(Article begins on next page)

Extended Reality Applications in Music Education: A Competency-Based Review Across Instructional Domains

Pietro Buccellato, Matteo Sacchetto and Cristina Rottondi
Department of Electronics and Telecommunications Politecnico di Torino Turin, Italy
pietro.buccellato@polito.it; matteo.sacchetto@polito.it cristina.rottondi@polito.it

***Abstract*—Extended Reality (XR) technologies, including augmented reality (AR), virtual reality (VR), and mixed reality (MR), have been increasingly explored in music education to support immersive, interactive, and embodied experiences. As research activity in this area grows, the resulting literature remains fragmented across application types, learning contexts, and target populations, making it difficult to obtain a coherent overview of current research trends. This paper presents a review of XR applications in music education. A total of 303 peer-reviewed records were identified through Scopus and Web of Science and processed according to the PRISMA guidelines, resulting in 25 studies included in the review. The selected studies were analyzed and categorized according to three main instructional domains (namely instrumental technique & performance skills, music theory & aural skills, and ensemble & conducting skills) using a competency-based perspective grounded in established structures of music education curricula. In addition, early-childhood & primary education and accessibility are examined as transversal dimensions cutting across application domains. The review identifies domain-specific research patterns, cross-category design approaches, and structural gaps in the current XR music-education landscape. Findings indicate that XR technologies are primarily explored as complementary tools supporting embodied interaction, with more mature research trajectories in instrumental performance and music theory, more limited development in ensemble and conducting skills applications, and transversal dimensions that remain underdeveloped.**

***Index Terms*—Extended reality, music education, immersive learning, music technology**

I. Introduction

In recent years, Extended Reality (XR) technologies, including Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), have been applied across a wide range of disciplines, from science and engineering to healthcare and arts education, owing to their ability to support immersive spatial interaction, multimodal feedback, and experiential engagement [1], [2].

Music education represents a particularly promising domain for XR-based learning. Musical skills often rely on embodied practice, perceptual acuity, and tacit knowledge acquired through demonstration, imitation, and situated rehearsal. Traditional music instruction frequently depends on in-person interaction to convey fine-grained motor gestures, temporal coordination, and expressive nuance, elements that are difficult to represent through static notation or verbal explanation alone. XR technologies offer new possibilities for visualizing, spatializing, and physically enacting musical information, potentially complementing established pedagogical approaches [3], [4].

As a result, a growing body of research has investigated XR applications for a variety of music-learning tasks, including instrumental practice, ear training, music theory instruction, ensemble rehearsal, and conducting. These systems employ diverse interaction paradigms, ranging from visual overlays on physical instruments to fully immersive virtual performance environments, and target learners at different stages and with different educational needs. A closer examination of the survey literature reveals recurring patterns in how XR technologies have been approached in relation to music. XR has been widely discussed in musical contexts such as performance and composition, but mostly without an explicit focus on education [5]. Within educational settings, several surveys concentrate on specific technological components, such as VR or AR, rather than on the XR spectrum as a whole, adopting systematic mapping review and bibliometric analysis approaches that emphasize publication metrics and thematic clusters rather than educational applications [6]. Additionally, XR in music education has been reviewed in relation to individual instruments [7].

To complement these perspectives, this paper presents a review of XR applications in music education, focusing on how XR systems support specific instructional goals and musical competencies. Within this context, the contributions of this paper are threefold:

- it provides a PRISMA-informed systematic mapping of XR research in music education using transparent selection criteria and traceable study selection reporting;
- it proposes an analytically derived, domain-based taxonomy informed by established structures of music education curricula [8], which supports structured comparison across XR studies beyond purely technology-centered categorizations;
- it identifies cross-domain patterns, limitations, and gaps that inform future research and design efforts in XR-based music education.

In line with this competency-based perspective, the paper analyzes peer-reviewed XR studies and organizes them into three primary instructional domains: *instrumental technique & performance skills*, *music theory & aural skills*, and *ensemble*

TABLE I: Instructional domains and reference musical competencies used for classification

Instructional Domain	Reference Knowledge and Skills
Instrumental Technique & Performance Skills	Musical instrument performance Posture, gesture, and motor coordination Vocal technique and voice training Real-time performance feedback Preparation for individual musical performance
Music Theory & Aural Skills	Ear training (e.g., pitch, intervals) Rhythmic literacy Harmony and basic voice-leading Music notation and score reading Scales, modes, and pattern recognition Musical structure and form
Ensemble & Conducting Skills	Collaborative ensemble performance Choral and orchestral conducting Coordination, cueing, and timing Collective musical expression

& conducting skills. In addition, the analysis considers *early-childhood & primary education* and *accessibility* as transversal dimensions that cut across instructional domains. These contributions are formalized through a set of Research Questions (RQs) that structure the selection, classification, and analysis of the included studies. Specifically, the review addresses the following RQs:

1) **RQ1**. *How has XR been applied in music education over the last decade, and which instructional domains and educational dimensions can be identified from the literature?*
2) **RQ2**. *What types of XR technologies, learner characteristics, educational objectives, and evaluation approaches are reported in XR-based music education studies?*
3) **RQ3**. *What recurring patterns, limitations, and research gaps emerge from existing XR applications in music education?*

The remainder of this paper is organized as follows. Section II describes the methodology adopted for literature selection and classification. Section III presents the results of the review, organized according to the instructional domains and transversal dimensions identified through the analysis. Section IV discusses the findings through the research questions, highlighting cross-domain patterns, methodological differences, and emerging limitations. Section V outlines future research directions. Section VI concludes the paper.

II. Methods

This section outlines the literature search procedure and the classification framework adopted to analyze XR-based studies in music education.

