A music recommendation system for constructed music-evoked episodic memories (CoMEEMs)

Raingeard de la Bletiere, P.; Neerincx, M.; Schaefer, R.S.; Oertel, C.

Int. J. Human–Computer Studies 208 (2026) 103717

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

A music recommendation system for constructed music-evoked episodic memories (CoMEEMs)

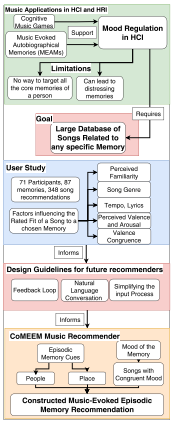
Paul Raingeard de la Bletiere [a,*], Mark Neerincx [a], Rebecca Schaefer [b], Catharine Oertel [a]

[a] Delft University of Technology, Van Mourik Broekmanweg 6, Delft, 2628 XE, the Netherlands
[b] Leiden University, Rapenburg 70, Leiden, 2311 EZ, the Netherlands

HIGHLIGHTS

- Interactive systems lack mood regulation functions for personalized affective support; music-linked memories enable this.
- CoMEEMs link music to existing memories using People, Place cues, and mood congruence.
- Familiarity and genre best predict song fit (small to medium effects).
- These are followed by valence, arousal, tempo, and lyrics (small effects).
- Familiar-sounding songs and user feedback should be prioritized for cognitive-affective support.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Music
Memory
Music recommender
Music evoked autobiographical memories
Human-computer interaction
User study

ABSTRACT

Music is widely used in human–computer interaction (HCI) to enhance engagement, sustain attention, and support cognitive stimulation. Yet its potential for deliberate mood regulation, particularly through personalized memory recall, remains largely unexplored.

Music-evoked autobiographical memories (MEAMs) are often elicited by well-known, favorite songs, yielding stronger mood effects than music without personal memory associations. However, songs can also trigger distressing memories, and will never capture all positive personal memories. Since happy personal memories can enhance mood, broader methods for retrieval are needed.

To address this, we introduce Constructed Music-Evoked Episodic Memories (CoMEEMs), a framework linking chosen episodic memories to music. By creating a personalized song-memory database, CoMEEMs enable autonomous mood regulation and communication in interactive systems, integrating memory cues—such as people and places—alongside mood congruence, to help choose songs with high mood regulatory impact.

In an experiment with 71 Dutch and French adults, participants described 87 positive memories and received song recommendations based on associated people and places, with and without mood matching. Results showed that song familiarity and genre were the strongest predictors of perceived fit, while valence, arousal, tempo, and lyrics played smaller roles. Mood congruence, especially in valence, significantly influenced song relevance. Participants emphasized the need for user input on emotional states and memory context. Based on these findings, we propose design guidelines to improve future music recommendation systems targeting memories.

* Corresponding author.
Email address: p.raingearddelabletiere@tudelft.nl (P. Raingeard de la Bletiere).

https://doi.org/10.1016/j.ijhcs.2025.103717
Received 11 April 2025; Received in revised form 25 November 2025; Accepted 15 December 2025
Available online 17 December 2025

1. Introduction

Music is widely integrated into interactive systems as a tool for entertainment, cognitive stimulation, and user engagement (Cuan et al., 2024; Tikkanen and Iivari, 2011; Samson et al., 2025). Its ability to capture attention and evoke emotional responses makes it particularly effective in maintaining user interest, enhancing interaction quality, and supporting cognitive functions such as memory, which are essential for learning, task performance, and psychological well-being. Interactive systems often use familiar songs or music-based games to sustain engagement (Tanner et al., 2023; Agres et al., 2019), digital games leverage dynamic soundtracks to enhance immersion and enjoyment (Farkas et al., 2022), and rehabilitation or learning systems use interactive music-making to motivate users (Baur et al., 2018). However, music in most interactive systems primarily serves motivation or enjoyment, rather than deliberate mood regulation or personalized reminiscence. Designing systems that tailor musical content to users' emotional states and personal histories could enhance engagement, satisfaction, and overall user experience (Tz-Han et al., 2023; Goodall et al., 2021).

Mood regulation, here defined as the deliberate use of stimuli to stabilize negative emotional states, is a critical but often overlooked dimension in interactive system design. While music's ability to evoke emotions and memories is well established (Janata et al., 2007), most systems do not actively tailor musical content to users' emotional needs or personal histories. This is despite growing evidence that tailored interventions can reduce stress and support emotional well-being (Goodall et al., 2021), with particularly strong effects observed in populations such as older adults. Interactive systems can leverage autobiographical memory, which is powerfully activated by music and can meaningfully influence mood and engagement across users (Dudzik et al., 2020). Despite this connection, memory's role in mood regulation has received limited attention in previous systems.

Linking memory and music, Music-Evoked Autobiographical Memories (MEAMs) stand out as naturally occurring phenomena where a familiar song triggers vivid, often emotionally salient personal memories (Belfi et al., 2020). MEAMs have been shown to foster communication, improve mood, and reduce depressive symptoms in diverse populations, including older adults (Baird et al., 2018; Völker, 2019). However, MEAMs are inherently unpredictable and uncontrollable, sometimes evoking distressing memories (Mehl et al., 2024). Moreover, relying solely on naturally occurring MEAMs means that certain core memories that are potentially crucial for mood regulation given their highly personal relevance (Fernández-Pérez et al., 2023) may remain inaccessible through music.

To address these limitations in interactive systems and MEAMs, we introduce **Constructed Music-Evoked Episodic Memories (CoMEEMs)**. Unlike spontaneous MEAMs, CoMEEMs involve deliberately linking songs with specific episodic memories—events tied to particular times and places (Schacter and Addis, 2007). By creating controlled, personally relevant song-memory connections, CoMEEMs provide a structured approach to evoke positive recollections and support mood regulation in interactive systems. We evaluate these CoMEEMs by measuring the perceived fit between a specific song and a given episodic memory. Using this measure of perceived fit, we aim to select songs that users can easily associate with their memories due to their high personal relevance, which generally enhances recall (Fernández-Pérez et al., 2023). This, in turn, would help them intuitively recall positive events and form emotional connections without requiring specialized training.

This study explores the feasibility of CoMEEMs and develops design guidelines for music recommender systems that actively link songs to episodic memories to support mood regulation. We examine three research questions:

- **RQ1:** What factors related to the music listening experience predict the perceived fit of a song to a specific episodic memory?
- **RQ2:** Which types of episodic memory cues determine the perceived fit of songs to memories?
- **RQ3:** Does congruence between song-evoked and memory-evoked moods influence this fit?

To account for natural variations in autobiographical memory and emotional responses, our study includes three age cohorts, capturing differences in memory content and musical associations. This design allows us to investigate how CoMEEMs can be tailored across users with diverse life experiences, informing the design of interactive systems that support memory recall and mood regulation more broadly.

This paper makes the following contributions:

- Introduction of the **Constructed Music-Evoked Episodic Memories** (CoMEEM) concept as an active, personalized approach to linking songs with specific episodic memories, complementing the unstructured nature of MEAMs.
- An empirical investigation of factors influencing perceived song-memory fit, focusing on subjective listening experiences, memory cues, and mood congruence.
- Design guidelines for music recommender systems aimed at mood regulation through episodic memory recall.

2. Related works

In the following section, we delve into related works in psychology that drive our design choices and hypotheses, as well as studies that have explored music recommendations in therapeutic settings, highlighting the advancements made in mood-based suggestions and memory integration in music recommender systems.

2.1. Song characteristics in MEAMs

Genre and familiarity are the two main factors influencing the frequency of MEAMs (Jakubowski and Francini, 2022). Greater engagement with a song's genre increases the likelihood of experiencing an MEAM (Jakubowski et al., 2024). Additionally, high familiarity with a song is associated with more frequent MEAMs, often aligning in valence with the emotional tone of the song (Jakubowski and Francini, 2022). Personal appreciation of a song also plays a significant role, as preferred songs are more likely to trigger memory retrieval (Kaiser and Berntsen, 2022). Familiarity and preference are interconnected; more familiar songs tend to be more liked, and stronger engagement with a genre increases enjoyment of songs within that genre (Jakubowski and Francini, 2022). However, self-rated familiarity only predicts MEAM strength for previously heard songs, with much less effect for unfamiliar songs (Kathios et al., 2024).

The era of a song also plays a critical role, with a higher incidence of MEAMs linked to music heard between the ages of 15–25, consistent with the memory bump effect (Zimprich, 2018). Research suggests that top-charting songs from this period have a 30 % chance of being associated with episodic memories (Janata et al., 2007).

In the context of CoMEEMs, familiarity may exert a weaker influence on memory association. Familiar songs, already linked to past MEAMs, may be less likely to form new memory connections. Additionally, while exposure to a familiar-sounding genre does not necessarily increase memory linkage (Kathios et al., 2024), familiarity is not expected to be a strong predictor of fit in CoMEEMs.

2.2. Personal characteristics and age factors in episodic memories and MEAMs

Age significantly affects episodic memory. Older adults tend to recall fewer and less specific episodic memories (Greene and Naveh-Benjamin, 2023; Kinugawa et al., 2013). Yet MEAM frequency remains stable across age groups and tends to be more positive in older individuals (Jakubowski et al., 2023; Mehl et al., 2024).

Personality traits also shape episodic memory recall; higher openness and lower neuroticism correlate with greater memory retrieval (Stephan et al., 2020). Beyond personality, cultural background, personal influences, rewards, emotional significance, stress, and activity levels all contribute to memory recall (Morales-Calva and Leal, 2024; Festini and McDonough, 2024).

For MEAMs, personal appreciation of music is linked to increased recall frequency, while greater auditory and visual imagery enhance memory vividness (Jakubowski et al., 2024). Despite these individual differences, age remains a key factor, with older adults reporting more positive MEAMs than younger individuals (Mehl et al., 2024).

Given the role of age and individual differences in episodic memory formation, similar effects are expected in CoMEEMs, where subjective perception and age likely influence the perceived fit between a song and a memory.

2.3. Episodic memory representation

Four key factors influence episodic memory: the location of the event, the time it occurred, the specific action or event, and the people involved (Tulving, 1984). Given their time-bound nature, models of episodic memories are often dynamic and kinematic (Andonovski, 2022; Jeunehomme et al., 2022).

