From Nodes to Notes: Conceptualisation of AI-Generated Music for Immersive Historical Reconstruction—MAKAIM



Guzden Varinlioglu, Jacopo de Berardinis, Nevin Sahin, and Ezgi Yurumez

Abstract The chapter explores the integration of AI-generated music to enhance immersive historical reconstructions. Informed by the music theory and modal systems of the period and drawing on **makam** traditions from later eras, it reconstructs the sound of a largely forgotten music heritage. AI-generated soundscapes, developed from historically grounded assumptions are integrated into photogrammetric models and interactive, game-based VR environments. The process involves curating a specialised dataset of **makam** music captions drawn from historical accounts and oral traditions, expanding it through large language model (LLM) augmentation under expert supervision, integrating contextual information from each caravanserai, generating audio samples using a fine-tuned text-to-music model, and applying acoustic simulations with measured impulse responses to reconstruct the historical soundscape. This study evaluates the impact of these reconstructions on user engagement, immersion, and perceived historical authenticity, demonstrating AI's potential to create multisensory, historically informed experiences within virtual heritage environments.

Keywords AI-generated music · Historical reconstructions · Makam music · Text-to-music model · Virtual reality

G. Varinlioglu (✉)
School of Architecture, University of Liverpool, Liverpool, UK
e-mail: guzden.varinlioglu@liverpool.ac.uk

J. de Berardinis
School of Computer Science and Informatics, University of Liverpool, Liverpool, UK
e-mail: jacodb@liverpool.ac.uk

N. Sahin
Department of Musicology, Hacettepe University Ankara State Conservatory, Ankara, Turkey
e-mail: nevin.sahin@hacettepe.edu.tr

E. Yurumez
Faculty of Arts and Design/Music, Yildiz Technical University, Istanbul, Turkey
e-mail: ezgi.yurumez@yildiz.edu.tr

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2026

A. Agkathidis et al. (eds.), *Architecture in the AI Era for Research, Practice, and Pedagogy*, SpringerBriefs in Architectural Design and Technology,
https://doi.org/10.1007/978-981-95-0760-3_9

1 Introduction

Integrating AI-generated music into historical reconstructions enhances immersive experiences by adding a sensory dimension beyond visual reconstructions, which have traditionally dominated heritage documentation. While photogrammetric models and game engines enable the virtual representation of architectural heritage, accurately selecting and recreating historical music remains a challenge. As an integral part of intangible heritage, sound requires careful reconstruction to ensure historical and cultural authenticity within VR environments.

This chapter explores the potential of AI-generated music to recreate the auditory dimensions of the Anatolian Middle Ages, specifically focussing on caravanserais—iconic architectural nodes along the Silk Roads. These structures, central to trade and cultural exchange during the Seljukid era (approx. 300 years), form the basis of a broader SILK project[1] dedicated to documenting, preserving, and presenting this architectural heritage along the Silk Roads. As part of the project, a large database of images and videos is collected and processed to create photogrammetric models of historic buildings. These models form the foundation of SILK VR, an immersive virtual experience that enables users to explore reconstructed architectural environments in detail. While the visual and spatial characteristics of these spaces have been studied extensively, their auditory heritage remains largely unknown. Limited resources exist on Seljukid music practices [1], making direct reconstruction or replay only possible through AI. This study explores AI-driven composition to generate music inspired by Eastern and Turkish compositions, characterised by microtonal scales, modal systems, and distinctive instrumentation that resonate with the Seljukid period.

The process begins with curating a specialised dataset of ***makam*** music captions based on historical accounts and oral traditions, expanded using a Large Language Model (LLM) with expert supervision. Contextual information specific to each caravanserai—such as architectural acoustics, spatial geometry, and historical function—is then incorporated to refine the input prompts. These enriched prompts are fed into a fine-tuned text-to-music generation model to produce audio snippets representing the reconstructed soundscape. Finally, acoustic simulation techniques, applying either measured or simulated impulse responses, are employed to recreate how the music would have sounded in each historical space [2]. By leveraging AI trained on surviving musical traditions and historical records, this research reconstructs plausible soundscapes, enhancing both user engagement and historical accuracy.

This interdisciplinary research, bridging historic preservation, architecture, and music, explores AI's role in reconstructing historical soundscapes by addressing key questions: Which generative AI systems best capture Seljuk auditory heritage? How

[1] The Sustainable Information Lifecycle of Kervansaray (SILK) project, directed by Assoc. Prof. Güzden Varinlioğlu, in the field of digital heritage, is funded by the U.S. Department of State, in partnership with Assoc. Prof. Takehiko Nagakura, of the Design Computation Lab, Department of Architecture at Massachusetts Institute of Technology (MIT). https://makingindigital.com/silk/.

much human intervention is needed to refine AI outputs? And does convolution enhance the authenticity of VR experiences? By integrating auditory elements into digital heritage, this study advances both methodology and historical understanding. It also demonstrates VR's potential in cultural heritage preservation.