A. Data Collection

This study adopts a review design supported by a systematic literature search, conducted in accordance with the PRISMA guidelines [9], to identify XR-based applications in music education. The literature search was conducted using two major academic databases, Scopus[1] and Web of Science[2], which were queried using boolean search strings combining XR-related and music-related terms, as detailed in the PRISMA flow diagram (Fig. 1). Searches were performed within the title, abstract, and keyword fields. The review was restricted to publications from the last ten years (2016-2025), reflecting recent developments in XR technologies and contemporary educational practices. All retrieved records were handled using a reference manager, where duplicate entries were removed prior to screening. Titles and abstracts were first screened to exclude studies not aligned with the scope of this work, and limited to English-language. Subsequently, full-text articles were assessed for eligibility, excluding non-empirical studies, design-focused contributions, and works that were primarily theoretical, mathematical, or simulation-based.

B. Classification Framework

Following the selection process, the included studies were analyzed through a qualitative classification framework grounded in instructional domains of music education. Rather than organizing XR systems according to technological characteristics (e.g., VR vs. AR) or target age groups, the analysis adopted a competency-based perspective based on established structures of music education curricula. Accordingly, the included studies were grouped into three primary instructional domains, reflecting what learners are expected to acquire through interaction with the system: *instrumental technique & performance skills*, *music theory & aural skills*, and *ensemble & conducting skills*. These domains align with broad structural areas of musical practice and research that are recurrently distinguished in comprehensive disciplinary overviews [8]. The reference musical competencies underlying each instructional domain are reported in Table I.

In addition to the primary instructional domains, two transversal dimensions were identified to capture learner-specific considerations that recur across the literature: *early-childhood & primary education* and *accessibility* for learners with sensory, motor, or cognitive impairments. These dimensions were treated as cross-cutting attributes that may characterize XR systems within any of the three primary domains, rather than as standalone categories. This distinction allowed the analysis to differentiate who the systems are designed for without conflating learner characteristics with the musical competencies being taught. *Early-childhood & primary-education* adaptations were identified in systems that explicitly targeted young learners (aged 4-11 years) through age-appropriate interaction design, simplified symbolic representations, or developmentally calibrated feedback. *Accessibility*-related adaptations were identified in systems designed to accommodate sensory impairments (e.g., hearing or visual impairments), motor limitations, or other special educational needs through multimodal feedback, alternative interaction paradigms, or assistive XR configurations.

[1]https://www.scopus.com/
[2]https://www.webofscience.com/

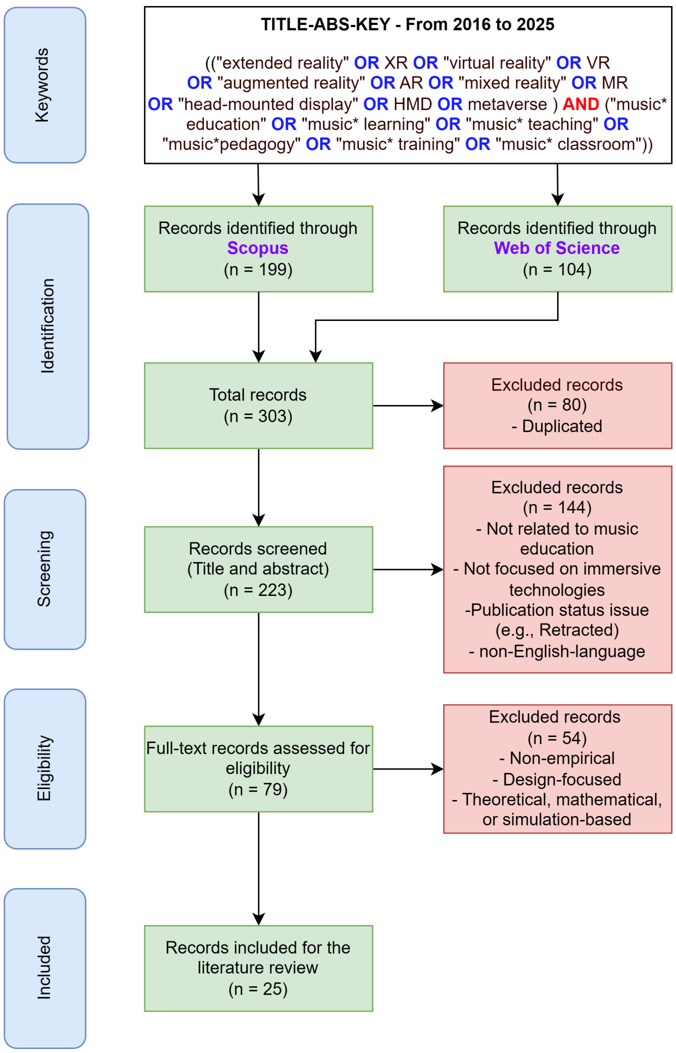

Fig. 1: PRISMA flow diagram illustrating the study selection process, from keyword-based identification of records to the final inclusion of papers.

III. Results

The results of the review are summarized in Fig. 1, which reports the number of records identified, screened, excluded, and included at each stage of the study selection process. Fig. 2 provides an overview of the distribution of the final selection of studies across instructional domains and transversal dimensions.

The following subsections examine studies within the identified instructional domains. Within each domain, transversal dimensions are presented where applicable.

A. XR for Instrumental Technique & Performance Skills

Studies within this application domain are organized according to the primary instrument or performance modality addressed. Key attributes are summarized in Table II.

An AR piano-instruction system is presented in [10], where virtual key overlays are superimposed onto a physical keyboard through an optical see-through headset. By reading MIDI files and projecting corresponding notes directly onto

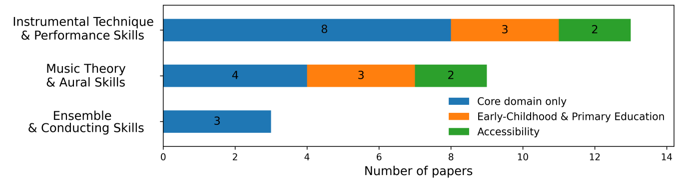

Fig. 2: Overview of XR studies in music education by instructional domain and transversal dimension.

the instrument, the system provides real-time spatial guidance for note identification, scale execution, and score–keyboard matching without evaluating the learner's input. A related system is described in [11], which integrates AR with mechanical actuation through an AR+IoT architecture. In this setup, an instructor wearing a head-mounted display (HMD) can remotely depress keys on a student's piano via servo-driven hardware, enabling real-time demonstration of fingering and technique. The system combines holographic visualization with physically executed keystrokes to support gesture modeling during remote instruction. Targeted motor challenges are addressed in the AR Piano system described in [12], which presents falling-note cues, labeled keys, and fingering guidance to support the development of left-hand technique.

