Each existing overview presents suggestions for future research and development. We complemented and deepened such suggestions for future works, pleading for pedagogical or musicological knowledge frameworks as possible foundations for the design and implementation of MusEdAR, focusing on visualization and movement. In our view, what is at stake is to lay out a strong interdisciplinary and research-based theoretical framework for pedagogical sound design and use of MusEdAR. This is important, as Garzón [148], referring to several studies, stated that there is a lack of pedagogical approaches for integrating AR into learning activities. According to the author ([148], p. 54), ''educational applications based on AR must transcend technological aspects, as the technology by itself does not ensure success in the learning process.''

We believe that the presented work might spur new and concrete directions in the domain of MusEdAR, whereby aspects such as musical understanding, creativity, participatory sense-making, and non-linearity in learning might be embraced. Thus, the focus on performance may be broadened to include the development of perceptual and responsive skills and music creation. Furthermore, it is important to find inspiration in existing classroom practices, especially when integrating movement. Please note that, while we believe the presented knowledge frameworks are important, they are not the sole knowledge bases that may need consideration. Others might be theories on cognitive load (see [62] and [143]), the cognitive theory of multimedia learning (e.g., [149]), or differential learning (see [150]).

Based on the different theoretical frameworks presented, we provided some concrete examples of the possible novel use of visualization and movement-based MusEdAR rather than focusing on general principles. While we acknowledge that the presented examples might involve important hardware and software challenges, we believe that conceptually designing possible futures may help steer the furthering of an exciting domain such as MusEdAR.

In addition to providing a solid pedagogical foundation, embedding MusEdAR into a clear theoretical and research-based framework can open new avenues for investigating creativity development, music perception, understanding, performance accuracy, and literacy.

ACKNOWLEDGMENT

The authors thank Elias Nijs for creating Figure 1. They also thank the editor and reviewers for their critical and constructive comments on this manuscript.

REFERENCES

[1] N. Jacobs, M. Lanzing, and M. Stevens, ''Thinking about technology: An introduction to philosophy of technology,'' in *Second Thoughts: First Introductions to Philosophy*, M. Sie and B. Engelen, Eds., Tilburg, The Netherlands: Open Press Tilburg University, Feb. 2023. [Online]. Available: https://doi.org/10.26116/second-thoughts-openpresstiu-jote

[2] P. Dourish, *Where the Action Is: The Foundations of Embodied Interaction*. Cambridge, MA, USA: MIT Press, 2001.

[3] C. Martin and H. Leurent, ''Technology and innovation for the future of production: Accelerating value creation,'' in *Proc. World Economic Forum*, Mar. 2017, pp. 1–38.

[4] V. Geroimenko, *Augmented Reality in Education*. Cham, Switzerland: Springer, 2020.

[5] A. J. Martingano and S. Persky, "Virtual reality expands the toolkit for conducting health psychology research," *Social Personality Psychol. Compass*, vol. 15, no. 7, Jul. 2021, Art. no. e12606.

[6] F. T. J. M. Zaal and R. J. Bootsma, "Virtual reality as a tool for the study of perception-action: The case of running to catch fly balls," *Presence*, vol. 20, no. 1, pp. 93–103, Feb. 2011.

[7] R. Vertucci, S. D'Onofrio, S. Ricciardi, and M. De Nino, "History of augmented reality," in *Springer Handbooks*, A. Y. C. Nee and S. K. Ong, Eds., New York, NY, USA: Springer, 2023, pp. 35–50.

[8] P. Q. Brito and J. Stoyanova, "Marker versus markerless augmented reality. Which has more impact on users?" *Int. J. Human–Computer Interact.*, vol. 34, no. 9, pp. 819–833, Sep. 2018.

[9] S. Leong, "Navigating the emerging futures in music education," *J. Music, Technol. Educ.*, vol. 4, no. 2, pp. 233–243, Feb. 2012.

[10] M. Billinghurst and A. Duenser, "Augmented reality in the classroom," *Computer*, vol. 45, no. 7, pp. 56–63, Jul. 2012.

[11] N. M. N. Affendy and I. A. Wanis, "A review on collaborative learning environment across virtual and augmented reality technology," *IOP Conf. Ser., Mater. Sci. Eng*, vol. 551, Aug. 2019, Art. no. 012050, doi: 10.1088/1757-899X/551/1/012050.

[12] H. Alhumaidan, K. P. Y. Lo, and A. Selby, "Co-design of augmented reality book for collaborative learning experience in primary education," in *Proc. SAI Intell. Syst. Conf. (IntelliSys)*, Nov. 2015, pp. 427–430.

[13] K. Apaydinli, "Content analysis of music education studies related to augmented reality technology," *J. Educ. Technol. Online Learn.*, vol. 6, no. 2, pp. 447–481, May 2023.

[14] L. Turchet, R. Hamilton, and A. Çamci, "Music in extended realities," *IEEE Access*, vol. 9, pp. 15810–15832, 2021.