2 Background and Literature

With digital tools transforming architectural preservation, photogrammetry has become a leading 3D documentation technique due to its affordability, accessibility, and detail capture [3]. High-resolution photographs of archaeological findings are processed using Structure-from-Motion (SfM) algorithms to generate accurate digital models [4]. This method ensures reliable artefact documentation. Neural Radiance Fields (NeRF) is an AI-based technique for synthesising highly realistic 3D scenes and objects from 2D image datasets. Unlike traditional photogrammetric methods, NeRF techniques are capable of reproducing subtle variations in lighting, shadows, and surface texture, making them particularly effective for visualising complex geometries and dynamic lighting conditions [5]. After generating the 3D photogrammetric model and completing the associated measurements, the high-polygon dataset is optimised for real-time rendering and interaction in a virtual reality setting.

Virtual reality is the presentation of a modelled, three-dimensional virtual environment to the user through headsets and similar tools. Thus, the user can have the feeling of existing in a virtual space [6]. Photogrammetry and virtual reality (VR) are increasingly utilised in the reconstruction and preservation of cultural heritage, providing immersive and interactive experiences while ensuring historical accuracy. Forte and Kurillo highlighted the use of VR in the virtual reconstruction of Çatalhöyük—Dig@IT [7]. Selimiye VR led to the creation of a virtual 3D model of the Selimiye Mosque, allowing users to explore the mosque interactively [8]. These applications showcase how the fusion of photogrammetry and VR offers both precise digital preservation and accessible, engaging cultural heritage experiences.

In recent years, AI-generated music has gained attention for its role in enriching virtual cultural heritage environments. Agostinelli et al. [9] explored how AI systems, trained with period-specific music sources, can generate music that reflects the historical context of heritage sites, adding an auditory dimension to VR environments. Başarır et al. [10] explore a Local Intelligence (LI) framework using a multimodal AI model to uncover how the tacit logic of folklore music can inform and enrich vernacular architecture. The project *Adakale Türküleri* uses AI and machine learning to rediscover and reconstruct traditional Turkish folk songs from Adakale, preserving its rich musical heritage [11].

Crucial for advancing research, comprehensive databases for Turkish ***makam*** music have been developed. Karaosmanoğlu [12] created SymbTr, a symbolic database of 1,700 pieces across 155 ***makams***, 100 ***usuls***, and 48 forms. Senturk and Chordia [13] compiled a Turkish folk music database, using it to demonstrate melodic prediction in the ***uzun hava*** form with Variable-Length Markov Models. Building

on these datasets, recent AI research has focussed on symbolic ***makam*** music generation. Parlak et al. [14] used LSTM networks to compose in Hicaz and Nihavent ***makams***, achieving 84% similarity to existing pieces. Aydıngün et al. [15] generated Turkish songs, including lyrics and melody, with recurrent neural networks. Kaşif and Sevgen [16] employed self-attention and LSTMs for stylistically consistent classical Turkish music generation [16]. However, these methods primarily rely on post-17th-century scores and symbolic data, limiting their ability to generate audio for earlier periods or unseen instruments, where characteristics must be inferred rather than directly observed.

2.1 Seljukid Architecture and Caravanserais

The history of Turkish architecture is deeply influenced by its diverse cultural heritage, spanning from the ancient Hittites to the Byzantine and Ottoman periods. Seljukid architecture in Anatolia, emerging between the eleventh and thirteenth centuries, represents a key moment in the evolution of Islamic architectural traditions. Drawing upon Byzantine, Persian, Central Asian, and Islamic sources, Seljuk architects developed distinctive building types—mosques, madrasas, mausoleums, and caravanserais—characterised by monumental scale, intricate ornamentation, and refined spatial organisation.

Caravanserais, known in Turkish as *han*, served as essential resting places for travellers during long journeys. Strategically located at intervals of approximately 30–40 km, roughly a day's travel, these structures formed a critical part of the trade infrastructure [17]. Efforts have been made to map the caravanserais, enabling spatial analyses of their locations and networks [18]. According to the SILK project, 322 Seljukid caravanserais have been documented along roads within the borders of modern-day Türkiye [19]. Varinlioglu and Balaban [20] applied AI to locate caravanserais using remote sensing and predictive models to identify and map these structures. Important works include Branning's research on Turkish hans (2017) [21] and a variety of scholarly contributions [22–25].