Specific elements of episodic memories serve as cues for recall. Lee and Dey (2007) identified four primary types of cues naturally used by individuals: people, places, objects, and actions, later expanded to include animal cues (Baumann et al., 2024). However, people, places, and actions remain the dominant recall cues. People cues were overall the most commonly used, suggesting that songs associated with individuals from a specific memory may enhance the perceived fit between the song and the memory.

Our study simplifies episodic memory modeling to explore foundational links between music and autobiographical recollections (Chater and Vitányi, 2003). We focus on people and place cues, as these have been central in prior MEAM research (Janata et al., 2007).

As it will be shown in the next section, beyond cues, mood also plays a crucial role in recall, influencing how individuals connect music to past experiences.

2.4. Mood-congruent episodic memories and music-induced mood

Mood is commonly described in terms of perceived valence, arousal, and dominance, though the latter is less frequently used (Posner et al., 2005). According to mood-congruent memory theory, a mood similar to the one experienced during memory formation enhances recall (Faul and LaBar, 2023). Valence-matching improves retrieval (Lewis et al., 2005), while arousal- matching enhances specificity and reduces false memories (Talarico et al., 2004; Mirandola and Toffalini, 2016). However, this relationship is complex, as valence and arousal interact in encoding and retrieval (Greene et al., 2010). To simplify this dynamic, we use Tulving's theory of episodic memory, which suggests that mood-matching at encoding and retrieval facilitates recall (Tulving, 1984), a process music can support when emotional states align during both phases (de l'Etoile, 2002).

Research indicates that a song's valence and arousal evoke corresponding listener moods (Västfjäll, 2001). Music is a powerful mood inducer, particularly when combined with visual stimuli (Jallais and Gilet, 2010). While mood induction studies typically use long excerpts (5–10 minutes) for sustained effects (Cheng et al., 2017), shorter clips can also elicit emotions, though mood tends to revert to baseline after a few minutes (Ribeiro et al., 2019). In memory-evoking songs, retrieval follows mood-congruence principles: valence aligns with memory positivity, and arousal correlates with intensity (Jakubowski and Francini, 2022). Therefore, we expect that songs with self-rated valence and arousal levels similar to those of a supplied memory will be perceived as a better fit for that memory.

While insights from cognitive psychology guide our hypotheses, we must also examine the technical capabilities of recommender systems to determine which design elements are most suitable for our use case.

2.5. Extending MEAMs, towards CoMEEMs

Music-evoked autobiographical memories (MEAMs) have been shown to support well-being by fostering communication, reducing depressive symptoms, and improving mood (Baird et al., 2018; Völker, 2019). Despite their therapeutic value, MEAMs are spontaneous and unpredictable: not all personally significant memories are accessible through music, and occasionally negative or irrelevant memories are triggered (Mehl et al., 2024). This unpredictability constrains their systematic use in interactive systems.

To address these limitations, we propose the concept of Constructed Music-Evoked Episodic Memories (CoMEEMs). In contrast to MEAMs, CoMEEMs deliberately link specific songs to chosen episodic memories—personally meaningful events tied to particular times and places (Schacter and Addis, 2007). This structured approach enhances predictability, ensures greater personal relevance, and provides a foundation for integrating music-based reminiscence into therapeutic and interactive contexts.

Fig. 1 summarizes this distinction. MEAMs rely on incidental memory activation, while CoMEEMs leverage cognitive principles such as the encoding specificity principle (Tulving and Thomson, 1973) and mood congruence (Faul and LaBar, 2023), together with episodic memory cues (e.g., people, places, objects, and actions) (Baumann et al., 2024; Lee and Dey, 2007). This controlled framework increases the likelihood that evoked memories are both positive and therapeutically useful.

In summary, MEAMs and CoMEEMs differ along three dimensions: (1) the *nature of recall* (spontaneous versus constructed), (2) *predictability and personal relevance* (uncontrolled versus targeted), and (3) *application potential* (incidental outcomes versus deliberate integration into recommenders and interactive systems). These distinctions position CoMEEMs as a novel conceptual bridge between psychological findings on memory and mood, and technical work in music recommender systems.

2.6. Technical works in music recommendation systems

2.6.1. Music recommendation systems based on LLMs and conversational agents

For personalization, Large Language Models (LLMs) have been used as recommenders, with zero- to few-shot learning achieving similar performance to state-of-the-art recommendation systems (Sanner et al., 2023)through combinations with collaborative filtering (Kim et al., 2024) and reinforcement learning (Wu et al., 2024). Tracking user interactions helps refine recommendations based on evolving preferences (Xi et al., 2024), while diversity-aware ranking improves variety and relevance (Carraro and Bridge, 2024). These strengths make LLMs valuable for both personalization and diversification, aiding in designing a system that targets a broad range of memories.

Beyond background recommendation, LLMs can also engage users through conversational recommendation, where interactive feedback refines suggestions. Hybrid (user- and system-generated feedback) and cascading critiquing methods enhance personalization (Jin et al., 2019; Cai et al., 2021) but introduce challenges, such as unpredictable system behaviors. To ensure reliability, we initially use LLMs only as background recommenders, with plans to expand into full conversational interactions.

2.6.2. Music recommendation for MEAMs

For emotion regulation, Mou et al. (2021) developed a music recommendation system that considers potential MEAMs in song selection. While not targeting specific episodic memories, they found that MEAMs have a stronger impact on emotion regulation than memory-free music. Rao et al. (2021) explored song recommendations from the reminiscence

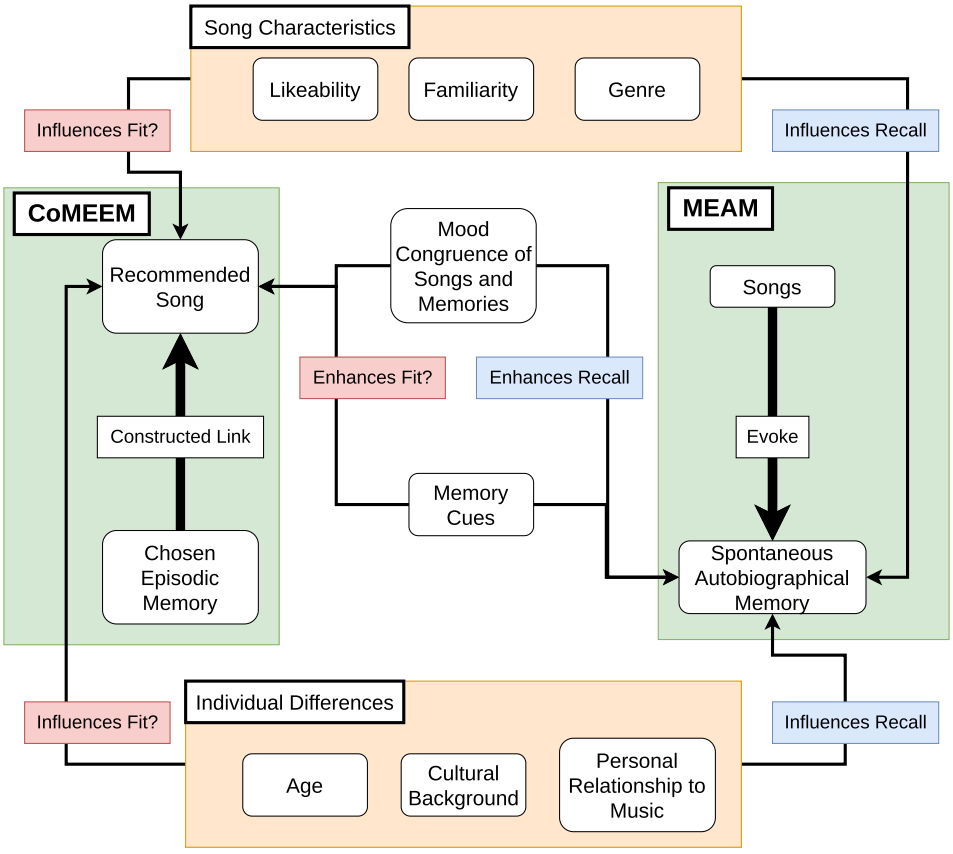


Fig. 1. Characterization of CoMEEMs and MEAMs. Mood Congruence, Memory Cues, Individual and Song Characteristics influence these concepts by enhancing memory recall (MEAMs) or potentially increasing perceived fit of songs to memories (CoMEEMs).

bump (ages 15–25) but found no conclusive effect on a better memory recall.

Beyond music, Sion et al. (2023) demonstrated that vibrotactile patterns can be linked to specific memories, but this also required training over several sessions. So far, no study has successfully linked specific memories to patterns—whether songs or vibrations—without several training sessions. This study aims to bridge that gap by investigating how well a music recommendation system can connect memories to existing songs without requiring extensive training.

To our knowledge, our study is the first one targeting specific memories and recommending songs, instead of recommending songs and assessing which memories are evoked by them. We take inspiration from LLM-based, therapeutic and MEAM music recommender systems to design our own recommender system. We test the following hypotheses taken from previous MEAM, music, and memory research:

- H1a (related to RQ1): Familiarity has minimal influence on perceived fit, contrary to the case of standard MEAMs.
- H1b (related to RQ1): Age and personal evaluation of songs are highly predictive of the fit of a song to a specific memory.
- H2 (related to RQ2): Songs related to People cues are generally associated with a higher fit to a memory compared to songs related to Place cues.
- H3 (related to RQ3): The alignment between song-evoked and memory-evoked mood increases the perceived fit of a song to a memory.

Hypotheses H1a and H1b refine RQ1 by testing specific factors from the music listening experience that may influence the perceived fit of a song to an episodic memory. H2 operationalizes RQ2 by examining whether certain types of memory cues—specifically People versus Place—differ in their association strength with songs. H3 corresponds to RQ3 and investigates whether emotional congruence between a song and a memory serves as a mechanism for perceived fit.

Testing these hypotheses can help us provide guidelines for future recommendation systems for CoMEEMs, their outcomes providing directions on how to personalize and optimize such interactions.