MR is applied to synchronous remote piano education in [13], where bi-directional visualization of teacher and student hands is achieved through virtual–physical keyboard alignment. The system supports real-time correction during lessons and includes an integrated "Piano Theory Hub" providing feedback on scales, chords, and intervals.

XR systems have also been explored for percussive training. An early system investigated virtual co-embodiment as a pedagogical strategy for drumming, allowing learners to perform one component of a polyrhythm while an autonomous agent executed the complementary part, with the aim of reducing bimanual coordination demands during practice [14]. Building on this line of work, a subsequent MR drumming tutor integrated first-person visual perspectives, haptic cues, and virtual demonstrations to support the learning of rudimental and polyrhythmic patterns [15].

In addition to percussive practice, XR-based approaches have been investigated for vocal performance training. A VR-based vocal training system presented in [16] investigates the effectiveness of immersive virtual reality for vocal music education through the design and experimental evaluation of an integrated training environment combining anatomical visualization, motion capture, real-time acoustic feedback, and simulated performance spaces. Complementarily, a qualitative 360° VR case study [17] explores how immersive visual metaphors influence vocal technique during practice, examining whether VR-supported imagery can facilitate embodied vocal control, particularly in breathing regulation, resonance, vibrato production, and tonal clarity.

1) Early-Childhood & Primary Education: Some XR systems addressing instrumental technique are designed for early-childhood learners, adapting motor tasks and interaction

TABLE II: XR-based music education – instrumental technique & performance skills studies.

Ref.	Technology (XR Device)	Learner Entry Level	Evaluation Protocol	Key Findings
[10]	AR (HoloLens, Vuforia).	Early-stage piano students.	No user study reported.	Anticipatory visualisation supports muscle memory in novice piano learning.
[11]	AR (Magic Leap) + IoT actuators.	University instructors and students.	Evaluation protocol proposed (no results).	Supports remote demonstration of piano technique.
[12]	AR (AR Piano application).	Intermediate and advanced conservatory students.	Pre/Post study (n = 297) using Piano Proficiency Test-2 (PPT-2).	Improves left-hand technique and reduces cognitive load.
[13]	MR (Meta Quest 2/Pro, passthrough).	Beginner to intermediate pianists.	Two qualitative studies (n = 15; n = 10).	Improves note accuracy and perceived co-presence.
[14]	VR (Meta Quest 2).	Musically novice users.	Pilot study (n = 15) using MIDI jitter analysis.	Co-embodiment simplifies early rhythmic learning.
[15]	MR (Meta Quest 2) + electronic pads.	Non-musicians and novices.	Controlled study (n = 20) comparing MR vs video.	MR training significantly improves timing accuracy.
[16]	VR (HTC Vive Pro 2) + motion sensors.	Mixed-experience vocal learners.	12-week experimental study (n = 150).	Improves pitch, rhythm, and timbre control.
[17]	VR (Samsung Gear VR, 360° video).	University vocal students.	Qualitative case study (n = 8).	Facilitates vocal reflex activation and comfort.
[18]	Spatial AR (projection + Disklavier).	Children (ages 7–13).	Preliminary observations (n = 8).	Promotes expressive playing and memorisation.
[19]	Mobile AR (Vuforia).	Primary school children.	Observational evaluation.	Enhances engagement and basic coordination.
[20]	Mobile AR (Android).	Preschool children (age $\sim$ 6).	Mixed-methods study (n = 60).	Improves learning outcomes and enjoyment.
[21]	XR (Meta Quest 3, passthrough).	VI musicians.	Need-finding + prototype evaluation (n = 10).	Enables simultaneous reading and playing.
[22]	VR (Quest 2 + Leap Motion).	Musicians with physical disabilities.	Five-week technology probe with co-design.	Passive haptics increase agency and stability.

modalities to developmental needs. Projection-based AR has been explored for early instrumental learning. Andantino [18] employs animated characters projected onto a physical piano keyboard and score display to guide children's expressive playing. The system supports timing, articulation, and musical expression through synchronized visual cues rather than symbolic notation alone.

Along similar lines, [19] targets primary-school beginners by using marker-based AR to support early flute practice: when note markers are recognized, a 3D flute is displayed and the corresponding fingerings are highlighted, coupling basic notation cues with action-oriented guidance during simple melodies.

At a more foundational level, the AR-based singing and solfège tutor presented in [20] focuses on early music learning prior to instrumental specialization, introducing basic pitch patterns and rhythmic sequences through animated characters, guided gestures, and tangible interaction within a playful learning environment.

2) Accessibility: Accessibility-oriented XR systems addressing instrumental technique and performance are less represented in the literature. XRMusic4VIP [21] targets Visually Impaired (VI) musicians through a hybrid XR workflow combining immersive VR and video-passthrough AR on a HMD. Immersive VR is used during a pre-practice phase to explore musical structure and navigate the score, while video-passthrough AR supports real-time performance by overlaying accessible visual representations of sheet music onto a physical keyboard.

Motor accessibility is addressed in [22], which focuses on the design and evaluation of virtual musical instruments in immersive VR rather than training on a specific acoustic instrument. Developed through participatory workshops with musicians with physical disabilities, the system explores gesture-sound mappings, spatial arrangement of controls, and co-located physical surfaces adapted to individual motor capabilities.

B. XR for Music Theory & Aural Skills

Studies in this domain are organized according to the targeted music-theoretical and aural competencies, progressing from basic pitch and rhythm perception tasks to interval recognition and harmonic reasoning. Key attributes are summarized in Table III.

A VR-based pitch-perception game is presented in [23], where traditional ear-training exercises are reframed as an action-based task. Learners identify tones produced by virtual entities and respond by selecting and triggering corresponding notes on a controller-embedded keyboard. The system combines immediate auditory and visual feedback, spatialized audio cues, and game mechanics to support pitch-class recognition in an engaging learning context.