Researchers, e.g. Erdmann [26] and Yavuz [27], classify caravanserai plans into four main types: open plans, open courtyards, closed central spaces, and mixed plans. While their architectural, economic and logistical importance, the cultural dimensions of these structures—particularly their association with musical performances—remained largely overlooked. Music, as a medium of interaction and cultural integration, reveals the broader social significance of caravanserais beyond their utilitarian purposes.

2.2 Seljukid Music and Music Creation

Music was an integral part of Seljuk's social life and contributed to the multidimensional atmosphere of caravanserais. However, being an oral tradition, the music was not well documented in terms of music notation, making sheet music collections appear as late as the seventeenth century [28], and the treatises aimed instead at digging into the science behind music, in line with the Ancient Greek understanding of Quadrivium. Hence the theoretical descriptions for this project could be taken from the treatises of the period, but well-documented music examples needed to be taken from either later periods (Ottoman music) or neighbouring geographies (Byzantine music).

Seljukid music, similar to neighbouring music of Anatolia, the Middle East, and the Balkans, has a modal nature. ***Makam***, the rules determining the melodic structure, and ***usul***, the rhythmic patterns, are the two key and complex systems to understanding this music [29]. Different from modes, the melodic scales of peculiar pitches are not enough to understand ***makams***; the melodic progression (***seyir***) should also be understood [30]. Being an oral tradition, ***makam*** music necessitated the master-student relationship named ***meshk*** in transmitting music, creating a gap between theory and practice and a lack of notated Seljukid repertoire [31].

Music during the Seljukid era permeated various facets of life, from military and religious practices to secular and courtly entertainment, including public spaces such as caravanserais [32]. The variety of music performances brought together a variety of instruments. The courtly and ceremonial music utilised chordophones such as ***ud*** and ***kopuz***, aerophones such as ***ney***, as well as membranophones such as ***def*** [33], which, compared to other instruments of the era, have been performed to date with minimal changes. One of the most notable instruments associated with Seljuk music was the ***kopuz***, a stringed instrument of Central Asian origin that remained popular among the nomadic Turks and continued to be used during the Seljuk period. The ***ney***, a reed flute later adopted by the Mevlevi dervishes, also emerged during this era and became a staple in both religious and secular music [34]. Because human voice is difficult to generate in AI modelling, the abovementioned instruments were selected due to the ease of access.

2.3 Design of the Virtual Experience—SILK VR

The project methodology integrates traditional photogrammetry and Neural Radiance Fields (NeRF) to document and analyse architectural heritage. Photogrammetric models are generated in Agisoft Metashape, a widely used tool in archaeology for its precision and usability. Approximately 5000–20,000 high-resolution photographs, depending on the building' size, are processed to capture the three—dimensional spatial data of architectural elements. Agisoft Metashape was used to align the images, calculate camera positions, and generates a detailed point cloud.

The software produces high-polygon surfaces, which are subsequently simplified to between 1 and 4 million polygons using its built-in decimation tools. Textures derived directly from the source photographs, resulting in a high-resolution output. These photogrammetric models are then optimised to reduce polygon counts for real-time applications such as Unity while maintaining essential geometric and material details. Texture maps are refined during optimisation to ensure seamless visual integration. Finally, a comparative analysis is conducted to evaluate the performance of NeRF-based reconstruction methods against photogrammetry outputs from Metashape and related software, in order to identify the most effective solutions for the project's documentation and visualisation objectives.

The project integrates photogrammetric models of historical structures into the game engine environment to create detailed, accurate representations. In Unity, the models are prepared for real-time interaction and visualisation, allowing users to explore and analyse the sites in virtual or augmented reality settings. The models are enhanced with interactive features, such as navigation tools, and hotspots for accessing additional information, designed to accommodate archaeologists, educators, and general users.

Traditional instruments and historically informed melodies composed by Ezgi Yurumez are used alongside environmental sounds, such as footsteps, wind, and ambient noise of trade routes, to align with the historical context. These auditory elements are synchronised with the visual reconstructions to create a fully multisensory experience.

The workflow follows a structured sequence: photogrammetric models are reconstructed and optimised, imported into Unity, and augmented with interactive features and spatialised soundscapes (Fig. 1). The outcome is a dynamic, interactive platform that transforms static architectural documentation into a tool for education and cultural engagement.

3 MAKAIM Conceptualisation Methods

The absence of notated scores or audio recordings from the Seljukid period presents a challenge in reconstructing the sonic environments of architectural heritage such as caravanserais. This study introduces MAKAIM, a methodological framework that employs **Multimodal Generative AI** to generate historically informed ***makam*** soundscapes. By integrating historical resources on music and the acoustic properties of architectural spaces, MAKAIM aims to reconstruct/generate intangible musical heritage with cultural and contextual fidelity. The conceptual framework is detailed in the following section and illustrated in Fig. 2.