3. The CoMEEM recommender system

3.1. Requirements

Following Socio-Cognitive Engineering, a framework for the design of human centered technologies, we design our recommender system by defining it through functions and corresponding effects (i.e., the claims) for the considered use cases, explicating the theoretical and empirical foundation (i.e, the design rationale and remaining research challenges or questions) (Sharples et al., 2002; Neerincx et al., 2019). The system aims to support mood regulation by linking songs to episodic memories, and rests on three central design claims:

- Cue-based association: Users can better associate songs and episodic memories when recommendations incorporate cues such as people, places, and moods. (Function: extract memory elements and their associated cues).
- Mood congruence: Matching the arousal and valence of a memory to the mood of a song increases the perceived fit between the song and memory. (Function: access and process memory mood to guide recommendation).
- Ontology-driven personalization: Building a growing database of songs enriched with formalized memory cues enables more personalized and reusable recommendations over time. (Function: create an ontology to store and retrieve memory–song links).

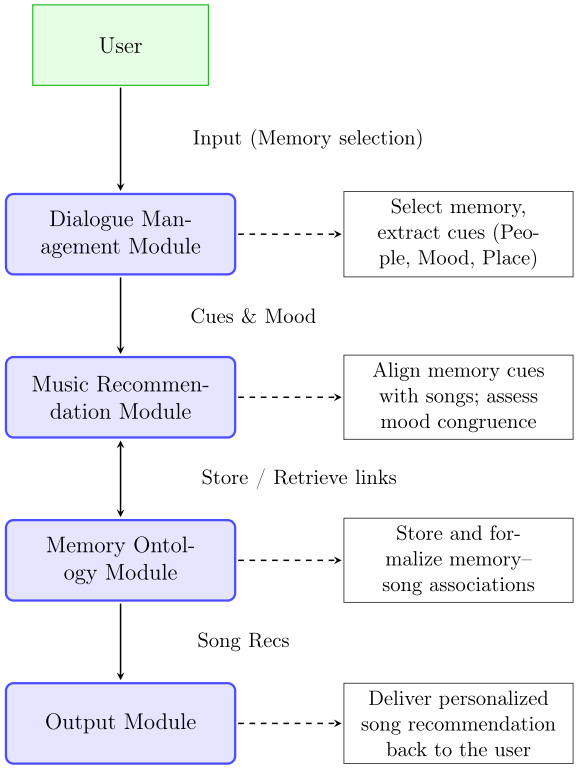


Fig. 2. Operational framework of the CoMEEM system. The four interconnected modules work sequentially to facilitate personalized music recommendations based on episodic memories.

These claims are realized through four interconnected modules (Fig. 2): (1) the Dialogue Management Module, which enables users to select memories and extracts associated cues; (2) the Music Recommendation Module, which aligns songs with memory cues and mood; (3) the Memory Ontology Module, which stores and organizes memory–song associations for future reuse; and (4) the Output Module, which delivers personalized song recommendations.

In the following subsections, we describe the operation of these modules in detail.

3.2. Dialogue management module

This module follows the structure outlined in Fig. 3.

The module operates using a rule-based approach, employing predefined questions to extract key memory cues and emotional context.

Users start by selecting a positive memory without an associated song, ensuring classical MEAMs are excluded. Any memory can be selected and entered in a free-text description, without any maximum length. While any memory can be stored in the ontology for future reference, those already linked to a song bypass the dialogue and recommendation modules.

The CoMEEM system considers not only autobiographical memories—core past events—but also recurring episodic memories. Autobiographical memories serve as an organizing framework for episodic memories (Conway, 2009), while CoMEEMs target any naturally recalled episodic event, including, although rare, recurring experiences (Means and Loftus, 1991).

Once a memory is selected, the user describes it and specifies whether it represents a unique event (e.g., a birthday, vacation) or a recurring experience (e.g., walking in the park, playing sports).

3.2.1. Cue extraction

The system then extracts key People and Place cues for song recommendations. Pre-existing associations between people or places and

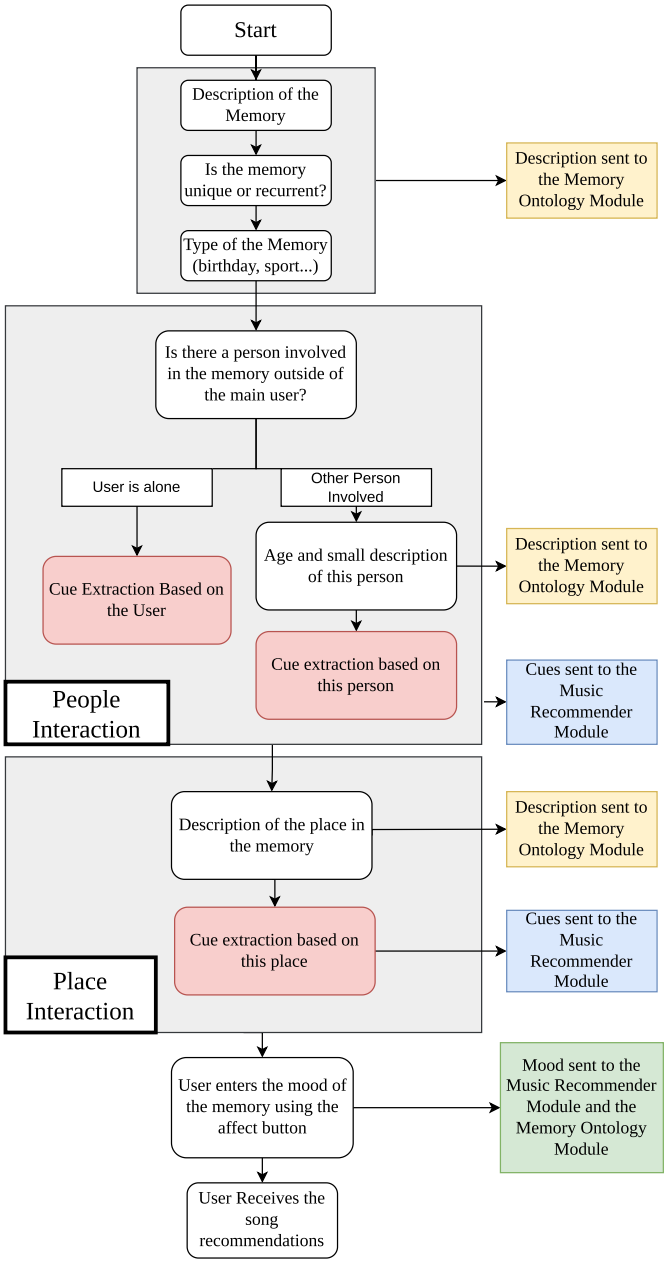


Fig. 3. Flowchart of the Dialogue Management Module. This module extracts People and Place cues, as well as the mood related to the chosen memory through a conversational agent.

specific artists or genres are considered, but direct links to particular songs are excluded since they are likely tied to other memories.

For People cues, the user is asked if someone else was present in the memory. If so, they specify whether that person is associated with a particular artist or genre. If no direct link exists, the recommendation module generates a probable cue (see Section 3.3). If multiple people are involved, the user selects the most important individual to simplify cue processing. Future versions of the dialogue management module could refine this approach by allowing for group-based distinctions.

If the memory involves only the user, they select a favorite artist, similar to (Fraile et al., 2019) methodology. If none exists, they choose a preferred genre or musical style.

For Place cues, the user states whether they associate the location with a specific genre or artist. If no clear association exists, the system generates a probable cue.

Table 1
Scenarios used for extracting relevant cues.

People cues	A specific person is part of the memory (if a group, then one member of that group)	An artist is associated with this person
		A genre is associated with this person
		Nothing is associated (cue generation from the system)
	Only the user is part of the memory	The user has a favorite artist
		The user has a favorite style of music (genre or more abstract)
Place cues	A specific place is part of the memory	An artist is associated with this place
		A genre is associated with this place
		Nothing is associated (cue generation from the system)

Table 2
Scenarios and Features Considered in the Prompt.

Scenario	Features taken into account in the Prompt
Favorite style of music of the user	Given cue (Preferred style of music), year of birth of the user + 23, type of event in the memory, cutoff year (the date of the memory)
Genre linked to the chosen person	Given cue (Genre), year of birth of the person + 23, country of the songs listened to by the person (or worldwide if international songs), type of event in the memory
No People Cue	Year of birth of the person + 23, country of the songs listened to by the person (or worldwide if international songs), type of event in the memory
Genre linked to the place	Given cue (Genre), name of the place, country of the place, year of the memory, type of event in the memory
No Place Cue	Name of the place, country of the place, year of the memory, type of event in the memory
Artist associated with a person	Given cue (Artist name), year of birth of the person + 23, type of event in the memory
Artist associated with a place	Given cue (Artist name), year of the memory, type of event in the memory

These user inputs form structured scenarios, summarized in Table 1.

3.2.2. Mood Extraction

Once memory cues are collected, users select their mood using the affect button (Broekens and Brinkman, 2009), an interface displaying a smiley face to express their emotional state during the memory (mood at encoding).

The affect button maps selections to valence, arousal, and dominance values ranging from −1 to 1. These values are then passed to the music recommender system to ensure song recommendations align with the user's recalled emotional experience.

3.3. Music recommendation module

This module selects songs based on extracted cues and generates missing cues when needed.

Following predefined scenarios, it generates a list of 10 recommended songs by prompting an LLM (GPT-4.0, selected for its strong performance in pilot tests). The prompt includes features derived from memory cues, as detailed in Table 2.

In Table 2, the "type of event in the memory" refers only to a general category (e.g., birthday, gym) rather than a full memory description, preserving user privacy by limiting personal data shared with the LLM.

To determine a target year for song selection, we use "year of birth + 23" for memories involving people, based on findings that music preferences peak at age 23 (Davies et al., 2022). For other cues, the target year is set to when the user was 20, aligning with the reminiscence bump. For users younger than 20 years old, the system chooses songs from its knowledge cutoff year (which depends on the chosen LLM). In our case this cutoff year was 2024.

After compiling features, the system prompts GPT-4.0 using the chain-of-thought method (Wei et al., 2024) to generate two lists of 10 songs—one based on a People cue and another on a Place cue. If an artist is specified, recommendations come exclusively from that artist's catalog. If only a place is provided, the LLM prioritizes songs popular in that location during the target year, increasing familiarity. Specific prompts are detailed in Appendix B.