A complementary approach is proposed in [24], which introduces a 3D "Music Playground" environment aimed at teaching fundamental rhythmic concepts. The system organizes learning activities into distinct interactive zones, including theory-focused areas, rhythm ear-training tasks, and game-like navigation challenges. Through narrative-driven exploration and physicalized interaction, symbolic rhythmic structures typically introduced in early theory curricula are embedded within a multimodal virtual environment.

Interval recognition, a core aural-skills competency, is examined in [26], where VR is used to present melodic intervals within a semicircular arrangement of virtual piano keys. By manipulating octave range and comparing mono and stereo-panned conditions, the study investigates how spatial audio presentation influences interval perception and testing performance.

At a more advanced theoretical level, functional harmony is addressed in HarmoSphereVR [25], which enables learners to construct a four-chord progression in C major by positioning color-coded voice blocks on a virtual keyboard. The system provides real-time notation rendering, range checking, and chord-level auralization, allowing users to explore relationships between harmonic rules, written notation, and sonic outcomes. Designed on the basis of interviews with harmony students, the system illustrates how XR environments can support harmonic reasoning through embodied interaction.

1) Early-Childhood & Primary Education: Some theory-oriented XR systems explicitly target young beginners, adapting symbolic content to the cognitive and perceptual needs of early-childhood learners. For example, [27] uses AR to teach introductory staff notation by allowing preschoolers to color musical notes and receive animated feedback synchronized to the melody, fostering early symbolic-literacy development through playful multimodal interaction. Similarly, ChordAR [28] enables children to build concrete LEGO-based chord representations and see them interpreted in AR, transforming abstract harmonic concepts into tangible learning experiences. The genre-learning environment in [29] also demonstrates that immersive conceptual exploration can be effective for primary-school learners, particularly when timbre, instrument-family roles, and spatial stage layouts are introduced through simplified interactive metaphors.

2) Accessibility: Accessibility-focused systems extend theoretical and aural-skills learning to users with sensory impairments. The SeeMusic platform [30], for instance, integrates AR and VR with auditory-rehabilitation principles to support hearing-impaired children in detection, discrimination, identification, and comprehension tasks through synchronized visual cues, animated syllables, and interactive rhythm-matching. MuCIcal Adventures [31] further demonstrates how VR can scaffold pitch, timbre, and spatial-audio discrimination for CI users by coupling spatialized sound with salient visual indicators and gamified repetition.

C. XR for Ensemble & Conducting Skills

Within this application domain, the reviewed studies follow a progression from leader-driven ensemble coordination and conducting practice toward more distributed, peer-to-peer collaborative performance models. Key attributes are summarized in Table IV.

An AR system supporting orchestral sectional rehearsal is presented in [32], where violinists practice unfamiliar excerpts by imitating the bowing gestures of their section leader presented in AR. Using HMD AR overlays, participants interact with either a 2D video representation or a 3D avatar displaying bowing patterns related to articulation, phrasing, and dynamics. The system enables the visualization of ensemble cues typically conveyed during in-person rehearsals through gesture-based guidance. VR has also been applied to conducting education: the system described in [33] employs a tracked baton with passive haptics, allowing novice conductors to rehearse in a virtual concert hall populated with animated instrumental sections. The environment includes a virtual music stand and real-time tempo estimation to support practice of timing, cueing, and expressive shaping within a simulated ensemble context. Beyond leader-centered ensemble structures, XR has been explored for peer-to-peer collaborative performance. In [34], a HMD AR interface visualizes a collaborator's hand positions, eye gaze, and digital-instrument interactions in real time. Musicians performing with electronic keyboard controllers can view annotated representations of their partner's actions, such as active keys and parameter adjustments, supporting coordination and mutual awareness during collaborative music-making.

1) Early-Childhood & Primary Education: No XR studies were identified that explicitly address early-childhood learning within ensemble or conducting contexts. The ensemble-oriented XR literature reviewed here focuses on adolescent or adult learners and does not include applications designed for preschool or early primary education.

2) Accessibility: Similarly, no XR studies were found that incorporate accessibility considerations within ensemble- or conducting-oriented XR applications. None of the reviewed systems address ensemble-based music learning or conducting training for users with sensory, motor, or cognitive impairments.

IV. Discussion

This section examines the results by addressing the three RQs introduced in Section 1.

A. RQ1 – XR applications and instructional domains

In relation to RQ1, the distribution of XR applications across instructional domains does not simply reflect differences in research volume, but points to a deeper bias in how XR is currently conceptualized within music education. The predominance of instrumental applications compared to theory and especially ensemble contexts suggests that XR has been primarily adopted in scenarios where learning can be framed as an individual interaction between user and system.

This becomes particularly evident when comparing how XR operates across domains. In instrumental contexts, systems such as AR piano overlays [10], MR-based remote instruction [13], or multimodal drumming tutors [15] establish a tight coupling between visual guidance, motor execution, and immediate feedback. This alignment represents a clear advantage, since it allows XR to intervene directly in the

TABLE III: XR-based music education – music theory & aural skills studies.