The conceptualisation of MAKAIM revolves around a four-stage process. First, a specialised dataset of ***makam*** music captions (Seljukid Makam Dataset-SMD) is curated, drawing from historical accounts and oral traditions, and is further expanded using a Large Language Model (LLM) with expert oversight. Second, contextual

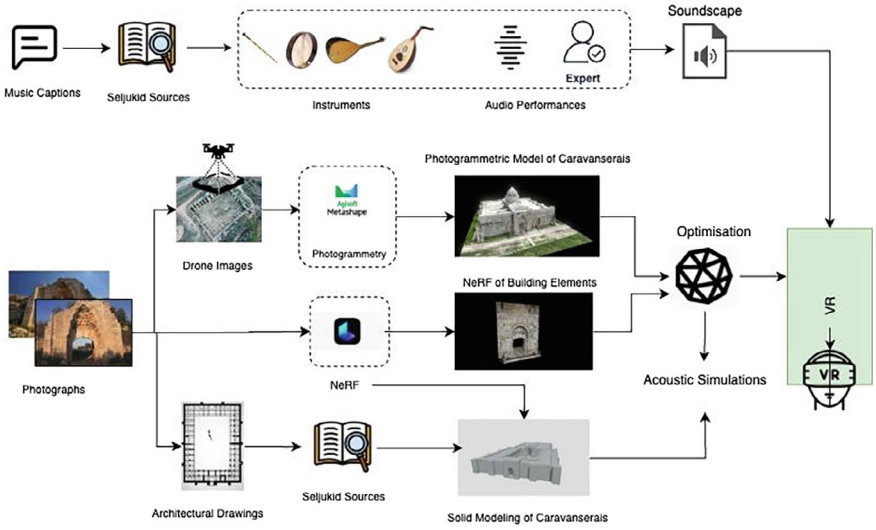

Fig. 1 Illustration of the SILK VR workflow

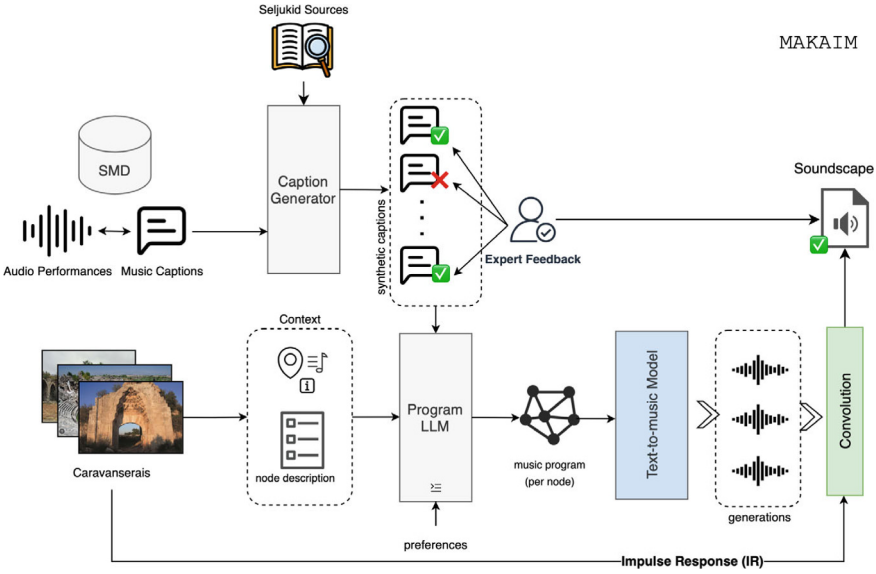

Fig. 2 Illustration of the MAKAIM workflow for a culturally aware generation

information specific to each caravanserai is integrated to augment the musical experience, considering both architectural features and historical usage. Third, these enriched prompts are fed into a state-of-the-art **text-to-music** generation model, selected and fine-tuned for this musical domain, to produce audio snippets representing the soundscape. Finally, these generated pieces undergo acoustic simulation, leveraging either measured or simulated impulse responses to recreate how the music would have sounded within the specific acoustic environment of each historical space.

3.1 The Seljukid Makam Dataset (SMD)

The generative approach for reconstructing Seljukid soundscapes relies on a curated dataset of descriptive music captions. The dataset is created by annotating a set of audio performances and improvisations by professional musicians speculating on the music of that period. This dataset aims to cover the core characteristics of Seljukid ***makam*** music and is created with the involvement of N=2 musicologists with expertise in this domain. The caption of *Ilahi* by N. Şahin, illustrated/played in the link provided, is shown in Fig. 3.