[15] J. A. Deja, S. Mayer, K. Č. Pucihar, and M. Kljun, "A survey of augmented piano prototypes: Has augmentation improved learning experiences?" *Proc. ACM Human-Comput. Interact.*, vol. 6, no. ISS, pp. 226–253, Nov. 2022.

[16] A. Simion, A. Iftene, and D. Gîfu, "An augmented reality piano learning tool," in *Proc. Int. Conf. Human-Comput. Interact.*, Bucharest, Romania, 2021, pp. 134–141.

[17] K. Marky, A. Weiß, and T. Kosch, "Supporting musical practice sessions through HMD-based augmented reality," 2021, *arXiv:2101.00874*.

[18] W. Molloy, E. Huang, and B. C. Wünsche, "Mixed reality piano tutor: A gamified piano practice environment," in *Proc. Int. Conf. Electron., Inf., Commun. (ICEIC)*. Auckland, New Zealand: IEEE, Jan. 2019, pp. 1–7.

[19] J. Chow, H. Feng, R. Amor, and B. C. Wünsche, "Music education using augmented reality with a head mounted display," in *Proc. 14th Australas. User Interface Conf.*, vol. 139, Adelaide, SA, Australia, Jan. 2013, pp. 73–79.

[20] S.-M. Chung, H.-G. Lin, T.-L. Tsou, C.-T. Wu, and C.-H. Huang, "Experiencing musical rhythm through interactive installation and AR/VR game development," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*. Santa Cruz de Tenerife, Spain: IEEE, Apr. 2018, pp. 384–388.

[21] L. Rigby, B. C. Wünsche, and A. Shaw, "PiARno—An augmented reality piano tutor," in *Proc. 32nd Austral. Conf. Human-Comput. Interact.* New York, NY, USA: ACM, Dec. 2020, pp. 481–491.

[22] S. Das, S. Glickman, F. Y. Hsiao, and B. Lee, "Music everywhere—Augmented reality piano improvisation learning system," in *Proc. Intern. Conf. New Interfaces Musical Expression (NIME)*. Copenhagen, Denmark: Aalborg Univ. Copenhagen, 2017, pp. 511–512.

[23] C. A. T. Fernandez, P. Paliyawan, C. C. Yin, and R. Thawonmas, "Piano learning application with feedback provided by an AR virtual character," in *Proc. IEEE 5th Global Conf. Consum. Electron.* Kyoto, Japan: IEEE, Oct. 2016, pp. 1–2.

[24] D. Hackl and C. Anthes, "HoloKeys-an augmented reality application for learning the piano," in *Forum Media Technology*. St. Pölten, Austria: St. Pölten Univ. Applied Sciences, 2017, pp. 140–144.

[25] F. Trujano, M. Khan, and P. Maes, "ARPiano efficient music learning using augmented reality," in *Innovative Technologies and Learning* (Lecture Notes in Computer Science), vol. 11003, T. Wu, Y. Huang, R. Shadiev, L. Lin, and A. Starčič, Eds., Cham, Switzerland: Springer, 2018, pp. 3–17.

[26] M. Weing, A. Röhlig, K. Rogers, J. Gugenheimer, F. Schaub, B. Könings, E. Rukzio, and M. Weber, "P.I.A.N.O.: Enhancing instrument learning via interactive projected augmentation," in *Proc. ACM Conf. Pervasive Ubiquitous Comput.* New York, NY, USA: Association for Computing Machinery, Sep. 2013, pp. 75–78.

[27] H. Zeng, X. He, and H. Pan, "FunPianoAR: A novel AR application for piano learning considering paired play based on multi-marker tracking," *J. Phys., Conf. Ser.*, vol. 1229, no. 1, May 2019, Art. no. 012072.

[28] Y. Lu, X. Wang, J. Gong, and Y. Liang, "ChordAR: An educational AR game design for children's music theory learning," *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1–9, Feb. 2022.

[29] M. Cai, M. A. Amrizal, T. Abe, and T. Suganuma, "Design and implementation of AR-supported system for piano learning," in *Proc. IEEE 8th Global Conf. Consum. Electron. (GCCE)*. Osaka, Japan: IEEE, Oct. 2019, pp. 49–50.

[30] D. Molero, S. Schez-Sobrino, D. Vallejo, C. Glez-Morcillo, and J. Albusac, "A novel approach to learning music and piano based on mixed reality and gamification," *Multimedia Tools Appl.*, vol. 80, no. 1, pp. 165–186, Jan. 2021.

[31] O. Cakmakci, F. Bèrard, and J. Coutaz, "An augmented reality-based learning assistant for electric bass guitar," in *Proc. 10th Int. Conf. HCI*, Crete, Greece, Jun. 2003.

[32] M. S. Del Rio-Guerra, J. Martin-Gutierrez, V. A. Lopez-Chao, R. Flores Parra, and M. A. Ramirez Sosa, "AR graphic representation of musical notes for self-learning on guitar," *Appl. Sci.*, vol. 9, no. 21, p. 4527, Oct. 2019.