Ref.	Technology (XR Device)	Learner Entry Level	Evaluation Protocol	Key Findings
[23]	VR (HTC Vive, SteamVR).	Students requiring long-term pitch training.	No study yet; expert interviews and field tests planned.	Shooting-based mechanic increases enjoyment in pitch training.
[24]	3D virtual world (desktop: keyboard/mouse).	Primary and secondary students.	Five-level training structure; no formal evaluation protocol; learning-effect evaluation under refinement.	"Music Playground" supports rhythm learning and imagination.
[25]	VR (Meta Quest 2; HarmosphereVR).	Secondary students (prior theory required).	Mixed-methods; interviews (n = 6).	Real-time note visualisation supports harmonic construction.
[26]	VR (HTC Vive) with stereo-panned audio cues.	Musicians and non-musicians.	User study (n = 27): System Usability Scale (SUS) and Game Experience Questionnaire (GEQ).	Spatial audio helps, but users cannot rely on it without training.
[27]	Mobile AR (printed staff verification).	Children (avg. age 6.2).	Pilot test with (n = 6).	Reduces initial barriers; supports spontaneous self-correction.
[28]	Mobile AR (mirror-based) + LEGO blocks.	Preschoolers (ages 4–7).	Observation + semi-structured interviews (n = 12).	Tangible AR enables imitation of complex chords (C, G, E).
[29]	Smartphone VR (Google Cardboard/GearVR; VR4EDU).	Primary students (age 10), incl. SEN.	Between-subject with control; one-way ANalysis Of VAriance (ANOVA).	Active VR navigation improves genre learning vs passive listening.
[30]	Mobile XR (AR textbooks + VR games).	Children (age 3+) with mild-to-moderate hearing loss.	Word Intelligibility by Picture Identification (WIPI) test (n = 23)	WIPI scores improve; accelerates auditory recovery/speech.
[31]	VR (Oculus Quest).	Adult Cochlear Implant (CI) users (6 months–26 years experience).	Iterative design (3 iterations) + one-week diary; SUS used.	Improves localisation habits; difficulty adaptation remains critical.

TABLE IV: XR-based music education – ensemble & conducting skills studies.

Ref.	Technology (XR Device)	Learner Entry Level	Evaluation Protocol	Key Findings
[32]	AR (HoloLens 2) + MoCap.	Amateur violinists ($\geq$12 yrs exp.).	Within-subject longitudinal study; kinematic metrics + presence questionnaires.	3D avatars increase presence vs 2D; imitation effects are subtle.
[33]	VR (HTC VIVE) + passive haptics.	Novice, university-aged participants.	A/B testing with Chi-square analysis; simple Beats Per Minute (BPM) estimation.	Passive haptics increase realism; visuals alter perceived physical properties.
[34]	AR (HoloLens 2) with body and MIDI visualisation.	Experienced electronic musicians.	2×2 within-groups study; SUS, NASA-Task Load Index (TLX), Situational Awareness Rating Questionnaire (SART) + interviews.	Improves situational awareness and mutual engagement.

perceptual–motor loop that underpins skill acquisition. At the same time, it introduces a limitation. Many of these systems implicitly assume that musical learning can be decomposed into discrete and correctable actions, which tends to privilege accuracy and coordination over expressive or interpretative dimensions. For example, anticipatory cues or falling-note visualizations [10], [12] reduce cognitive load and support execution, but they also risk encouraging procedural reproduction rather than deeper musical understanding.

A different logic emerges in music theory and aural skills applications. Systems such as VR pitch-training environments [23] or exploratory platforms like "Music Playground" [24] do not aim to stabilize motor performance, but rather to transform abstract musical concepts into perceptual and interactive experiences. This shift opens interesting possibilities, particularly in terms of engagement and intuitive learning. However, it also exposes a lack of pedagogical consistency. Many of these systems appear to function as engaging environments rather than structured learning tools. For instance, while HarmoSphereVR enables users to explore harmonic relationships through embodied interaction [25], it assumes prior theoretical knowledge, which highlights a tension between accessibility and conceptual depth.

The limitations become even more apparent in the ensemble and conducting domain. Existing XR systems, including AR-based gesture imitation [32], VR conducting environments with passive haptics [33], and AR interfaces for collaborative awareness [34], focus on specific components of ensemble interaction rather than on the full complexity of collective performance. These approaches can enhance perception of gestures, timing, or partner actions, but they do not fully address the distributed and adaptive nature of ensemble coordination. Unlike instrumental practice, ensemble performance cannot be reduced to a closed interaction loop between user and system, since it relies on shared intentionality and continuous mutual adaptation. Current XR solutions therefore capture fragments of ensemble practice without fully supporting its emergent dynamics.

The analysis of transversal dimensions further reinforces this interpretation. Early-childhood applications are mostly confined to instrumental and theory contexts, where interaction can be simplified and gamified [18], [19], [20]. Accessibility-oriented systems similarly focus on enabling individual interaction with musical material, for instance through accessible score visualization [21] or adaptive gesture-based interfaces [22]. Their absence in ensemble and conducting contexts suggests that XR research has not yet addressed how inclusive and collaborative musical experiences can be supported simultaneously. Overall, this points to a broader tendency in which XR is primarily used to support individualized learning processes, while its potential for collective and socially grounded music-making remains largely unexplored.

B. RQ2 – XR technologies, learners, and evaluation approaches

With respect to RQ2, the variability observed across technologies, learner populations, and evaluation approaches reflects a set of underlying trade-offs that are not yet fully resolved within the field.

From a technological perspective, one of the most evident tensions lies between immersion and accessibility. High-end XR systems used in instrumental and vocal training, such as VR environments with motion capture and real-time acoustic analysis [16] or MR platforms with precise hand tracking [13], enable detailed feedback and support fine-grained motor learning. These systems can significantly enhance performance accuracy and control. However, they also introduce practical limitations related to hardware cost, setup complexity, and user comfort. Issues such as tracking imprecision, headset weight, or fatigue can directly affect usability and limit long-term adoption.

In contrast, mobile and lightweight solutions commonly adopted in theory-oriented and early-childhood applications [27], [28], [29] are easier to deploy and more accessible to a broader audience. These systems are effective in promoting engagement and lowering entry barriers, especially for young learners. At the same time, they often provide a more limited form of interaction, with reduced immersion and less precise feedback. This creates a clear trade-off where accessibility is achieved at the expense of interaction richness, and where engagement does not always translate into deeper learning.

A similar tension can be observed in the choice of learner populations. Instrumental studies often focus on well-defined groups, such as conservatory students [12] or novices [15], which allows for controlled experimentation and clearer measurement of learning outcomes. While this increases internal validity, it also reduces generalizability, since the results are closely tied to specific user profiles. On the other hand, studies involving heterogeneous populations [16] or users with specific needs [21], [22] provide a more realistic representation of learning contexts, but introduce variability that complicates evaluation and comparison across studies.