Each caption provides a textual description of a ***makam*** or a homogeneous musical segment (tempo, rhythm, instrumentation) from a meaningful, self-contained pattern. These prompts capture specific aspects of ***makam*** music, including function, style, melodic articulation, instruments, voices, tempo, and emotional character. We begin with a small expert-curated dataset of ***makam*** annotations and expand it using an LLM, integrating explicit knowledge from Seljukid sources and implicit knowledge from the model. This process generates new captions describing plausible ***makams*** or shorter segments aligned with historical musical accounts. Music experts then manually review these captions, identifying inaccuracies in melodic progressions, rhythmic patterns, and instrumentation, suggesting edits, and providing feedback for refinement.

Performance: Segâh Niyaz İlahisi
Caption: Attributed to Sultan Veled (d. 1312), religious music performed towards the end of the whirling ceremony of dervishes. The available recording is performed by ney (aerophone), kudüm (membranophone), kemençe (chordophone) and male voices. The melody starts in a 14/8 beat section (ca. 100 bpm), followed by a vocal improvisation accompanied by ney, subtly in tempo. It ends with an accelerando section in 6/8, starting around 150 bpm and ending around 220 bpm. The melodic progression of *makam* Segâh (starting around the 1st degree, emphasizing 3rd degree, sharpening 4th degree when resting on 5th degree, flattening the 1st degree when resting on 7th degree lower octave. Microtonal, the interval between 1st and 2nd degrees broader than a minor second, the interval between sharpened 7nd degree lower octave and 1st degree narrower than a minor second) is followed throughout the composed parts, sheet music for this piece is available at https://neyzen.com/nota_arsivi/01_ilahiler/082_segah/semi_ruhuna_cismimi_pervane_dusurdum.pdf.

Fig. 3 The caption of ***ilahi*** (religious music in Turkish Sufi culture) was retrieved from https://www.youtube.com/watch?v=8CZYNqE-DMg

3.2 Music Scene Contextual Augmentation

Currently, the VR experience allows users to navigate immersive 3D reconstructions of caravanserais to explore architectural details and functions, but a complete sensory experience, including sound, has yet to be developed (Fig. 1). This presents both a challenge and an opportunity to better contextualise the simulated music. The VR experience currently includes around eight caravanserais, with each exploration taking 3–5 min on average. Users can teleport between nodes or walk to simplify navigation. New music plays each time a user jumps to a new node, which is selected to highlight key points, such as the entrance door with prominent decoration, the connected courtyard, and enclosed spaces.

To enhance the generated soundscapes, we incorporate contextual information from the architectural spaces where the music is experienced. This serves two purposes: (i) it ensures consistent ***makam*** selection throughout the experience, aligning with the characteristics of each space while maintaining stylistic coherence; and (ii) it enriches the experience by potentially adding audio tracks (not necessarily music). This step aims to bring the unique character of each caravanserai to life by blending its architectural details, historical context, and social activities to create a culturally relevant soundscape. Alongside the music caption, it aggregates information from various sources. An example of a context caption for Evdir Han is provided in Fig. 4.

3.3 Text-to-Music Soundscape Generation

Our soundscape generation process relies on state-of-the-art **text-to-music** audio models. The decision to rely on audio music generation is due to the repertoire gap of the Seljukid period, on which some reconstruction attempts could be observed (see Uslu [1]). Once annotated with their captions, the acoustic manifestations of these intended reconstructions (the Seljukid Dataset) contribute training material (audio-caption pairs) to fine-tune a generative model for this particular domain.

We are experimenting with two open music models: **MusicGen 3.1b** by **Meta AI** [35]; **AudioLDM 2** [36]; and a proprietary model for comparison, namely **Stable Audio AudioSparx 2.0** by **Stability AI**. While we advocate responsible development and the use of open music models, **Stable Audio** provides a more user-friendly solution for fast prototyping.

The simulation generation process involves two steps. First, in **Program Creation**, the context caption of the chosen caravanserai and a selection of music captions are provided to an LLM, which generates a program outlining the overall musical experience within the caravanserai. This program specifies ***makam*** descriptions, musical structures, ***usuls***, instruments, emotional qualities, and contextual augmentations for each node. Next, in **Soundscape Generation**, each node-level prompt is fed into a text-to-music model, which generates audio snippets.



Caravanserai: Evdir Han
Caption: Located 18 km north of Antalya, the port city, it was built during the reign of Sultan Izzeddin Keykavus I (1210-1219) and stands as one of the largest Seljuk caravanserai in Anatolia. The han's unique design features four iwans surrounding a central open courtyard, resembling caravanserais found in Central Asia and Iran, with adaptations for Antalya's warm climate. Constructed from finely-cut stone, it showcases classic Seljuk architectural elements like pointed arches and vaulted ceilings supported by piers. The structure includes 14 cells on the east and west sides and 11 on the north and south, with much of the southern side remaining intact despite some degradation. The entrance is marked by a large pointed arch with decorative motifs, including geometric patterns and arabesques, typical of Seljuk style, serving both as a focal point and a functional element, providing access to the courtyard while ensuring protection and privacy. Following the "Open Courtyard" (OC) design, the rooms are distributed around the central open courtyard, with no covered sections, making the courtyard the core of the structure.