[33] J. R. Keebler, T. J. Wiltshire, D. C. Smith, S. M. Fiore, and J. S. Bedwell, "Shifting the paradigm of music instruction: Implications of embodiment stemming from an augmented reality guitar learning system," *Frontiers Psychol.*, vol. 5, p. 471, May 2014.

[34] J. Martin-Gutierrez, M. S. Del Rio Guerra, V. Lopez-Chao, R. H. Soto Gastelum, and J. F. Valenzuela Bojórquez, "Augmented reality to facilitate learning of the acoustic guitar," *Appl. Sci.*, vol. 10, no. 7, p. 2425, Apr. 2020.

[35] X. Xiao, B. Tome, and H. Ishii, "Andante: Walking figures on the piano keyboard to visualize musical motion," in *Proc. Intern. Conf. New Interfaces Musical Expression (NIME)*, Jun. 2014, pp. 629–632.

[36] X. Xiao, P. Puentes, E. Ackermann, and H. Ishii, "Andantino," in *Proc. The 15th Int. Conf. Interact. Design Children*. New York, NY, USA: ACM, Jun. 2016, pp. 37–45.

[37] G. Santini, "Augmented piano in augmented reality," in *Proc. Int. Conf. New Interfaces Musical Expression (NIME)*, Jul. 2020, pp. 411–415.

[38] T. Yamabe, H. Asuma, S. Kiyono, and T. Nakajima, "Feedback design in augmented musical instruments: A case study with an AR drum kit," in *Proc. IEEE 17th Int. Conf. Embedded Real-Time Comput. Syst. Appl.*, vol. 2. Toyama, Japan: IEEE, Aug. 2011, pp. 126–129.

[39] O. Gali, S. B. Ercan, G. Atherton, L. Cross, B. Sayis, P. Heaton, and N. Pares, "Mediating interpersonal synchronization in children through a full-body mixed reality system: Analysis of the pre-interactive mandala experience," *PRESENCE, Virtual Augmented Reality*, vol. 32, pp. 35–51, Dec. 2023.

[40] T. Hopkins, S. C. C. Weng, R. Vanukuru, E. A. Wenzel, A. Banic, M. D. Gross, and E. Y.-L. Do, "AR drum circle: Real-time collaborative drumming in AR," *Frontiers Virtual Reality*, vol. 3, Aug. 2022, Art. no. 847284.

[41] L. Gerry, S. Dahl, and S. Serafin, "ADEPT: Exploring the design, pedagogy, and analysis of a mixed reality application for piano training," in *Proc. 16th Sound Mus. Comp. Conf. (SMC)*. Malaga, Spain: Sound and Music Computing Network, May 2019, pp. 241–249.

[42] V. Farinazzo Martins, L. Gomez, and A. G. D. Corrêa, "Teaching children musical perception with MUSIC-AR," *EAI Endorsed Trans. e-Learn.*, vol. 2, no. 5, p. e3, Mar. 2015.

[43] H. Guclu, S. Kocer, and O. Dundar, "Application of augmented reality in music education," *Eurasia Proc. Sci. Technol. Eng. Math.*, vol. 14, pp. 45–56, Dec. 2021.

[44] Y. Zhang, S. Liu, L. Tao, C. Yu, Y. Shi, and Y. Xu, "ChinAR: Facilitating Chinese Guqin learning through interactive projected augmentation," in *Proc. 3rd Int. Symp. Chin. CHI*. New York, NY, USA: Association for Computing Machinery, Apr. 2015, pp. 23–31.

[45] D. Johnson, D. Damian, and G. Tzanetakis, "Evaluating the effectiveness of mixed reality music instrument learning with the theremin," *Virtual Reality*, vol. 24, no. 2, pp. 303–317, Jun. 2020.

[46] F. Huang, Y. Zhou, Y. Yu, Z. Wang, and S. Du, "Piano AR: A markerless augmented reality based piano teaching system," in *Proc. 3rd Int. Conf. Intell. Human-Machine Syst. Cybern.*, vol. 2. Hangzhou, China: IEEE, Aug. 2011, pp. 47–52.

[47] C. Kerdvibulvech, "An innovative real-time mobile augmented reality application in arts," in *Proc. 4th Int. Conf. Aug. Real. Virt. Real. Comp. Graph. (AVR)*. Ugento, Italy: Springer International Publishing, Jun. 2017, pp. 251–260.

[48] M.-L. Juntunen, "Ways to enhance embodied learning in Dalcroze-inspired music education," *Int. J. Music Early Childhood*, vol. 15, no. 1, pp. 39–59, Jul. 2020.

[49] G. Santini, "Composition as an embodied act: A framework for the gesture-based creation of augmented reality action scores," in *Proc. 17th Sound Mus. Comp. Conf. (SMC)*, 2020, pp. 357–363.

[50] F. Avanzini, A. Baratè, M. Cottini, L. A. Ludovico, and M. Mandanici, "Developing music harmony awareness in young students through an augmented reality approach," in *Proc. 4th Int. Conf. Comput.-Human Interact. Res. Appl.*, 2020, pp. 56–63.