To refine recommendations, the system analyzes the mood of each track using the Spotify API (Spotify Ltd., 2024). All songs are ranked by their Euclidean distance from the memory's mood. The best match from each scenario (person and place) is selected for recommendation.

The system operates without a memory feature, as this functionality is disabled in the LLM; consequently, each event is treated independently in generating recommendations. Future developments could enhance performance by incorporating long-term interaction modeling, where memory plays a central role.

The Spotify API's valence and energy metrics are used as proxies for valence and arousal, as done in previous research (Panda et al., 2021). Although alternative models such as Essentia (Alonso-Jiménez et al., 2020) were considered, real-time processing of full-song imports was not feasible due to speed and fair-use limitations.

To prevent hallucinations, all generated songs are verified against Spotify. If a song does not exist, it is removed. If more than half of the songs are unrecognized—common with niche artists—the system selects tracks from the artist's top Spotify songs instead.

The user receives two final recommendations: one based on the People cue and one on the Place cue. Each includes a 30-second preview or a full song link from Spotify or Deezer (Deezer SA, 2024), depending on availability. These platforms together comprise a catalog of approximately 190 million licensed tracks (about 100 M on Spotify and 90 M on Deezer). If the user approves a recommendation, it is stored in the Memory Ontology Module for future use.

3.4. Memory ontology module

The memory ontology is implemented as a Neo4j graph database (Neo4j, Inc., 2024), with individual files for each user. It enables the replay of songs linked to specific memories and serves as a database representing user memories. The ontology structure builds on that by Lim et al. (2013), emphasizing predefined place and people cues.

Its superclasses include Memory, People and Place. The Memory Class stores user-selected memories, descriptions, associated moods, chosen place and People cues, and linked songs. If no Place or People cue is provided, the song's artist is added as a direct cue. The People Class represents mentioned individuals along with their birthdates for future recommendations. The Place Class represents referenced locations.

Previously mentioned entities in the CoMEEM recommender system are matched with new mentions, merging identical places or people into a unified graph representation.

An example of a memory ontology graph is shown in Fig. 4.

This memory ontology is crucial for retaining the user's interactions and can serve as a foundation for future learning mechanisms, providing prior knowledge to enhance song recommendations.

4. Experimental method

The CoMEEM recommender system, integrating the previously described modules, was tested in an experiment designed to assess its

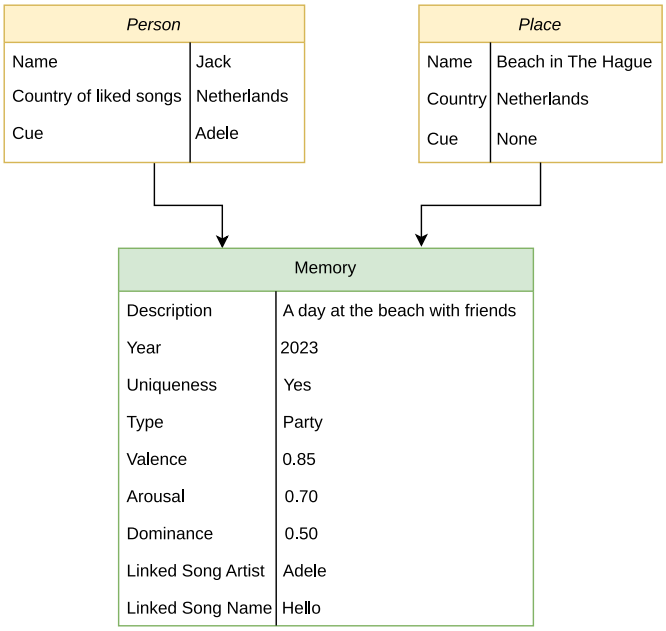


Fig. 4. Memory Ontology Example. The memory is represented through its related Person and Place Classes.

effectiveness and validate the hypotheses derived from the research questions.

4.1. Participants

Seventy-one participants (32 men, 38 women, 1 non-binary/other) were recruited across three age groups: 18–25 (27 participants, M = 23.4, SD = 1.85), 40–60 (22 participants, M = 53, SD = 5.65), and 65+ (22 participants, M = 74.8, SD = 9.2). They were drawn from France (24) and the Netherlands (47) via personal and university networks. This controlled sample enabled a systematic analysis of age-related differences in music recommendation responses. Age groups were defined to maximize within-group homogeneity while maintaining sufficient separation between groups to capture meaningful differences across distinct life stages, which would not be possible with a single continuous age scale. The 18–25 band follows the emerging adulthood literature that identifies ages 18–25 as a distinct developmental phase (Arnett, 2000). Middle adulthood (40–60) reflects commonly used definitions of midlife in lifespan psychology (Lachman et al., 2014), while 65+ aligns with public-health and demographic conventions for older adults (World Health Organization, 2015). The experiment was preregistered on OSF (Raingeard De La Bletiere et al., 2024) and approved by the Delft University of Technology's Ethics Committee (approval numbers *4122* and *4967*).

4.2. Design

The CoMEEM system was deployed as a web-based platform, accessible to all participants. It introduced two additional recommendations featuring the most mood-incongruent songs from the LLM-generated list of ten, ensuring a controlled contrast between mood-congruent and mood-incongruent choices.

The experiment included four recommendation conditions: (1) two based on people associated with the memory (most and least mood-congruent songs) and (2) two based on place (most and least mood-congruent songs). This design allowed for a systematic evaluation of mood congruence in perceived song-memory fit while incorporating both episodic memory cues.

The study was conducted in France and the Netherlands, in both French and Dutch.

4.3. Measures

4.3.1. Main measures

The primary outcome measure was perceived song-memory fit, assessed on a scale from 1 to 100 ("How appropriate is this song to form an association with your memory?"). We classify scores above 75 as an arguably good fit.

Secondary outcome measures evaluated fit across four episodic memory components: event, people, place, and year. Participants also rated the influence of ten predefined factors on song-memory fit using a scale from −50 (negative impact) to +50 (positive impact).

These factors, derived from prior research (Li and Hu, 2023; Panda et al., 2023), excluded technical elements like harmony, which were not understood by participants. Mood, familiarity, and preference were adapted from Music-Evoked Autobiographical Memories (MEAMs) studies (Jäncke, 2008; Belfi et al., 2020; Jakubowski and Francini, 2022), while artist, genre, and lyrics were included based on music recommendation perception studies (Kamehkhosh and Jannach, 2017). Participants could also introduce additional relevant categories.

The final set of factors included: (1) Song characteristics: Tempo, valence, energy, instrumentation, and lyrics; (2) Other factors: Familiarity, genre, artist, year of release, and personal preference; (3) User-defined categories, if applicable.

To further assess familiarity, participants answered two questions from Jakubowski and Francini (2022): "How familiar are you with the genre of this song?" and "Have you heard this specific song before?"

To assess the system's ability to match songs to memory-evoked moods, participants reported their emotions while listening using the affect button (Broekens and Brinkman, 2009), capturing valence, arousal, and dominance scores.

4.3.2. Control measures

Participants provided demographic data (age, gender, education level) alongside measures assessing individual differences that may influence responses: (1) Musical preferences, assessed using the Short Test of Music Preferences (STOMP) questionnaire (Rentfrow and Gosling, 2003). The STOMP categories were also used to structure the genre selection process during the interaction; (2) Perceived importance of music in daily life, rated from 1 to 100; (3) Frequency of Music-Evoked Autobiographical Memories (MEAMs), as well as life periods most associated with MEAMs, measured with a 7-point Likert scale; (4) Self-reported frequency of memory problems and frequency of experiencing depressed mood episodes, also measured through a 7-point Likert scale.

4.3.3. Qualitative evaluation

To refine the CoMEEM recommendation system, qualitative feedback was collected. Participants could comment on whether the recommended songs evoked different memories and suggest improvements.

After receiving all four recommendations, they reflected on: (1) The personal significance of the chosen memory; (2) The perceived effectiveness of the recommendations, particularly in identifying a suitable song for the memory; (3) Suggestions for improving the system's interaction and recommendation process.

This qualitative data offers insights into users' memory selection patterns and informs iterative system enhancements.

The full multilingual questionnaire is provided in Appendix C.

4.4. Procedure

Each participant accessed the experiment website, providing one to two episodic memories, each linked to four song recommendations, resulting in 87 memories and 348 recommended songs. These songs form a necessarily small subset of the full Spotify/Deezer catalogs (190 M tracks). They were drawn through the memory-elicitation process, ensuring diversity across genres and decades.

The experimental procedure followed a structured sequence. After accessing the website and completing an informed consent form, an initial questionnaire was administered to gather demographic and control measures. Next, the dialogue management module guided participants in selecting a memory, identifying episodic memory cues, and specifying its associated mood. After this, the recommendation module generated song suggestions based on these inputs, and participants received four randomly ordered recommendations, and were asked to rate their perceived fit with the memory, and evaluate song characteristics influencing the match. Finally, the highest-rated song was added to the participant's Memory Ontology Module, and a final questionnaire gathered qualitative feedback on improving the recommendation process.

Participants could complete the procedure independently online or request researcher assistance if needed.

4.5. Data analysis

We tested four main hypotheses (H1a–H3) regarding the factors influencing the perceived fit of songs to episodic memories.

- H1a (RQ1): Familiarity has minimal influence on perceived fit, contrary to the case of standard MEAMs.
- H1b (RQ1): Age and personal evaluation of songs are highly predictive of the fit of a song to a specific memory.
- H2 (RQ2): Songs related to People cues are generally associated with a higher fit to a memory compared to songs related to Place cues.
- H3 (RQ3): The alignment between song-evoked and memory-evoked mood increases the perceived fit of a song to a memory.