Fig. 4 The caption of the architectural features of Evdir Han

This depends on the model's generation capabilities, as the prompt may be divided into segments (e.g. **MusicGen** produces 30 s snippets; hence, for a 3 min piece, approximately 6 segments would be required).

Given that current music models are predominantly biased towards Western music, which is a larger problem in music information retrieval [37], careful crafting of prompts is needed to emphasise the unique characteristics of Seljukid ***makam*** music (e.g. microtonal inflexions, specific rhythmic patterns) to guide the model towards more authentic results. Also, preliminary audio generations are subject to iterative refinement by the active involvement of experts in Turkish classical music.

3.4 Acoustic Simulation

To enhance the immersive experience, the project simulates how the AI-generated music would have sounded within the unique acoustic environment of each caravanserai. This involves two steps: capturing the acoustic "fingerprint" of each space

and then using that data to create auralisations of the generated music. Acoustic data is acquired either through direct Impulse Response (IR) measurements, involving the recording of a brief, broadband sound's reverberation within the space, or through acoustic simulations based on detailed 3D models of the caravanserais. This is done using Rhino with the Pachyderm plugin, which models sound propagation based on assigned material properties.

The second step integrates the AI-generated music with the captured acoustic data through a process called convolution. This combines the "dry" AI-generated music with the impulse response of a specific caravanserai, resulting in an auralisation that simulates how the music would sound if played in that space. This process effectively "places" the music within the historical acoustic environment.

4 System Evaluation

The research explores four key questions surrounding the integration of music and acoustics in a VR experience focussed on caravanserais. **RQ1** investigates the ability of a model to generate music that is both historically accurate and acoustically relevant to traditional sounds. **RQ2** examines whether contextualising the caravanserai setting enhances the music generation process and adds value to the overall user experience. **RQ3** explores the impact of the ***makam*** programme on creating a more coherent musical experience during the exploration of the caravanserais. Finally, **RQ4** looks into the role of acoustic simulation, assessing its contribution to enhancing the immersive quality of the VR experience. These questions collectively aim to increase understanding of how music, context, and acoustics can shape the effectiveness of virtual heritage experiences.

The evaluation of MAKAIM's generated music centres on a structured listening study designed to assess its historical plausibility, given the absence of verifiable audio recordings from the Seljukid period (RQ1). Two expert musicologists, specialising in Turkish classical music and knowledgeable about the Seljukid era, will participate in a blind evaluation. The evaluation will examine specific musical dimensions, including ***makam*** scale accuracy, ***seyir***, ***usul***, instrumental timbre authenticity, and stylistic appropriateness. Experts will rate each generated audio stimulus on a 7-point Likert scale across these dimensions and provide qualitative feedback. The stimuli will encompass a variety of makams, ***usuls***, and instrumental combinations, generated with variations in prompts and model parameters. Control stimuli, consisting of contemporary recordings of Turkish classical music, will also be included for comparison. Data analysis will involve descriptive statistics and inter-rater reliability metrics (e.g. Cohen's Kappa, ICC) to assess the system's performance and the agreement between experts.

The iterative refinement process will leverage the results of the listening study to improve the music generation process. Feedback from the expert evaluations, both quantitative ratings and qualitative comments, will be used to adjust prompts, fine-tune the AI model, and refine the acoustic simulation parameters. For instance, if a

particular ***makam's seyir*** is consistently rated as inaccurate, the prompts associated with that ***makam*** will be revised, or the model may undergo further fine-tuning. This iterative feedback loop is crucial to enhancing the historical plausibility of the generated music. Additionally, user studies within the VR environment, addressing research questions RQ2 to RQ4, will assess the impact of contextualisation, ***makam*** programming, and acoustic simulation on the overall user experience, providing further data for refinement.

4.1 Technical Challenges, Limitations, and Ethical Considerations

Reconstructing Seljukid-era soundscapes with AI presents several technical challenges. The most significant is the scarcity of data on Seljukid ***makam*** music, coupled with the fact that existing text-to-music models are primarily trained on Western music, potentially leading to stylistic inaccuracies due to fundamental differences in musical structures, such as microtonality. While fine-tuning on limited data and related musical traditions can help, it may not fully address these differences. Additionally, accurately representing the timbral properties of historical instruments poses a challenge, requiring careful prompt engineering and potentially leveraging techniques like timbral manipulation. The accuracy of the acoustic simulation is also crucial, relying on the quality of 3D models, material properties, and the sensitivity of impulse response measurements to environmental factors.