[51] F. Permana, H. Tolle, F. Utaminingrum, and R. Dermawi, "Development of augmented reality (AR) based gamelan simulation with leap motion control," *Int. J. Interact. Mobile Technol. (iJIM)*, vol. 13, no. 12, p. 120, Dec. 2019.

[52] D. M. Huber, *The Midi Manual. A Practical Guide to Midi Within Modern Music Production*. New York, NY, USA: Routledge, Oct. 2020.

[53] J. H. Wallevik, "TappyBeatsXR: Using a modern mixed reality approach to develop a virtual music creation environment for non-musicians," Ph.D. dissertation, Dept. Comput. Sci., Elect. Eng. Math. Sci., Western Norway Univ. Appl. Sci., Bergen, Norway, 2022.

[54] D. Hoppe, M. Sadakata, and P. Desain, "Development of real-time visual feedback assistance in singing training: A review," *J. Comput. Assist. Learn.*, vol. 22, no. 4, pp. 308–316, Aug. 2006.

[55] L. Nijs, "Moving together while playing music: Promoting involvement through student-centred collaborative practices," in *Becoming Musicians—Student Involvement and Teacher Collaboration in Higher Music Education*, S. Gies and J. Sætre, Eds., Oslo, Norway: Norges musikkhøgskole, 2019, ch. 1, pp. 239–260.

[56] L. Nijs and M. Leman, "Interactive technologies in the instrumental music classroom: A longitudinal study with the music paint machine," *Comput. Educ.*, vol. 73, pp. 40–59, Apr. 2014.

[57] G. Wulf and R. Lewthwaite, "Attentional and motivational influences on motor performance and learning," in *Art in Motion: Musical and Athletic Motor Learning and Performance*, vol. 1, A. Mornell, Ed., Lausanne, Switzerland: Peter Lang, 2009, pp. 95–117.

[58] G. Wulf, N. Mcconnel, M. Gärtner, and A. Schwarz, "Enhancing the learning of sport skills through external-focus feedback," *J. Motor Behav.*, vol. 34, no. 2, pp. 171–182, Jun. 2002.

[59] K. Ng, O. Larkin, T. Koerselman, and B. Ong, "i-Maestro gesture and posture support: 3D motion data visualisation for music learning and playing," in *Proc. EVA London Int. Conf.*, J. P. Bowen, S. Keene, and L. MacDonald, Eds. London, U.K.: London College of Communication, Jul. 2007, pp. 11–13.

[60] P.-T. Yu, Y.-S. Lai, H.-H. Tsai, and Y.-H. Chang, "Using a multimodal learning system to support music instruction," *Educ. Technol. Soc.*, vol. 13, no. 3, pp. 151–162, Jul. 2010.

[61] T. de Jong, "Cognitive load theory, educational research, and instructional design: Some food for thought," *Instructional Sci.*, vol. 38, no. 2, pp. 105–134, Mar. 2010.

[62] J. Sweller, J. J. G. van Merrienboer, and F. G. W. C. Paas, "Cognitive architecture and instructional design," *Educ. Psychol. Rev.*, vol. 10, no. 3, pp. 251–296, 1998.

[63] R. E. Mayer, R. Moreno, M. Boire, and S. Vagge, "Maximizing constructivist learning from multimedia communications by minimizing cognitive load," *J. Educ. Psychol.*, vol. 91, no. 4, pp. 638–643, 1999.

[64] M. Leman, *The Expressive Moment: How Interaction (with Music) Shapes Human Empowerment*. Cambridge, MA, USA: MIT Press, 2016.

[65] M. Bremmer and L. Nijs, "Embodied music pedagogy. A vision on the dynamic role of the body in music education," in *European Perspectives on Music Education: Music is What People Do*, T. Buchborn, S. Schmid, G. Brunner, and T. De Baets, Eds., Helbling, 2022, pp. 29–46.

[66] M. Bremmer and L. Nijs, "Embodiment in music education," *J. Rech. Educ. Art.*, vol. 2, Feb. 2024.

[67] G. Lakoff and M. Johnson, *Metaphors We Live by*. Chicago, IL, USA: Univ. Chicago Press, 2008.

[68] L. Nijs, N. Grinspun, and S. Fortuna, "Developing musical creativity through movement: Navigating the musical affordance landscape," *Creativity Res. J.*, pp. 1–25, Jan. 2024.

[69] E. Rietveld and J. Kiverstein, "A rich landscape of affordances," *Ecological Psychol.*, vol. 26, no. 4, pp. 325–352, Oct. 2014.

[70] A. Schiavio and H. De Jaegher, "Participatory sense-making in joint musical practice," in *The Routledge Companion to Embodied Music Interaction*, M. Lesaffre, P.-J. Maes, and M. Leman, Eds., New York, NY, USA: Routledge, 2017, pp. 31–39.

[71] J. Y. Chow, K. Davids, R. Hristovski, D. Araújo, and P. Passos, "Nonlinear pedagogy: Learning design for self-organizing neurobiological systems," *New Ideas Psychol.*, vol. 29, no. 2, pp. 189–200, Aug. 2011.