We used multilevel modeling to identify key predictors of song–memory fit, focusing on H1a, H1b, and H3. The model evaluated ten candidate predictors—such as valence, arousal, tempo, lyrics, genre, and familiarity—along with control variables including age group (discretised from 0 to 2 to account for differences in the analysis), gender, country, depressive symptoms, memory issues, and the personal importance of music. Mood matching variables were computed as the Euclidean distance between memory and song evaluations along the valence, arousal, and dominance dimensions. Predictors were added stepwise, retaining only those that reduced the Bayesian Information Criterion (BIC). Significance was assessed using p-values corrected with the False Discovery Rate (FDR) procedure to control false positives (Benjamini and Hochberg, 1995). Effect sizes were assessed using the f^2 metric (Aiken and West, 1991; Lorah, 2018), where $f^2 = \frac{R_2^2 - R_1^2}{1 - R_2^2}$ and R_2^2 denotes variance explained with a factor included.

In parallel, statistical learning methods were used to evaluate how well these same features could predict fit scores. We trained a support vector machine, random forest classifier, and gradient boosting regressor to assess predictive accuracy under different modeling conditions. Incrementally introducing predictors—particularly familiarity, genre, and mood—allowed us to comprehensively test H1a and H3, while exploring H1b through variability in model performance across individuals.

To directly assess H2, a one-way ANOVA was conducted on fit ratings, comparing songs linked to people versus those linked to places. Additional scenario-based analyses were used to examine how different cue interpretations influenced perceived fit, providing further evidence for H2 and clarifying participant associations between memory cues and specific music features.

Supplementary analyses provided additional context. System performance was evaluated by comparing the proportion of high-fit recommendations to earlier findings on Music-Evoked Autobiographical Memories (MEAMs) (Janata et al., 2007). Participant feedback was thematically analyzed to identify improvement opportunities and subjective experiences related to fit judgments. Finally, we categorized the types of memory events participants selected and explored how these thematic contexts related to fit outcomes, offering further insight into H1b and the variability of fit perception across individuals.

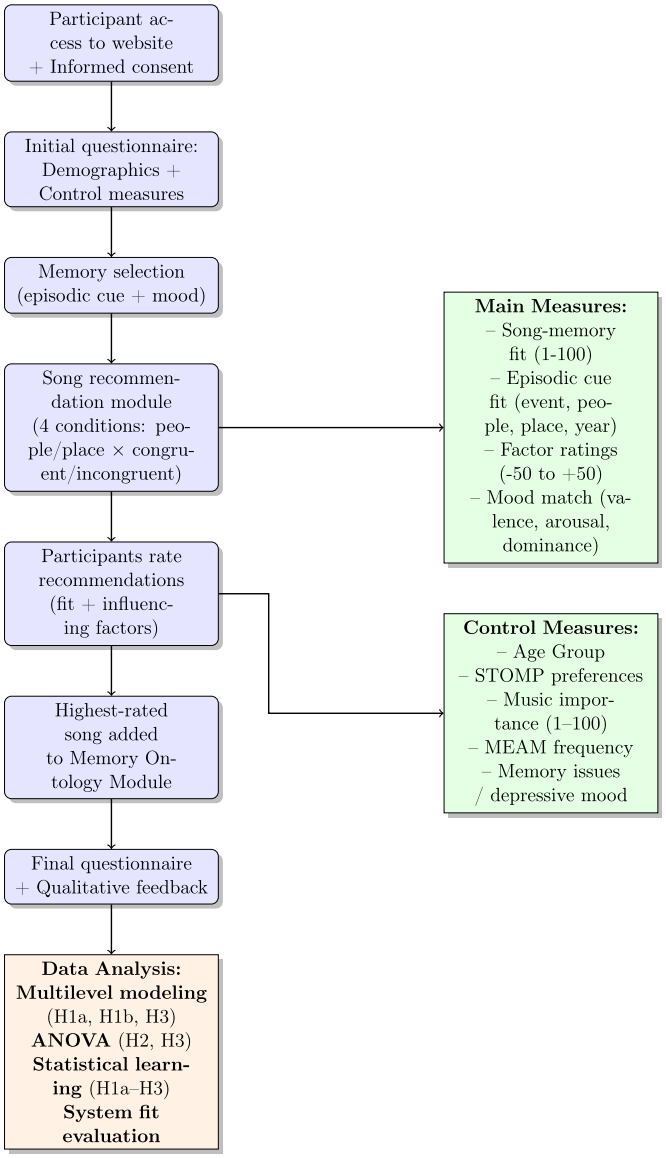


Fig. 5. Overview of the experimental procedure in the CoMEEM study.

Qualitative responses were translated using the Seamless M4T model (Loïc Barrault, 2023) and manually validated by two native Dutch speakers and one native French speaker to ensure accuracy. Raw data are available in a restricted 4TU repository (Raingeard de la Bletiere et al., 2025b), while a separate open dataset contains anonymized, non-personal data (Raingeard de la Bletiere et al., 2025a). An overview of the experimental procedure, variables, and data collection flow is presented in Fig. 5.

5. Results

5.1. Assessing our predictors

5.1.1. Multilevel modeling

We first examined how song and memory features predicted perceived fit using a multilevel model, with the goal of evaluating Hypotheses H1a, H1b, and H3. Predictors included valence, arousal, familiarity, tempo, lyrics, genre, and mood matching (valence from the affect button). Their contributions are summarized in Table 3. The effects of the predictors are summarized in Fig. 6.

Table 3
Predictors of the multilevel model with standardized beta coefficients (β), standard errors (SE), unadjusted and FDR-corrected p-values.

Predictor	β	SE	Unadjusted p	FDR-corrected p
Valence	0.21	0.06	0.0002	0.0004
Arousal	0.16	0.06	0.007	0.008
Familiarity	0.27	0.05	5.94×10^{-9}	4.75×10^{-8}
Tempo	0.19	0.06	0.0007	0.001
Lyrics	0.21	0.05	5.05×10^{-5}	0.0001
Genre	0.27	0.05	1.26×10^{-7}	5.02×10^{-7}
Mood Matching Valence (Affect Button)	−0.17	0.05	0.0009	0.001

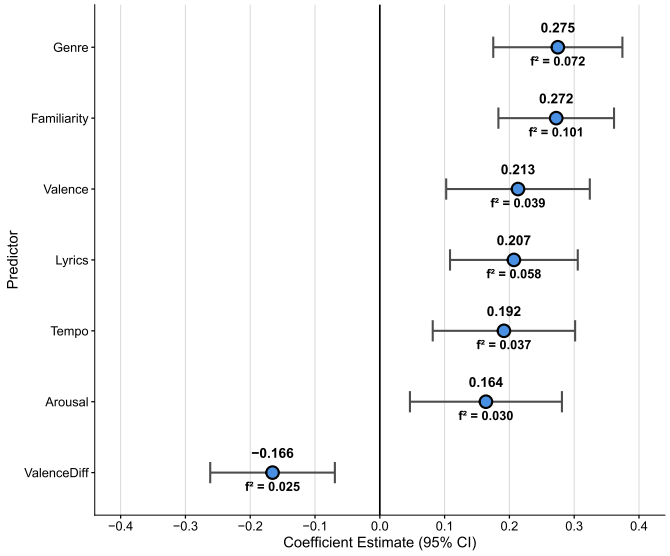


Fig. 6. Fixed Effects of the Multilevel Model. Estimated effects are displayed with 95 % confidence intervals. All retained predictors have confidence intervals that do not cross zero, indicating statistically reliable effects.

Effect size analysis clarified the relative importance of each predictor. Familiarity ($f^2 = 0.10$) and genre ($f^2 = 0.07$) showed the largest effects, indicating small to medium impacts, while valence, arousal, tempo, and mood matching each had small effects. These findings contradict H1a: instead of having minimal influence, familiarity emerged as one of the strongest predictors of fit. Similarly, the effect of mood congruence (valence matching) partially supports H3: valence alignment between song and memory influenced fit, though effects were smaller than those of other predictors.

The intraclass correlation coefficient (ICC = 0.21) indicated a moderate influence of individual differences, and age was not a significant predictor of fit, meaning H1b was not supported. Personal evaluations contributed moderately, but demographic age did not. The model's fitted R^2 of 0.74 indicates a strong fit to the data.

These findings highlight familiarity and genre as key factors for well-fitting song recommendations, rejecting H1a, alongside other musical features. Notably, familiarity influenced perceived fit to the memory more strongly than prior knowledge of the song. Assessing collinearity, we found a low correlation between perceived familiarity and song or genre recognition (0.28 and 0.29, respectively), suggesting that while previous exposure plays a role in higher ratings (see Fig. 7), its effect is not solely due to prior exposure, but to the perceived usefulness of familiarity.

5.1.2. Machine learning prediction

We evaluated three machine learning models—Random Forest, Support Vector Machine (SVM), and Gradient Boosting—to predict fit scores. Model performance was assessed under three conditions:

- Continuous fit value prediction, where predictions were considered accurate if they fell within 15 points of the actual score. This threshold corresponds to approximately half a standard deviation of observed ratings (SD ≈ 30), which is commonly considered a meaningful difference in subjective scales, following Cohen's effect size conventions (Cohen, 2013) and the "half-SD rule" in psychometrics (Norman et al., 2003).
- Three-category classification of fit (low, medium, high).
- Four-category classification of fit (low, medium-low, medium-high, high).

For the continuous predictions, we used K-Fold cross-validation (5 folds). For the categorical predictions, we used Stratified K-Fold cross-validation (5 folds) to ensure class proportions were preserved in each fold and to provide robust estimates of model performance. Feature scaling was applied for SVM via a standardization pipeline, while tree-based models could handle predictors.

The results for the Mean Square Error of classifiers in the continuous scenario are summarized in Table 4, and the accuracies are summarized in Table 5.

The dataset included a large number of predictors with relatively small effect sizes and a hierarchical structure. Such characteristics tend to dilute the signal-to-noise ratio, making it difficult for SVM models to identify consistent separating boundaries. In contrast, Random Forests are more robust to noisy or diffuse signals, which likely contributed to their performance.

In this case, the predictors that increased accuracy when introducing them gradually are the highlighted valence and arousal, the highlighted familiarity and genre, the lyrics, and the distance between the mood of the song and the memory on the three axes (valence, arousal, and dominance). The strong contribution of familiarity contradicts H1a, which proposed that familiarity would have minimal influence on fit. Additionally, the impact of mood distance supports H3. The lack of contribution from age-related factors indicates H1b is not supported, consistent with multilevel results.