The use of AI in this project raises ethical and sociocultural considerations. It is essential to acknowledge that the generated soundscapes are speculative reconstructions, not authentic reproductions. Given that ***makam*** music is a living tradition with deep cultural significance, the project is mindful of cultural sensitivities to avoid perpetuating stereotypes while engaging with contemporary practitioners. Responsible use of AI models is also crucial. While the models used are open-source, they were trained on datasets that may contain copyrighted material. Therefore, the generated music in this project is intended solely for cultural and educational purposes within the immersive experience.

5 Conclusions

This project conceptualises MAKAIM—a novel, AI-driven framework designed to reconstruct and reimagine the historical soundscapes of Seljukid caravanserais. By integrating expert-curated music captions, contextual information derived from architectural and historical sources, state-of-the-art generative text-to-music models, and acoustic simulations, MAKAIM offers a unique approach to reviving and experiencing lost musical traditions.

While the project faces technical challenges, including the inherent bias of current music models towards Western music and the limitations of accurately capturing the timbral properties of historical instruments, these are mitigated through a combination of careful prompt engineering, model fine-tuning, and iterative expert review. The incorporation of contextual information from the historical venues and the use of spatial audio techniques further enhances the realism of the auralisations, contributing to a multisensory and historically grounded VR experience.

The evaluation framework, centred on the expert assesment of musical plausibility across various dimensions, ensures that the generated soundscapes align with historical understanding and musicological knowledge. This approach, coupled with the planned user studies in the VR environment, will provide valuable insights into the effectiveness of MAKAIM in creating a sense of presence, authenticity, and engagement.

Ultimately, this research demonstrates the potential of Generative AI and immersive technologies to contribute to cultural heritage preservation by enabling the reconstruction of lost soundscapes. By bridging tangible and intagible dimensions of heritage, MAKAIM offers a pathway to enriching our understanding of the past, making the legacies heritage of Seljukid music accessible to contemporary audiences within a historically informed and immersive context.

Acknowledgements We would like to express our gratitude and acknowledgement to Prof. Nagakura for the training he provided on photogrammetry, S. C. Seyedian Karadagli, H. Tu, X. Zhang, I. Ercan, O. Borekci, H. Kaleli, M. Ozteker, B. Bekiroglu for data collection, and V. Yucel, H. Tu, and K. Maes for VR implementation. The fieldwork for this project is funded by the U.S. Department of State, U.S. Mission Turkiye Program, while the VR implementation and music research is funded by the RDIF Grant at the Liverpool University School of Architecture. The academic research on Seljukid Caravanserais by Dr. Varinlioglu is funded by Fulbright, Chevening, and TUBITAK.

References

1. Uslu R (2010) Selçuklu topraklarında müzik [Music in Seljuk Lands]. Konya Valiliği, İl Kültür ve Turizm Müdürlüğü, Konya
2. Varinlioglu G, Jayakrishnan D, Mukarrama M, Gulshan S, Puliekandy A, de Berardinis J (2025) Ambisonic acoustic simulations for historical reconstruction. In: 43rd Education and research in computer aided architectural design in Europe (eCAADe), pp 29–38
3. Kersten TP, Lindstaedt M (2012) Image-based low-cost systems for automatic 3D recording and modelling of archaeological finds and objects. Progress in cultural heritage preservation. Springer, Berlin Heidelberg, pp 1–10
4. Verhoeven G (2011) Taking computer vision aloft: archaeological three-dimensional reconstructions from aerial photographs with PhotoScan. Archaeol Prospect 18(1):67–73
5. Croce V, Caroti G, De Luca L, Piemonte A, Véron P (2023) Neural radiance fields (NeRF): review and potential applications to digital cultural heritage. Int Arch Photogramm Remote Sens Spat Inf Sci 48:453–460