[72] W. Bowman, "Cognition and the body: Perspectives from music education," in *Knowing Bodies, Moving Minds*, 1st ed., L. Bresler, Ed., Dordrecht, The Netherlands: Springer, 2004, pp. 29–50.

[73] J. Shepherd, "How music works. Beyond the immanent and the arbitrary an essay review of music in everyday life," *Act. Critic. Theory Music Educ.*, vol. 1, no. 2, pp. 2–18, Dec. 2002.

[74] L. Nijs and M. Bremmer, "Embodiment in early childhood music education," in *Music in Early Childhood: Multi-disciplinary Perspectives and Inter-disciplinary Exchanges. International Perspectives on Early Childhood Education and Development*, S. Young and B. Ilari, Eds., Cham, Switzerland: Springer, Jul. 2019, ch. 5, pp. 87–102.

[75] M. Clayton, R. Sager, and U. Will, "In time with the music: The concept of entrainment and its significance for ethnomusicology," *ESEM CounterPoint*, vol. 1, pp. 1–45, Nov. 2004.

[76] J. Phillips-Silver and L. J. Trainor, "Hearing what the body feels: Auditory encoding of rhythmic movement," *Cognition*, vol. 105, no. 3, pp. 533–546, Dec. 2007.

[77] L. Nijs, M. Bremmer, and J. van den Dool, *Singewing Space, Augmented Blended-Learning Approach to Music Learn.* Amsterdam, The Netherlands: Lectoraat Kunsteducatie, 2018.

[78] S. Gallagher and R. Lindgren, "Enactive metaphors: Learning through full-body engagement," *Educ. Psychol. Rev.*, vol. 27, no. 3, pp. 391–404, Sep. 2015.

[79] M. Antovic, "Metaphor in music or metaphor about music: A contribution to the cooperation of cognitive linguistics and cognitive musicology," in *Metaphors We Study: Contemporar*, 2014.

[80] R. Lindgren, M. Tscholl, S. Wang, and E. Johnson, "Enhancing learning and engagement through embodied interaction within a mixed reality simulation," *Comput. Educ.*, vol. 95, pp. 174–187, Apr. 2016.

[81] M. Johnson, "The philosophical significance of image schemas," in *From Perception to Meaning*, B. Hampe, Ed., Berlin, Germany: Walter de Gryurter, 2005, pp. 15–34.

[82] H.-Y. Pai and C.-M. Huang, "How to use camera view changing and morphing technology to embody the visual representation of metaphor in augmented reality," in *Proc. Knowl. Innov. Design Culture*. Singapore: World Scientific, Oct. 2021, pp. 177–181.

[83] K. Wilkie, S. Holland, and P. Mulholland, "Towards a participatory approach for interaction design based on conceptual metaphor theory: A case study from music interaction," in *Music and Human-Computer Interaction* (Springer Series on Cultural Computing), S. Holland, K. Wilkie, P. Mulholland, and A. Seago, Eds., London, U.K.: Springer, Apr. 2013, pp. 259–270.

[84] L. Zbikowski, "Music, metaphor, and creativity," in *Performing Metaphoric Creativity Across Modes and Contexts*, L. Hidalgo-Downing and B. K. Mujic, Eds., Amsterdam, The Netherlands: John Benjamins, Jun. 2020, ch. 3, pp. 43–70.

[85] G. Boal-Palheiros and J. Wuytack, "Effects of the 'musicogram' on children's musical perception and learning," in *Proc. 9th Int. Conf. Music Percept. Cogn.*, Bologna, Italy, Aug. 2005, pp. 1264–1271.

[86] S. Dixon, W. Goebl, and G. Widmer, "The 'air worm': An interface for real-time manipulation of expressive music performance," in *Proc. Int. Comp. Music Conf. (ICMC)*, Barcelona, Spain, Sep. 2005.

[87] L. Nijs, B. Moens, M. Lesaffre, and M. Leman, "The music paint machine: Stimulating self-monitoring through the generation of creative visual output using a technology-enhanced learning tool," *J. New Music Res.*, vol. 41, no. 1, pp. 79–101, Mar. 2012.

[88] L. Nijs, "Dalcroze meets technology: Integrating music, movement and visuals with the music paint machine," *Music Educ. Res.*, vol. 20, no. 2, pp. 163–183, Mar. 2018.

[89] E. Gordon, *Learning Sequences in Music: A Contemporary Music Learning Theory*. G.I.A., 2007.

[90] L. H. Baer, J. L. N. Thibodeau, T. M. Gralnick, K. Z. H. Li, and V. B. Penhune, "The role of musical training in emergent and event-based timing," *Frontiers Human Neurosci.*, vol. 7, p. 191, May 2013.

[91] F. C. Nather, J. L. O. Bueno, E. Bigand, and S. Droit-Volet, "Time changes with the embodiment of another's body posture," *PLoS ONE*, vol. 6, no. 5, May 2011, Art. no. e19818.

[92] K. Overy and I. Molnar-Szakacs, "Being together in time: Musical experience and the mirror neuron system," *Music Perception*, vol. 26, no. 5, pp. 489–504, Jun. 2009.