These differences are reflected even more strongly in evaluation methodologies. Instrumental applications tend to adopt more rigorous approaches, including controlled experiments and quantitative performance metrics [12], [15], [16]. This allows for objective assessment of improvements in timing, pitch, or motor coordination. However, even in these cases, evaluation often remains confined to controlled settings and may not fully capture how skills transfer to real-world musical practice.

In theory and ensemble applications, evaluation is more frequently based on qualitative methods, usability studies, or small-scale pilot experiments [25], [26], [34]. These approaches provide valuable insights into user experience and system design, but they offer limited evidence regarding long-term learning outcomes. As a result, a divergence emerges in which technologically sophisticated systems are not always accompanied by equally robust pedagogical validation, while more accessible or engaging systems lack rigorous evaluation frameworks.

Transversal dimensions further complicate this picture. Early-childhood studies emphasize engagement, enjoyment, and emotional response [19], [20], often relying on observational or mixed-methods approaches. Accessibility-oriented research adopts iterative and participatory methodologies [21], [22], which are essential for inclusive design but typically involve small samples and limited statistical validation. These choices are justified by the characteristics of the target populations, yet they contribute to a broader fragmentation in evaluation practices.

Taken together, these observations suggest that XR music education is not only diverse, but also characterized by a lack of alignment between technological capabilities, learner diversity, and evaluation strategies. Different strands of research tend to optimize for specific objectives, such as immersion, accessibility, or engagement, without fully reconciling these dimensions within a coherent framework.

C. RQ3 – Patterns, limitations, and research gaps

In relation to RQ3, the patterns and limitations identified across the reviewed studies reveal a set of underlying tensions that shape the current development of XR in music education. One of the most evident patterns is the strong focus on individual skill acquisition, particularly in instrumental contexts where XR can effectively support motor learning through aligned visual and haptic feedback [10], [12], [15]. This has led to measurable improvements in performance accuracy and control. However, this emphasis also introduces a limitation, as learning is often framed in terms of quantifiable metrics such as timing precision or pitch accuracy. More complex dimensions of musical practice, including expressivity, interpretation, and interaction, remain less explored within XR environments. This limitation becomes particularly visible in the relative scarcity of ensemble and conducting applications. Existing systems address specific aspects of collaborative performance, such as gesture imitation [32], perceived realism [33], or awareness of partner actions [34], but do not fully capture the adaptive and relational nature of ensemble interaction. This suggests that current XR approaches are more suited to structured and controlled interaction scenarios than to open-ended, socially embedded musical practices.

Another critical issue concerns evaluation. While some studies demonstrate strong methodological rigor through controlled and longitudinal designs [12], [16], a significant portion of the literature remains at an exploratory stage [11], [17], [25]. Many systems are evaluated primarily in terms of usability, engagement, or perceived realism, rather than long-term learning outcomes. This creates a disconnect between technological innovation and pedagogical impact, where systems may appear effective without clear evidence of sustained educational benefits.

Technological constraints further contribute to these limitations. Issues such as limited field of view, tracking instability, and hardware discomfort directly affect the quality of interaction and may hinder prolonged use. In some cases, attempts to increase realism or provide richer feedback result in higher cognitive load, particularly in collaborative settings where multiple information layers are presented simultaneously [34]. This highlights a persistent trade-off between providing informative feedback and maintaining usability.

Transversal dimensions reveal additional gaps. Accessibility-oriented systems demonstrate the potential of XR to support inclusive music learning [21], [22], but remain limited in scope and are primarily focused on individual interaction. Their absence in ensemble contexts indicates that inclusive design has not yet been extended to collaborative musical environments. Similarly, early-childhood applications emphasize engagement and creativity [18], [19], [20], but do not address how collaborative skills can be developed at early stages of learning.

More broadly, the reviewed studies suggest that XR is currently positioned as a complementary tool that enhances specific aspects of music learning rather than as a technology that reshapes pedagogical structures. Many systems are designed around available hardware and interaction paradigms, without a strong connection to curriculum design or competency progression. As a result, XR applications often remain isolated interventions rather than integrated components of music education.

Overall, the field appears to be evolving within a set of unresolved tensions, including those between individual and collective learning, immersion and accessibility, engagement and rigor, and technological innovation and pedagogical grounding. Addressing these tensions will be essential for moving from experimental prototypes toward more coherent and impactful XR-based learning environments.

V. Future Directions

The present review highlights a research landscape that is both dynamic and unevenly consolidated. XR applications in music education demonstrate promising results across instrumental practice, theoretical training, and collaborative contexts; however, the field remains characterized by methodological fragmentation, technological constraints, and domain-specific imbalances. Future developments appear to require a shift from exploratory experimentation toward pedagogically grounded and adaptive validated systems. The future directions identified in this review are organized along three cross-cutting priorities that apply across the three instructional domains as well as the two transversal dimensions.

A. From Experimental Prototypes to Validated Educational Tools

Across instructional domains, a recurrent limitation concerns the heterogeneity and scale of evaluation protocols. While some contributions report controlled studies and longitudinal designs, a substantial portion of the literature remains grounded in pilot testing, usability assessment, qualitative case studies, or proof-of-concept validation. In several instances, particularly within theory-oriented and ensemble-based systems, presence, engagement, or perceived realism are assessed more frequently than long-term learning outcomes. Future research should prioritize methodological consolidation. Longitudinal studies are needed to determine whether improvements in timing accuracy, motor coordination, pitch control, or harmonic reasoning persist beyond initial exposure and are not primarily attributable to novelty effects. Comparative designs contrasting XR-based instruction with traditional pedagogical approaches and with alternative digital tools would help isolate the specific added value of immersive technologies.

Moreover, larger and more diverse participant cohorts are necessary to enhance generalisability. Several ensemble and conducting applications, for example, have been evaluated with small samples or within highly controlled laboratory settings. Extending validation to broader populations would strengthen claims regarding educational effectiveness.

In parallel, future studies may benefit from the development of standardized performance metrics tailored to XR contexts. Objective indicators, such as fine-grained temporal accuracy, kinematic analysis of gesture imitation, or structured ear-training benchmarks, could support cross-study comparability and facilitate cumulative evidence building. Moving from technology validation to pedagogical validation represents a central step in the maturation of XR-based music education.