6. Bruno F, Bruno S, De Sensi G, Luchi ML, Mancuso S, Muzzupappa M (2010) From 3D reconstruction to virtual reality: a complete methodology for digital archaeological exhibition. J Cult Herit 11(1):42–49
7. Forte M, Kurillo G (2010) Cyber-archaeology: experimenting with teleimmersive archaeological reconstruction. Virtual Archaeol Rev 1(2):21–25
8. Kersten TP et al (2017) The Selimiye Mosque of Edirne, Turkey—an immersive and interactive virtual reality experience using HTC Vive. Int Arch Photogramm, Remote Sens Spat Inf Sci XLII-5/W1:403–409
9. Agostinelli A et al (2023) MusicLM: generating music from text. arXiv. https://doi.org/10.48550/ARXIV.2301.11325. Accessed 2 Feb 2025
10. Başarır L, Çiçek S, Koç M (2024) Local intelligence: time to learn from AI. Arch Sci Rev 1–16
11. Güray C, Güvener D, Yıldız Ş (2023) Adakale'nin Saklı Türkülerinin Markov Zinciri Tabanlı Stokastik Yöntem ile Yeniden Yapılandırılması. Etnomüzikoloji Dergisi 6(3):38–58
12. Karaosmanoğlu MK (2012) A Turkish makam music symbolic database for music information retrieval: SymbTr. In: Proceedings of the 13th international society for music information retrieval conference, 8–12 October 2012, Porto, Portugal. ISMIR, pp 223–228
13. Sertan S, Chordia P (2011) Modeling melodic improvisation in Turkish folk music using variable-length Markov models. In: Proceedings of the 12th international society for music information retrieval conference, pp 269–274
14. Parlak IH et al (2021) Deep learning for Turkish makam music composition. Turk J Electr Eng Comput Sci 29(7):3107–3118
15. Aydıngün A et al (2020) Turkish music generation using deep learning. In: 2020 28th signal processing and communications applications conference (SIU). IEEE, pp 1–4
16. Kaşif A, Sevgen S (2024) Classical Turkish music composition with LSTM self-attention. J Innov Sci Eng 8(1):25–35
17. Akalın Eryavuz Ş (2022) Kervansaray. In: Türkiye Diyanet Vakfı İslam Ansiklopedisi. Türkiye Diyanet Vakfı, Istanbul. https://islamansiklopedisi.org.tr/kervansaray. Accessed 16 Jan 2025
18. Tuncer OC (2007) Anadolu Kervan Yolları. Vakıflar Genel Müdürlüğü Yayınları
19. Varinlioglu G, Börekçi Ö (2025) Towards a geospatial and temporal experience of islamic architecture: caravanserais of Anatolia. In: Jameel S, Clary DC (eds) Disruptive technologies and Muslim societies: from AI and education to food and fintech. World Scientific Publishing, Singapore, pp 85–102
20. Varinlioglu G, Balaban Ö (2021) Artificial intelligence in architectural heritage research: simulating networks of caravanserais through machine learning. In: The Routledge companion to artificial intelligence in architecture. Routledge, pp 207–223
21. Branning K (2001–2020) Turkish Hans. http://www.turkishhan.org/. Accessed 1 Feb 2025
22. Özergin KM (1965) Anadolu'da Selçuklu kervansarayları. Tarih Dergisi 15(20):141–170
23. İlter İ (1969) Tarihi Türk Hanları: Caravanserais. Karayolları Genel Müdürlüğü
24. Acun H (ed) (2007) Anadolu Selçuklu Dönemi Kervansarayları. Kültür ve Turizm Bakanlığı, Ankara
25. Johansen EM (2016) Rūmselçuk caravanserais of Anatolia. Else Marie Johansen, Copenhagen
26. Erdmann K, Erdmann H (1961) Das Anatolische Karavansaray des 13. Jahrhunderts. Verlag Gebr, Mann, Berlin
27. Yavuz AT (1997) The concepts that shape Anatolian Seljuq caravanserais. Muqarnas 14:80–95
28. Popescu-Judetz E (1996) Türk Musıki Kültürünün Anlamları. Pan Yayıncılık, İstanbul, p 17
29. Signel K (2006) Makam: Türk Sanat Musikisinde Makam Uygulaması. YKY, İstanbul, p 35
30. Aksoy B (n.d.) Turkish classical music: the makam phenomenon in Ottoman Turkish music. http://www.turkishmusicportal.org/en/types-of-turkish-music/turkish-classical-music-the-makam-phenomenon-in-ottoman-turkish-music. Accessed 22 Jan 2025
31. Aydemir M (2010) Turkish Music Makam Guide. Pan, İstanbul, pp 13–14
32. Kara S (2014) Selçuklu Türkiye'sinde Eğlence Türü Olarak Bezm ve Musiki. Bilig 68:169–182
33. Uslu R (2000) Musicians in the Ottoman Empire and Central Asia in the 15th century according to an unknown work of Aydınlı Nahifi. In: The great Ottoman Turkish civilization, vol IV, pp 548–555

34. Uslu R (2015) Is an echo of Seljuk music audible: a methodological research. In: Greve M (ed) Writing the history of Ottoman music. Nomos, Würzburg, pp 241–246
35. Copet J et al (2024) Simple and controllable music generation. In: Advances in neural information processing systems, vol 36
36. Liu H, Yuan Y, Liu X, Mei X, Kong Q, Tian Q, Plumbley MD (2024) Audioldm 2: learning holistic audio generation with self-supervised pretraining. IEEE/ACM Trans Audio, Speech, Lang Processing
37. Holzapfel A, Sturm B, Coeckelbergh M (2018) Ethical dimensions of music information retrieval technology. Trans Int Soc Music Inf Retr 1(1):44–55