[93] B. Van Kerrebroeck, K. Crombé, S. Wilain, M. Leman, and P.-J. Maes, "The virtual drum circle: Polyrhythmic music interactions in extended reality," 2023, *arXiv:2308.01889*.

[94] J. Bruner, "The course of cognitive growth," *Amer. Psychol.*, vol. 19, no. 1, pp. 1–15, 1964.

[95] G. Pezzulo and P. Cisek, "Navigating the affordance landscape: Feedback control as a process model of behavior and cognition," *Trends Cogn. Sci.*, vol. 20, no. 6, pp. 414–424, Jun. 2016.

[96] J. Krueger, "Affordances and the musically extended mind," *Frontiers Psychol.*, vol. 4, p. 1003, Jan. 2014.

[97] A. Peñalba, L. Martínez-Álvarez, and A. Schiavio, "The active musical room: Fostering sensorimotor discoveries and musical creativity in toddlers," *J. Res. Music Educ.*, vol. 69, no. 2, pp. 128–151, Jul. 2021.

[98] V. P. Glăveanu, "What can be done with an egg? Creativity, material objects, and the theory of affordances," *J. Creative Behav.*, vol. 46, no. 3, pp. 192–208, Dec. 2012.

[99] E. W. Large and M. R. Jones, "The dynamics of attending: How people track time-varying events," *Psychol. Rev.*, vol. 106, no. 1, pp. 119–159, 1999.

[100] C. T. Kello, G. G. Anderson, J. G. Holden, and G. C. Van Orden, "The pervasiveness of 1/f scaling in speech reflects the metastable basis of cognition," *Cogn. Sci.*, vol. 32, no. 7, pp. 1217–1231, Oct. 2008.

[101] H. De Jaegher and E. Di Paolo, "Participatory sense-making," *Phenom. Cogn. Sci.*, vol. 6, no. 4, pp. 485–507, Dec. 2007.

[102] P.-J. Maes, V. Lorenzoni, B. Moens, J. Six, F. Bressan, I. Schepers, and M. Leman, "Embodied, participatory sense-making in digitally-augmented music practices: Theoretical principles and the artistic case 'SoundBike,'" *Crit. Arts*, vol. 32, no. 3, pp. 77–94, May 2018.

[103] P. E. Keller and B. H. Repp, "Multilevel coordination stability: Integrated goal representations in simultaneous intra-personal and inter-agent coordination," *Acta Psychologica*, vol. 128, no. 2, pp. 378–386, Jun. 2008.

[104] E. Pacherie, "How does it feel to act together?" *Phenomenology Cogn. Sci.*, vol. 13, no. 1, pp. 25–46, Mar. 2014.

[105] H. Liebermann-Jordanidis, G. Novembre, I. Koch, and P. E. Keller, "Simultaneous self-other integration and segregation support real-time interpersonal coordination in a musical joint action task," *Acta Psychologica*, vol. 218, Jul. 2021, Art. no. 103348.

[106] C. Hermans, "Participatory sense-making: Rhythm, repetition and affective resonance in dance," in *Proc. Lecture Conf. Worlding Brain*. Amsterdam, The Netherlands: University of Amsterdam, 2016.

[107] M. Bremmer and L. Nijs, "The role of the body in instrumental and vocal music pedagogy: A dynamical systems theory perspective on the music teacher's bodily engagement in teaching and learning," *Frontiers Educ.*, vol. 5, p. 79, Jun. 2020.

[108] K. M. Newell, "Change in motor learning: A coordination and control perspective," *Motriz Rio Claro*, vol. 9, no. 1, pp. 1–6, 2003.

[109] I. Renshaw and J.-Y. Chow, "A constraint-led approach to sport and physical education pedagogy," *Phys. Educ. Sport Pedagogy*, vol. 24, no. 2, pp. 103–116, Mar. 2019.

[110] K. S. Rosengren and G. S. Braswell, "Constraints and the development of children's drawing and writing skills," in *Development of Movement Coordination in Children: Applications in the Fields of Ergonomics, Health Sciences, and Sport*, G. J. P. Savelsbergh, K. Davids, J. v. d. Kamp, and S. Bennet, Eds., New York, NY, USA: Routledge, 2003, ch. 6, pp. 56–74.

[111] T. Hopper, "Constraints-led approach and emergent learning: Using complexity thinking to frame collectives in creative dance and inventing games as learning systems," *Open Sports Sci. J.*, vol. 5, no. 1, pp. 76–87, Sep. 2012.

[112] M. F. Levin and M. Demers, "Motor learning in neurological rehabilitation," *Disability Rehabil.*, vol. 43, no. 24, pp. 3445–3453, Nov. 2021.

[113] S. Fortuna and L. Nijs, "Children's representational strategies based on verbal versus bodily interactions with music: An intervention-based study," *Music Educ. Res.*, vol. 22, no. 1, pp. 107–127, Jan. 2020.