B. From Static Interfaces to Adaptive and Inclusive Learning Environments

A second priority concerns the evolution from static XR interfaces toward adaptive and user-sensitive learning environments. The reviewed studies highlight recurring challenges related to cognitive load, hardware ergonomics, and interaction reliability. Limited FOV, headset weight, intrusive controllers, and tracking instability may constrain prolonged practice sessions or complex motor tasks. In collaborative contexts, dense visual overlays risk introducing distraction rather than facilitating coordination.

Future systems should incorporate adaptive strategies capable of dynamically regulating information density, feedback modality, and task difficulty. Visualization layers might adjust according to user workload, progressively simplifying or enriching guidance to prevent overload. Performance-based modulation, grounded in measurable indicators such as timing deviation, gesture stability, or pitch accuracy, could support personalized progression across skill levels.

Such adaptive principles are particularly relevant when considering the variability inherent in early learners and in users with sensory or motor impairments. Age-calibrated interaction paradigms, simplified symbolic representations, and emotionally supportive feedback may help sustain engagement among young students while mitigating frustration. In accessibility-oriented contexts, robust sensing and flexible interaction mappings remain critical, especially where atypical physiologies or perceptual constraints challenge standard tracking algorithms. At the hardware level, improvements in ergonomic comfort, sensing robustness, and deployability will remain central to sustained adoption. However, technological refinement should be guided by pedagogical relevance rather than by immersion metrics alone. Advancing XR design thus entails aligning interface sophistication with cognitive, developmental, and perceptual diversity across learner populations.

C. From Isolated Skill Training to Networked Music Practice

The majority of reviewed XR applications focus on individual skill acquisition, instrumental technique, pitch recognition, harmonic construction, or gesture rehearsal. Ensemble and conducting contexts remain comparatively underrepresented, and multi-user interaction is often limited to small-scale or highly structured scenarios. Yet musical practice is inherently relational, involving synchronization, shared intentionality, and collective expression.

Future research should therefore expand beyond isolated training modules toward networked grounded music-making environments. Multi-user XR platforms enabling synchronous rehearsal, remote collaboration, and shared performance spaces represent a key direction. Such systems would need to address real-time synchronization, latency management, and coherent representation of non-verbal cues, including gesture, gaze, and expressive dynamics.

In ensemble-oriented contexts, the integration of multimodal feedback, combining visual guidance, haptic cues, and motion capture, could support more nuanced motor imitation and coordination training. However, the educational value of increased perceptual realism must be empirically tested: higher presence does not automatically translate into improved collective performance. Long-term studies examining how immersive rehearsal environments influence ensemble timing, expressivity, and mutual awareness are therefore needed.

Furthermore, extending XR validation to less controlled and more authentic contexts, such as student orchestras, community ensembles, or remote collaborative settings, would strengthen the real-world relevance of these systems.

By connecting individual practice with collaborative performance and by embedding XR systems within real musical ecosystems, future developments may move the field from technologically advanced prototypes toward integrated educational infrastructures for music education.

VI. Conclusion

This paper presented a competency-based review of XR applications in music education, structuring the literature according to core instructional domains of formal music training. By addressing the proposed research questions, the study provided a structured and comparative perspective on how XR systems are currently designed, applied, and evaluated across domains.

In particular, the analysis highlights a clear predominance of XR applications targeting instrumental technique and performance skills, where systems are primarily designed to support perceptual–motor alignment and execution. Music theory and aural skills applications show a broader variety of approaches, often focusing on the transformation of abstract concepts into interactive and exploratory environments. In contrast, ensemble and conducting contexts remain comparatively underexplored, with existing systems addressing only partial aspects of collaborative musical interaction.

Early-childhood & primary education and accessibility emerge as transversal dimensions that remain fragmented across the literature, indicating the need for more integrated and inclusive design approaches. Overall, XR technologies show significant potential to enrich music learning; however, the field is still evolving from exploratory implementations toward more coherent, pedagogically grounded, and scalable educational frameworks.

Acknowledgment

The authors would like to thank the XR Lab of the LINKS Foundation for their valuable support and the enlightening discussions that contributed to this work. This publication is part of the project PNRR-NGEU which has received funding from the MUR- DM 118/2023.

Funding

This publication is part of the project PNRR-NGEU which has received funding from the MUR-DM 118/2023. We acknowledge the support of the MUSMET project funded by the EIC Pathfinder Open scheme of the European Commission (grant agreement no. 101184379). This manuscript reflects only the authors' views and opinions and neither the European Union nor the European Innovation Council can be considered responsible for them.

References

[1] O. Emma, "Extended reality in the digital age: A literature review on technology-driven learning environments," *Computers & Education: X Reality*, vol. 8, p. 100127, 2026.

[2] G. M. Serna-Mendiburu and C. R. Guerra-Tamez, "Shaping the future of creative education: The transformative power of VR in art and design learning," *Frontiers in Education*, vol. 9, p. 1388483, 2024.

[3] L. Nijs and B. Behzadaval, "Laying the foundation for augmented reality in music education," *IEEE Access*, vol. 12, pp. 100 628–100 645, 2024.

[4] S. Yu, Q. Liu, M. C. Johnson-Glenberg, M. Han, J. Ma, S. Ba, and L. Wu, "Promoting musical instrument learning in virtual reality environment: Effects of embodiment and visual cues," *Computers & Education*, vol. 198, p. 104764, 2023.

[5] L. Turchet, R. Hamilton, and A. Çamci, "Music in extended realities," *IEEE Access*, vol. 9, pp. 15 810–15 832, 2021.

[6] X. Zhang and W. A. J. W. Yahaya, "A systematic review of augmented reality in music education," in *Enhancing Learning Experiences With Digital Tools: AI, ChatGPT, and Virtual and Augmented Reality*. IGI Global Scientific Publishing, 2024, pp. 1–22.