Although the models outperformed random guessing, they fell short of zero-shot learning accuracy. This suggests that incorporating a feedback loop could improve the system's recommendation accuracy over time.

5.2. Scenario analyses

Through one-way ANOVAs, we find that songs related to people received higher average fit scores than those related to places, with a statistically significant difference and a small to medium effect size, $F(1, 348) = 5.843, p = .02, \eta^2 = .016$. This finding supports H2: person-based memory cues enhance perceived fit more strongly than place-based cues.

To test H3, which predicted that mood-congruent music would increase fit, we compared mood-matching and non-mood-matching conditions. The difference was not significant, $F(1, 348) = 1.17, p = .28, \eta^2 = .003$, suggesting no reliable support for H3 in this analysis. Standard deviations across both analyses were SD = 33.7. The fit values for each condition are summarized in Fig. 8(a).

This result contradicts the findings for H3 obtained from the multilevel model, which may be explained by individual-level variability that the ANOVA did not account for, as it did not adjust for within-person or within-memory differences.

Scenario-level analyses showed that participants often linked a specific artist to the person in their memory, but rarely formed strong associations with places. This further reinforces the idea that person-based memory cues tend to elicit more meaningful or concrete associations than place-based ones, in line with H2. No significant differences were observed between different scenario types and the fit of resulting songs, but the pattern of responses in people-based scenarios suggests more frequent alignment with music-related attributes. The values of fit depending on scenarios are summarized in Figs. 8(b) and (c).

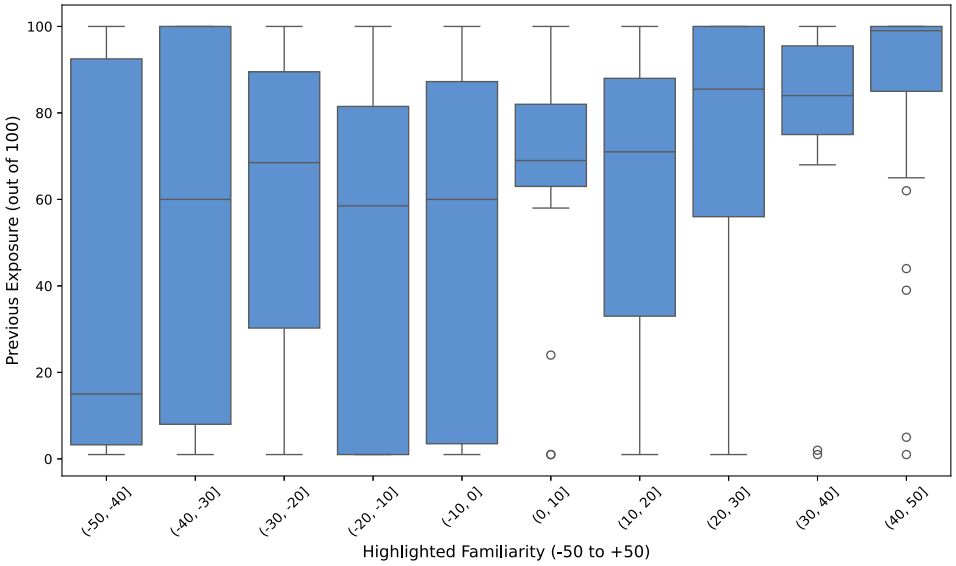


Fig. 7. Familiarity vs Previous exposure of songs. This graph shows a relationship between previous exposure and perceived influence of familiarity on the fit only for positive perceived influence by the user.

Table 4
Table of the Mean Absolute Error in the continuous scenario.

	SVM	Random forest	Gradient boosting
Mean Absolute Error	17.9	14.3	14.8
Mean Square Error	534	500	516

Table 5
Table of the accuracy in all 3 scenarios.

Accuracy	SVM	Random forest	Gradient boosting
Continuous	0.36	0.61	0.60
3 Categories	0.70	0.67	0.66
4 Categories	0.58	0.52	0.51

5.3. Recommendation performance

The system's performance, measured by overall fit scores, averaged 46.9 with a median of 54.0 and a standard deviation of 33.65, indicating a skew toward extreme values, particularly lower ratings. However, on average, 51.2 % of recommendations received scores above 50/100 (SD = 14.72), and 25 % exceeded 75/100 (SD = 8.16). The full distribution of ratings is shown in Fig. 9.

When considering only the highest-rated song per memory, the mean fit score increased to 77.2 (CI [72.4, 82.0], SD = 22.4), suggesting that most participants received at least one song they considered a strong match for their memory.

Among participants who did not receive a strong match (fit score < 75), 46 % identified a specific song they believed would have been a better fit, while 88 % suggested either a specific song or a genre they found more suitable. These findings highlight the potential benefit of integrating a feedback loop into the system, and indicate that participants valued more familiar songs in these cases (H1a).

Additionally, participants generally developed a clearer sense of what type of song best suited their memory. In the final part of the study, participants rated whether they had a better idea of what type of song would match their memory through our recommendations on a slider scale, resulting in an average score of 70 and a median of 76, suggesting that most found the process informative and exploratory.

Hallucinations were infrequent: among the 87 memories, only 5 contained more than 5 hallucinated tracks out of 10. These cases occurred when the system associated a recommendation with an artist not recognized in the database, such as orchestras or secondary contributors not listed as primary track artists. A recovery strategy, querying the artist on Spotify and retrieving their top 10 tracks, proved effective. Future work could integrate such Spotify insights directly into the LLM prompting process.

5.4. Qualitative analysis

Participants most commonly described memories related to family events ($N = 21$), events with friends ($N = 20$), holidays ($N = 17$), and love or romantic relationships ($N = 13$). Less frequent but still notable categories included work or study ($N = 9$), daily life activities ($N = 9$), and sports ($N = 4$). Several memories spanned multiple categories—for example, a holiday involving both friends and family. Across the sample, memories were generally rated as emotionally meaningful, with an average importance rating of 74 ($SD = 23$).

To explore factors that may influence high perceived fit, we examined the participants who assigned the highest average fit ratings (defined as scores above 90, $N = 22$). No strong individual differences such as age, musical preferences, or the general importance of music in daily life emerged to explain these high scores, coherent with the analyses of the multilevel and predictive models, showing further evidence against H1b: personal evaluations mattered, but age and broader demographics did not.

However, memory type did appear to influence fit outcomes.Work- and education-related memories were overrepresented among high-fit cases (25 % vs. 15 % in the full dataset). In contrast, memories tied to family and relationships were underrepresented (10 % of high-fit memories vs. 20 % overall). These patterns suggest that the specificity and flexibility of memory cues play a role in shaping perceived fit, which extends H2: while people cues predicted higher fit than place cues, broader contextual factors (e.g., work/study memories) also influenced fit. Participants often reported already having specific songs or genres in mind for family- or relationship-related memories, narrowing the range of potential recommendations. In contrast, work and education memories provided broader interpretive space, which may have facilitated stronger matches when algorithmic suggestions aligned.

Additional qualitative feedback was provided by 37 participants. Many emphasized the importance of emotional resonance in recommendations (H3), with $N = 8$ requesting more space to describe their emotions and $N = 12$ asking for the ability to elaborate on contextual

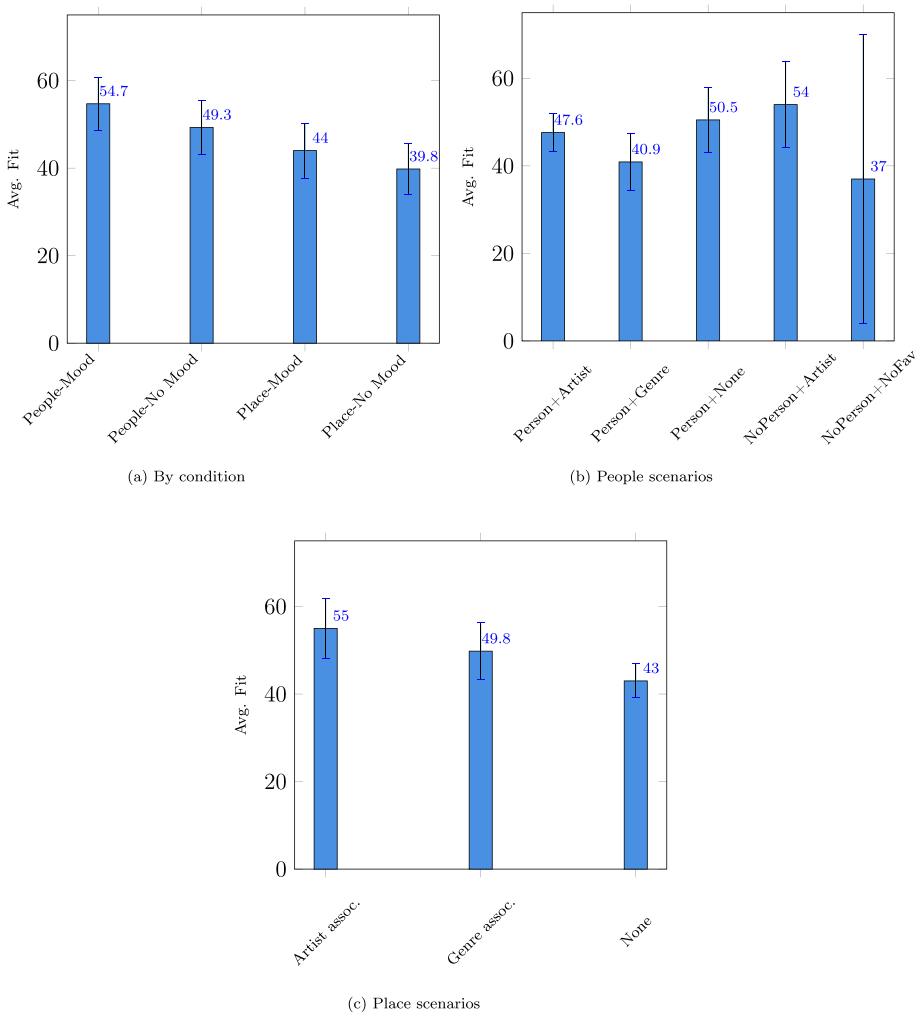


Fig. 8. Average Fit Values across conditions: (a) overall condition comparison, (b) people scenarios, and (c) place scenarios. Values are displayed above the corresponding error bars.

details of the memory. A notable portion ($N = 13$) expressed a desire for recommendations that better matched their favorite songs or artists, indicating that familiarity remains a salient factor for many users, further contradicting H1a's prediction of minimal influence.