[114] S. Fortuna and L. Nijs, "The influence of discrete versus continuous movements on children's musical sense-making," *Brit. J. Music Educ.*, vol. 39, no. 2, pp. 183–202, Jul. 2022.

[115] L. Oppici, A. Dix, and S. Narciss, "When is knowledge of performance (KP) superior to knowledge of results (KR) in promoting motor skill learning? A systematic review," *Int. Rev. Sport Exercise Psychol.*, vol. 17, no. 1, pp. 182–207, Oct. 2021.

[116] A. Sheaff, *A Constraints-Led Approach to Swim Coaching*. New York, NY, USA: Taylor & Francis, Jul. 2023.

[117] C. J. Larche and M. J. Dixon, "The relationship between the skill-challenge balance, game expertise, flow and the urge to keep playing complex mobile games," *J. Behav. Addictions*, vol. 9, no. 3, pp. 606–616, Oct. 2020.

[118] A. R. Addessi, "A device for children's instrumental creativity and learning: An overview of the MIROR platform," *Frontiers Psychol.*, vol. 11, Nov. 2020, Art. no. 516478.

[119] A. Khatchatourov, F. Pachet, and V. Rowe, "Action identity in style simulation systems: Do players consider machine-generated music as of their own style?" *Frontiers Psychol.*, vol. 7, p. 474, May 2016.

[120] U. Bronfenbrenner, *The Ecology Human Development: Experiments by Nature Design*. Cambridge, MA, USA: Harvard Univ. Press, Sep. 1979.

[121] R. A. Schmidt, "Principles of practice for the development of skilled actions: Implications for training and instruction in music," in *Art in Motion. Musical and Athletic Motor Learning and Performance*, A. Mornell, Ed., Frankfurt, Germany: Peter Langer, 2008, pp. 41–67.

[122] Y. K. Ro, A. Brem, and P. A. Rauschnabel, "Augmented reality smart glasses: Definition, concepts and impact on firm value creation," in *Augmented Reality and Virtual Reality*, T. Jung and M. Dieck, Eds., Cham, Switzerland: Springer, 2018, pp. 169–181.

[123] P. Enlöf, C. Romare, P. Jildenstål, M. Ringdal, and L. Skär, "Smart glasses for anesthesia care: Initial focus group interviews with specialized health care professionals," *J. PeriAnesthesia Nursing*, vol. 36, no. 1, pp. 47–53, Feb. 2021.

[124] H. Shin and G. Gweon, "Supporting preschoolers' transitions from screen time to screen-free time using augmented reality and encouraging offline leisure activity," *Comput. Hum. Behav.*, vol. 105, Apr. 2020, Art. no. 106212.

[125] T. K. Oswald, A. R. Rumbold, S. G. E. Kedzior, and V. M. Moore, "Psychological impacts of 'screen time' and 'green time' for children and adolescents: A systematic scoping review," *PLoS ONE*, vol. 15, no. 9, Sep. 2020, Art. no. e0237725.

[126] R. K. Sawyer, "Learning music from collaboration," *Int. J. Educ. Res.*, vol. 47, no. 1, pp. 50–59, Jan. 2008.

[127] C. Pidel and P. Ackermann, "Collaboration in virtual and augmented reality: A systematic overview," in *Proc. 7th Int. Conf. Augmented Reality, Virtual Reality, Comput. Graphics*. Lecce, Italy: Springer, Sep. 2020, pp. 141–156.

[128] T. A. Syed, M. S. Siddiqui, H. B. Abdullah, S. Jan, A. Namoun, A. Alzahrani, A. Nadeem, and A. B. Alkhodre, "In-depth review of augmented reality: Tracking technologies, development tools, AR displays, collaborative AR, and security concerns," *Sensors*, vol. 23, no. 1, p. 146, Dec. 2022.

[129] Z. Dong, J. Zhang, X. Bai, A. Clark, R. W. Lindeman, W. He, and T. Piumsomboon, "Touch-move-release: Studies of surface and motion gestures for mobile augmented reality," *Frontiers Virtual Reality*, vol. 3, Aug. 2022, Art. no. 927258.

[130] M. Kim, S. H. Choi, K.-B. Park, and J. Y. Lee, "User interactions for augmented reality smart glasses: A comparative evaluation of visual contexts and interaction gestures," *Appl. Sci.*, vol. 9, no. 15, p. 3171, Aug. 2019.

[131] S. Zhao, H. Zhang, C. S. Mishra, S. Bhuyan, Z. Ying, M. T. Kandemir, A. Sivasubramaniam, and C. Das, "HoloAR: On-the-fly optimization of 3D holographic processing for augmented reality," in *Proc. 54th Annu. IEEE/ACM Int. Symp. Microarchitecture*. New York, NY, USA: ACM, Oct. 2021, pp. 494–506.

[132] O. Danielsson, M. Holm, and A. Syberfeldt, "Augmented reality smart glasses in industrial assembly: Current status and future challenges," *J. Ind. Inf. Integr.*, vol. 20, Dec. 2020, Art. no. 100175.