[7] K. Wilson and P. E. Pfeiffer, "Feedback in augmented and virtual reality piano tutoring systems: a mini review," *Frontiers in Virtual Reality*, vol. 4, p. 1207397, 2023, mini Review. [Online]. Available: https://doi.org/10.3389/frvir.2023.1207397

[8] G. E. McPherson and G. F. Welch, Eds., *The Oxford Handbook of Music Education*. Oxford, UK: Oxford University Press, 2012.

[9] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and D. Moher, "The prisma 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, 2021.

[10] D. Hackl and C. Anthes, "Holokeys: An augmented reality application for learning the piano," in *Proceedings of the International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, 2017, pp. 140–144.

[11] A. Stanbury, I. Said, and H. J. Kang, "Holokeys: Interactive piano education using augmented reality and iot," in *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology (VRST '21)*, 2021, pp. 1–3.

[12] D. Lu, "The impact of augmented reality on motor skills in piano education: Improving the left-hand technique with ar piano," *International Journal of Human–Computer Interaction*, vol. 41, no. 13, pp. 8398–8408, 2025.

[13] V. Amm, K. Chandran, L. Engeln, and M. McGinity, "Mixed reality strategies for piano education," *Frontiers in Virtual Reality*, vol. 5, pp. 1–16, 2024.

[14] J. Pinkl and M. Cohen, "Vr drumming pedagogy: Action observation, virtual co-embodiment, and development of drumming "Halvatar"," *Electronics*, vol. 12, no. 17, p. 3708, 2023.

[15] J. Pinkl, J. Villegas, and M. Cohen, "Multimodal drumming education tool in mixed reality," *Multimodal Technologies and Interaction*, vol. 8, no. 70, pp. 1–16, 2024.

[16] L. Zong, "Evaluation on the effect of enhancing vocal music training experience with virtual reality technology," *International Journal of Web-Based Learning and Teaching Technologies*, vol. 20, no. 1, 2025.

[17] S. Doganyigit and O. F. Islim, "Virtual reality in vocal training: A case study," *Music Education Research*, vol. 23, no. 3, pp. 391–401, 2021.

[18] X. Xiao, P. Puentes, E. Ackermann, and H. Ishii, "Andantino: Teaching children piano with projected animated characters," in *Proceedings of the 2016 ACM Conference on Interaction Design and Children (IDC)*, 2016, pp. 379–390.

[19] H. Guclu, S. Kocer, and O. Dundar, "Application of augmented reality in music education," in *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)*, vol. 14, 2021, pp. 45–56, presented at IConTech 2021: International Conference on Technology.

[20] X. Wei, R. Qiu, and R. Liu, "Using augmented reality to facilitate music learning for preschool children," in *Proceedings of the 31st International Conference on Computers in Education (ICCE)*, 2023, pp. 338–342.

[21] J. Anken, F. Köhler, M. Brehmer, and P. Reipschläger, "Xrmusic4vip: Enabling simultaneous sheet music reading and playing for visually impaired musicians through extended reality," in *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, 2025.

[22] D. Mills, F. Schroeder, and J. D'Arcy, "Facilitating accessibility in virtual reality musical environments," in *Proceedings of the International Computer Music Conference (ICMC)*, 2022, DOI not available, pp. 116–123.

[23] J. Yang and C. Cheng, "A pitch perception training game in vr environment for enhancing music learning motivation," in *2020 IEEE International Conference on Consumer Electronics – Taiwan (ICCE-Taiwan)*, 2020, pp. 1–2.

[24] Y. Feng, S. Li, and Y. Wei, "Design of a game-based learning environment for basic rhythmic patterns in music," in *Third International Conference on Computer Science and Communication Technology (ICCSCT 2022)*, ser. Proc. of SPIE, vol. 12506, 2022, p. 125065E.

[25] W. Stachurska, A. Witoszek-Kubicka, and M. Igras-Cybulska, "Harmosphere vr: Enhancing harmonic learning in music schools through virtual reality," in *Proceedings of the 2024 International Conference*, 2024.

[26] C. Fletcher, V. Hulusic, and P. Amelidis, "Virtual reality ear training system: A study on spatialised audio in interval recognition," in *2019 IEEE International Conference (publication via IEEE Xplore)*, 2019, pp. 1–4.

[27] A. G. D. Correa, B. H. V. Lemos, M. Nascimento, and R. D. Lopes, "Ar musical app for children's musical education," in *2016 IEEE International Symposium on Consumer Electronics (ISCE)*, 2016, pp. 1–2.

[28] Y. Lu, X. Wang, J. Gong, and Y. Liang, "Chordar: An educational ar game design for children's music theory learning," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–9, 2022.

[29] E. Degli Innocenti, M. Geronazzo, D. Vescovi, R. Nordahl, S. Serafin, L. A. Ludovico, and F. Avanzini, "Mobile virtual reality for musical genre learning in primary education," *Computers & Education*, vol. 139, pp. 102–117, 2019.

[30] L. Li, X. Chen, Y. Qin, and H. Cai, "Seemusic: Xr-enhanced musical language training application for hearing-impaired children," in *Proceedings of the 14th International Conference on Educational and Information Technology (ICEIT)*, 2025, pp. 333–338.

[31] E. S. Hansen, S. M. K. Kristiansen, S. Serafin, and A. Adjorlu, "Virtual reality music training for ci users: Insights and design guidelines," in *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2025, pp. 653–658.

[32] A. Campo, A. Michałko, B. Van Kerrebroeck, and M. Leman, "Dataset for the assessment of presence and performance in an augmented reality environment for motor imitation learning: A case-study on violinists," *Data in Brief*, vol. 51, p. 109663, 2023.

[33] A. Barmpoutis, R. Faris, L. Garcia, L. Gruber, J. Li, F. Peralta, and M. Zhang, "Assessing the role of virtual reality with passive haptics in music conductor education: A pilot study," in *Proceedings of the International Conference on Human-Computer Interaction*, ser. LNCS, vol. 12190, 2020, pp. 1–11.

[34] Y. Wang, M. Xi, M. Adcock, and C. P. Martin, "Seeing the sound: Supporting musical collaboration with augmented reality," in *Proceedings of the 2025 ACM Creativity & Cognition Conference (C&C '25)*, 2025, pp. 1–14.