6. Discussion

6.1. Summary of findings

This study introduced a Constructed Music-Evoked Episodic Memory (CoMEEM) recommender system, designed to link songs to specific personal memories. We tested four hypotheses: (H1a, RQ1) familiarity would play only a minimal role in perceived fit, (H1b, RQ1) age and personal evaluations would predict fit, (H2, RQ2) people-based cues would yield higher fit than place-based cues, and (H3, RQ3) mood congruence would increase perceived fit.

6.1.1. RQ1: What subjective factors predict the perceived fit of a song to a specific episodic memory?

Our findings contradict H1a which was not supported. Across all methods, familiarity consistently emerged as a major predictor of fit. Both statistical models and machine learning analyses showed that familiarity explained variance above and beyond other factors, and participants frequently mentioned familiarity in qualitative feedback. Importantly, the effect was not limited to prior exposure but reflected the perceived usefulness of familiarity in assessing fit, extending prior MEAM research (Janata et al., 2007; Kathios et al., 2024).

H1b was also not supported. Unlike studies identifying age-related MEAM retrieval patterns (Mehl et al., 2024), our findings suggest CoMEEM-based fit is age-independent. Age was not predictive of fit in any analysis, while personal evaluations contributed moderately. Individual differences accounted for some variability, but demographic age alone did not play a significant role. Additionally, the lack of a significant likeability effect contradicts previous MEAM research emphasizing its role in recall (Jakubowski and Francini, 2022). Though no single factor had a strong effect size, our multilevel model confirms that these factors collectively shape perceived fit, reinforcing MEAM perspectives on the interplay of emotion, familiarity, and personal preferences rather than demographics (Jakubowski and Francini, 2022).

6.1.2. RQ2: Which type of episodic memory cues enhances the fit of a song to a memory?

H2 was supported. Songs tied to people were perceived as a better fit than those tied to places, in line with previous work highlighting the stronger recall potential of social versus spatial cues (Lee and Dey, 2007; Baumann et al., 2024). Broader contextual categories (e.g., work- and education-related memories) further reinforced the importance of social and situational context. Participants often associated artists with people in their memories, whereas such associations were rare for places. This

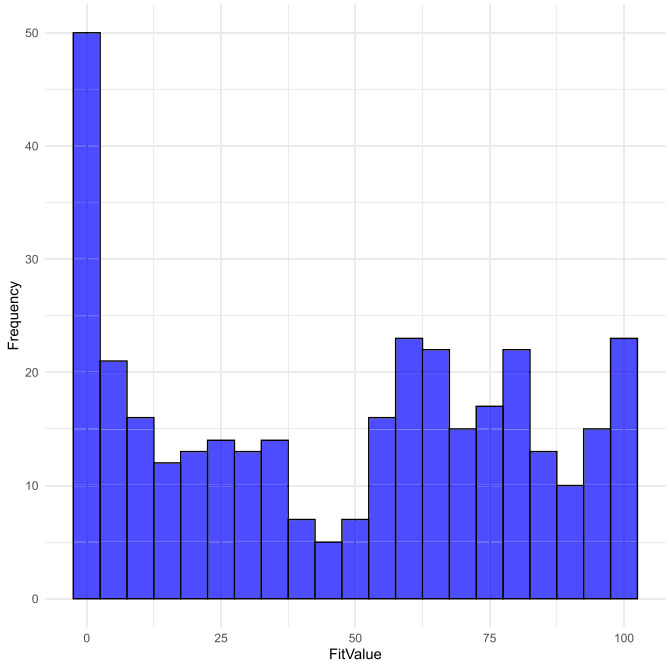


Fig. 9. Histogram of Fit Ratings. This graph indicates a skew towards extreme negative values.

reinforces the idea that person-based cues provide richer anchors for song–memory connections.

Future research should balance memory scenario distributions, as some were more common than others. Additionally, incorporating object or action cues, commonly used for recall (Lee and Dey, 2007), may help identify more diverse songs, potentially improving the fit.

6.1.3. RQ3: Does the congruence between song-evoked and memory-evoked mood influence fit?

H3 received partial support. Valence alignment between song and memory increased perceived fit, consistent with mood-congruent memory theory (Faul and LaBar, 2023; Lewis et al., 2005). Arousal and dominance congruence were weak or inconsistent, appearing only as minor contributors in machine learning models. This aligns with prior findings suggesting that arousal affects memory specificity rather than perceived fit (Talarico et al., 2004; Mirandola and Toffalini, 2016), and that dominance is rarely a meaningful dimension of musical mood (Brinker et al., 2012).

6.1.4. System performance

System performance was moderate: approximately 25 % of recommendations achieved high-fit levels. While this is lower than optimal MEAM studies, which often maximize memory recall by pre-selecting familiar songs (Kathios et al., 2024), our system's novelty lies in recommending music for specific memories rather than eliciting any memory (Janata et al., 2007). This specificity has potential applications in music therapy, stress recovery, or everyday reflective practices.

Participants also highlighted the importance of familiarity and emotional resonance in fit judgments. Incorporating participant feedback into future iterations, such as allowing users to refine recommendations based on favorite artists or providing richer emotional context, could substantially improve system performance.

6.2. Design implications

Based on our empirical findings, we propose an operational framework for designing music recommendation systems that support associations between songs and episodic memories. Our framework is presented in Fig. 10.

The framework is grounded in three core principles, each explicitly tied to the hypotheses and results of the study:

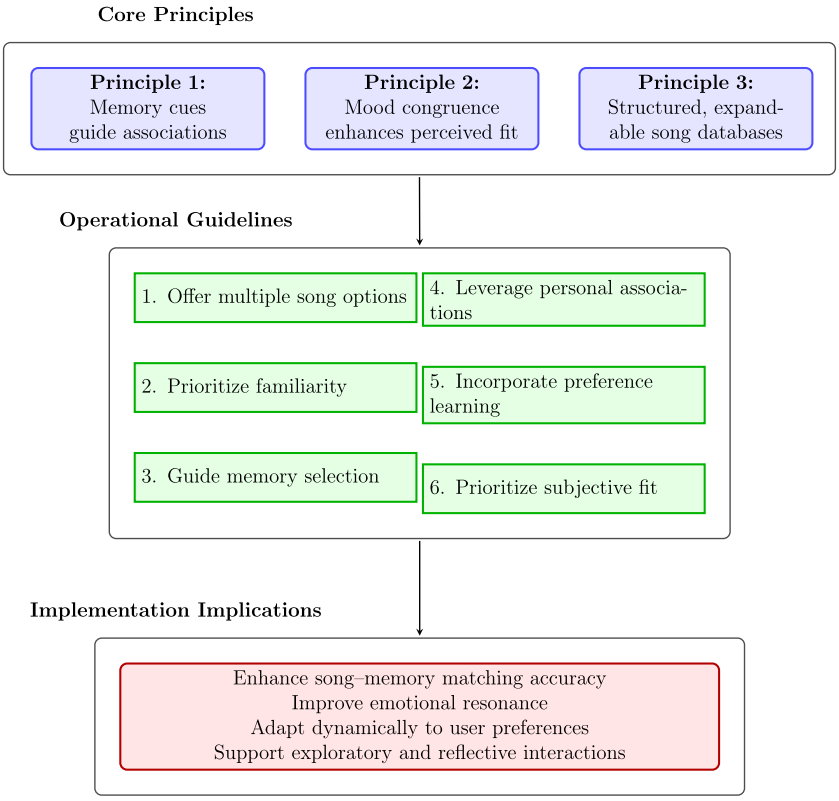


Fig. 10. Proposed operational guidelines and principles.

1. **Structured, expandable song databases support personalization (H1a/RQ1):** Maintaining a growing and formally structured database of songs, annotated with memory-relevant cues, allows systems to generate more precise and personalized recommendations over time. Familiarity and genre emerged as major predictors of fit, supporting the need for structured song metadata.
2. **Memory cues guide associations (H2/RQ2):** Cues from episodic memory elements, particularly People cues, enable users to link songs with personal memories effectively. This principle is supported by our findings which show that songs linked to People cues were consistently rated as better fits than songs linked to Place cues.
3. **Mood congruence enhances perceived fit (H3/RQ3):** Aligning the valence of a song with the valence of the memory improves perceived fit. While arousal congruence was less consistently influential, valence alignment consistently supported stronger memory-song associations.

These principles translate into six operational design guidelines, summarized in Table 6 and Fig. 10, which can inform both system architecture and interaction design.

Implementation Implications: By following these operational principles, future CoMEEM systems can:

- Enhance song–memory matching accuracy.
- Improve emotional resonance through valence-aligned recommendations.
- Adapt dynamically to individual user preferences while maintaining system flexibility.
- Support exploratory and reflective interactions with personal memories.

6.3. Limitations

6.3.1. Technical limitations

The Spotify API was unreliable in aligning songs with participants' emotional experiences (Valence MAE = 0.73, MSE = 0.79; Arousal MAE = 0.72, MSE = 0.78). Because the H3 analysis uses self-rated mood, not the Spotify valence estimates, the unreliability of the API does not propagate into the H3 results. However, as mood matching significantly predicted perceived fit, alternative mood classification methods should be explored.

Previous research (Panda et al., 2021) also reported Spotify's mood classification limitations, though our study found even greater discrepancies. This may stem from participants linking mood to memories before answering other questions or limitations in the affect button, despite its validation (Broekens and Brinkman, 2013). This has been shown to already be the case in previous studies, where there was an effect of activated memory on emotional state when looking at and listening to multimedia and videos (Dudzik et al., 2020). To improve accuracy, human mood input should complement algorithmic sorting.

Participants followed a rule-based scenario, answering multiple questions before receiving recommendations. While effective for data collection, this approach may be too demanding for individuals with cognitive impairments. To enhance accessibility, the system should be simplified by reducing the number of required responses to two or three and adopting a conversational interface over a rigid rule-based structure.