[133] L. Lauer, K. Altmeyer, S. Malone, M. Barz, R. Brünken, D. Sonntag, and M. Peschel, "Investigating the usability of a head-mounted display augmented reality device in elementary school children," *Sensors*, vol. 21, no. 19, p. 6623, Oct. 2021.

[134] B. Munsinger, G. White, and J. Quarles, "The usability of the Microsoft HoloLens for an augmented reality game to teach elementary school children," in *Proc. 11th Int. Conf. Virtual Worlds Games Serious Appl. (VS-Games)*. Vienna, Austria: IEEE, Sep. 2019, pp. 1–4.
[135] R. Schmidt and G. Wulf, *Continuous Concurrent Feedback Degrades Skill Learning: Implications for Training and Simulation*. New York, NY, USA: Routledge, Mar. 2017, pp. 375–391.
[136] C. J. Winstein and R. A. Schmidt, "Reduced frequency of knowledge of results enhances motor skill learning," *J. Experim. Psychol., Learn., Memory, Cognition*, vol. 16, no. 4, pp. 677–691, 1990.
[137] M. Sadakata, D. Hoppe, A. Brandmeyer, R. Timmers, and P. Desain, "Real-time visual feedback for learning to perform short rhythms with expressive variations in timing and loudness," *J. New Music Res.*, vol. 37, no. 3, pp. 207–220, Sep. 2008.
[138] W. Thorpe, J. Callaghan, and J. van Doorn, "Visual feedback of acoustic voice features in singing training," in *Proc. 9th Austral. Speech Sci. Technol. Conf.*, Hong Kong, Jul. 2002, pp. 3–5.
[139] F. Paas, A. Renkl, and J. Sweller, "Cognitive load theory and instructional design: Recent developments," *Educ. Psychologist*, vol. 38, no. 1, pp. 1–4, Jan. 2003.
[140] J. S. Bruner, *The Process of Education*. Harvard, MA, USA: Harvard Univ. Press, 2009.
[141] J. S. Bruner, *Toward a Theory of Instruction*. Harvard, MA, USA: Harvard Univ. Press, 1974, Paper CNC64.
[142] A. Elby, "What students' learning of representations tells us about constructivism," *J. Math. Behav.*, vol. 19, no. 4, pp. 481–502, 2000.
[143] J. Buchner, K. Buntins, and M. Kerres, "The impact of augmented reality on cognitive load and performance: A systematic review," *J. Comput. Assist. Learn.*, vol. 38, no. 1, pp. 285–303, Feb. 2022.
[144] C. Fueger, L. E. Sergio, S. Heuer, L. Petrovska, and W. E. Huddleston, "Remote concussion history does not affect visually-guided reaching in young adult females," *Concussion*, vol. 4, no. 3, p. CNC64, Nov. 2019.
[145] S. Vercruysse, M. Gilat, J. M. Shine, E. Heremans, S. Lewis, and A. Nieuwboer, "Freezing beyond gait in Parkinson's disease: A review of current neurobehavioral evidence," *Neurosci. Biobehavioral Rev.*, vol. 43, pp. 213–227, Jun. 2014.
[146] J. E. Reiser, E. Wascher, and S. Arnau, "Recording mobile EEG in an outdoor environment reveals cognitive-motor interference dependent on movement complexity," *Sci. Rep.*, vol. 9, no. 1, p. 13086, Sep. 2019.
[147] C. Hasty, *The Image of Thought and Ideas of Music*. New York, NY, USA: Routledge, Oct. 2016, pp. 1–22.
[148] J. Garzón, "An overview of twenty-five years of augmented reality in education. multimodal technologies and interaction," *Multimodal Technol. Interact.*, vol. 5, no. 7, p. 37, Jul. 2021.
[149] J. M. Krüger and D. Bodemer, "Application and investigation of multimedia design principles in augmented reality learning environments," *Information*, vol. 13, no. 2, p. 74, Feb. 2022.
[150] W. I. Schöllhorn, "Applications of systems dynamic principles to technique and strength training," *Acta Acad. Olymp. Eston.*, vol. 8, pp. 67–85, Jan. 2000.

LUC NIJS received the M.A. degree in music performance (clarinet) from the Royal Conservatoire of Brussels, Belgium, in 1996, the M.A. degree in philosophy, in 2008, and the Ph.D. degree in systematic musicology from Ghent University, Belgium, in 2012. He is currently a Professor of early childhood music education and the Head of the Bachelor in Music Education with the University of Luxembourg, Luxembourg. He has published articles on music educational technology in journals, such as *Computers and Education* and *Interacting With Computers*. His current research interests include embodied music pedagogy, music education technology, and creativity.

BAHAREH BEHZADAVAL received the M.A. degree in educational technology from Allameh Tabataba'i University, Tehran, Iran, in 2021. She started her research at Allameh Tabataba'i University, in 2018. She is currently a Flutist, a Singer, and a General Music Teacher in Iran. She has focused on educational technology in music education. As a Music Teacher and an Educational Technologist, she produced a series of instructional and musical videos for Persian language children under the name of Tutak Kids Music.