6.3.2. Methodology limitations

Although participants only interacted with 340 tracks, this subset emerged naturally through the memory-elicitation process and spans diverse genres, eras, and affective contexts. The CoMEEM system itself is not limited to this subset: it interfaces with the entire Spotify and Deezer catalogs (around 190 million tracks in total). Thus, while our study necessarily involved a manageable subset for evaluation, the mechanisms for dialogue, feature extraction, and recommendation scale to the full library. Moreover, participants were recruited across multiple age groups, capturing inter-generational variation in music–memory associations. While recruitment was conducted in only 2 countries, the cognitive mechanisms under study (autobiographical memory recall and emotional linkage to music) are not culturally specific, supporting the plausibility of broader generalization.

Most participants (88 %) identified a better-fitting song or genre after listening to recommendations, highlighting the need for a feedback mechanism to refine suggestions based on user preferences.

The system assumes memories are linked to a specific person, but some participants associated their memories with groups. Expanding the system to accommodate collective experiences and shared musical preferences would improve relevance.

Two participants encountered familiar songs that evoked negative memories. To mitigate this, future versions should allow users to flag or

Table 6
Operational guidelines for CoMEEM-based music recommendation systems. Each guideline operationalizes the core principles identified in the study.

#	Operational guideline
1	**Offer multiple song options.** Presenting users with a curated set of candidate songs enables exploratory interaction and user-driven selection of the best-fitting track. This approach enhances agency, consistent with designs in serious games for older adults and people with dementia (Agres et al., 2019; Tanner et al., 2023), and in music therapy and social robot systems (Samson et al., 2025). Systems could implement this by offering the top-3 valence-matched tracks ranked by cue relevance.
2	**Prioritize familiarity.** Even though CoMEEMs rely less on spontaneous associations than MEAMs, familiar songs increase perceived safety and emotional resonance, especially in older adults (Canapa et al., 2025; Baur et al., 2018). Systems could query a user's listening history or request known "safe" songs to avoid distressing associations (Rao et al., 2021; Mou et al., 2021).
3	**Guide memory selection.** Structured prompts (e.g., "Think of a time with a close friend" or "Recall a meaningful place from your youth") help users retrieve relevant memories. Similar cue-based approaches are used in cognitive training games (Agres et al., 2019) and reminiscence therapy apps (Samson et al., 2025). Systems could implement guided memory selection through chatbot-style interactions or card-based UIs.
4	**Leverage personal associations beyond memories.** Users often associate songs with people even without direct event links. Social cues are strong drivers of recall (Janata et al., 2007; Baumann et al., 2024). Systems could use known social data (e.g., family members or caregivers) or invite users to tag songs with social metadata. In interactions, the agent might ask follow-ups such as "Does this song remind you of someone?"
5	**Incorporate preference learning.** Tracking user preferences prevents mismatches and increases system adaptability over time (Kim et al., 2024; Wu et al., 2024). Systems should log explicit ratings and implicit signals (e.g., skipped songs, facial expressions) to learn iteratively. This aligns with hybrid recommender approaches in therapeutic music applications (Fraile et al., 2019).
6	**Prioritize subjective fit over objective metrics.** Emotional and memory fit is subjective. Our findings, consistent with Jakubowski and Francini (2022); Mou et al. (2021), show that users' own ratings predict song-memory fit better than audio features. Systems should prioritize subjective inputs (e.g., "Does this song feel right for the memory?") over audio similarity alone, using critique-based feedback loops in LLM-based recommenders (Cai et al., 2021; Jin et al., 2019).

exclude songs before listening. The system could then learn to avoid such songs in future recommendations. Notably, these instances did not prevent participants from identifying suitable songs among other recommendations.

Lyrical content, a key factor in song fit, was not considered. Future enhancements should integrate lyric analysis tools to align lyrics with memory themes and ensure context-aware filtering for emotional consistency.

Finally, incorporating action cues could refine recommendations by treating the "type of event in the memory" (see Section 3.3) as a cue associated with specific artists and genres.

6.4. Future work

Future work should advance the system along several dimensions. A key priority is the integration of a feedback loop to enable adaptive recommendations that evolve with user input over time. Expanding the range of memory cues beyond people and places to include objects and actions could enrich the contextual basis for recommendations. In parallel, incorporating lyric analysis and other content-based features would enhance context-aware song matching. Another important direction is the design of simplified, cognitively accessible interfaces that preserve personalization while ensuring usability for diverse populations. Extending evaluations to larger and more cross-cultural participant groups will be essential to validate the system's generalizability. Furthermore, embedding memory within the LLM framework would allow the recommender to accumulate experience, progressively refining its personalization strategies.

Such developments would be particularly valuable in dementia care, where music-evoked memories strongly influence mood. They could also support human–robot interaction systems, enabling robots to personalize user experiences in moments of stress—for instance, by targeting songs linked to upcoming events or anticipated visits from family members. Finally, collaboration with music therapists could guide the personalization process and facilitate the reuse of therapeutic songs across multiple sessions.

7. Conclusion

This study explored a music recommendation system designed to evoke Constructed Music-Evoked Episodic Memories (CoMEEMs) and its implications for assistive and interactive technologies. The system's effectiveness, measured by perceived song-memory fit, demonstrated that most participants identified at least one strong match, highlighting its potential for enhancing targeted memory recall.

Our findings reveal key predictors of fit, emphasizing subjective factors in song-memory associations. Familiarity emerged as the strongest predictor, contrary to our initial hypothesis. However, this effect stemmed from the perceived usefulness of familiarity rather than prior exposure, distinguishing CoMEEMs from MEAMs. Individual differences such as age did not significantly influence fit, while genre, valence, arousal, and lyrics played a role, though none dominated. This suggests that fit arises from an interplay of multiple subjective factors.

For episodic memory cues, songs linked to People cues were perceived as a better fit than those associated with Place cues, supporting our hypothesis. Participants frequently connected artists to individuals in their memories, whereas such associations were weaker for locations.

Mood congruence findings confirm that valence alignment between a song and memory enhances fit. While arousal and dominance congruence were not significant in multilevel models, their effects improved learning model accuracy, suggesting small but meaningful contributions.

Future work will focus on adaptive learning mechanisms to refine recommendations based on user feedback. We will explore conversational interfaces to simplify input, improving accessibility for individuals. Additionally, we will expand research into stress recovery applications, where music recommendations could serve as an intervention for emotional well-being.

By structuring music-memory linkages, we aim to advance music-based memory retrieval in interactive systems. Future systems incorporating user feedback and improved mood-matching algorithms could deliver personalized and meaningful musical experiences, enhancing interactions.

CRediT authorship contribution statement

Paul Raingeard de la Bletiere: Writing – original draft, Visualization, Software, Methodology, Investigation, Data curation, Conceptualization. **Mark Neerincx:** Writing – review & editing, Supervision. **Rebecca Schaefer:** Writing – review & editing, Supervision. **Catharine Oertel:** Writing – review & editing, Supervision.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Gemma2-9B in order to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was funded by the Dutch Research Council (NWO) through the QoLEAD (Quality of Life by use of Enabling AI in Dementia) Project (project number: KICH1.GZ02.20.008). Additional support from Alzheimer Nederland is gratefully acknowledged. We acknowledge the work of Tygo Huizer and Sophie Haayen, who have greatly contributed to the recruitment of participants and the data collection.

Appendix A. Example of memory data filled in

See Table A.7.

Appendix B. Prompt for using cues in LLM recommendations

Base prompt:

- *If the user does not have a specific artist in mind* ***Prompt A1:*** "You are an assistant working on trying to link one of my memories with songs. You will have to give me a list of 10 songs that would remind me of my memory in the context I will give you. Choose songs released around the given year. If the year is after your cutoff knowledge choose songs from the last year you have (for example 2023). Think step by step, and at then end give your final answer as a list of python tuples with (song name, artist name) after an 'output' keyword, enclosed in double quotes. example: output = [("STAY", "Justin Bieber"), ("Hello", "Adele")]. Just put the main artist, without mentioning featuring artists or collaborators."
- *If the user has a specific artist in mind* ***Prompt A2:*** "You are an assistant working on trying to link one of my memories with songs. You will have to give me a list of 10 songs from the given artist that would remind me of my memory in the context I will give you. If possible, choose songs released around the given year, and before the cutoff year if one is given. If the given year is after your cutoff knowledge choose songs from the last year you have (for example 2023). Think step by step, and at then end give your final answer as a list of python tuples with (song name, artist name) after an 'output' keyword, enclosed in double quotes. example: output = [("STAY", "Justin Bieber"), ("Hello", "Adele")]. Just put the main artist exactly as given by the user if there is one, without mentioning featuring artists or collaborators. You can recommend less than 10 songs if

Table A.7
Personal memories connected to music, people, and places. Each row describes a recollected experience, including the country and year it took place, a brief memory description, the person associated with the memory, any musical artist linked to them, their birth year, the relevant music genre (if any), and a brief description of the place or event where the memory occurred.

Country	Year	Memory description	Person name	Person artist	Birth year	Genre	Event
Belgium	1972	Good feeling on a motorcycle trip from North Holland to Maastricht	Andy	Muse	1975		Belgium – Travel by moped
Germany	1986	During the celebration of the World Cup in Inzell, Hein Vergeer was crowned the winner.	Robert	KC and the Sunshine Band	1979		Inzell – Skateboarding Trip
United Kingdom	2023	I was on a long cycle ride… through the hills of Kent… the whole world was open to me.	me	Everything except loud, hysterical screaming music	1976	Classical	Kent – Cycling trip
France	2019	Meeting	Charlie		1995	Rock	Le Touquet – Seminar

necessary. If you do not know the artist or cannot recommend any songs, return 'output = None".

Example of a standardized user prompt for people if a person is mentioned Prompt B1: "favorite country for music:[Chosen Country or International], artist or genre(if given):[Favorite artist or genre], year:[Year of Birth + 20], context: [Short memory description], cutoff year:[Year of the memory]. This describes the person in my memory, their preferred music and country for this music, the (artist/genre) you need to find songs from, the cutoff year for recommended songs, and the context of the memory."

Example of a standardized user prompt for people if the user is alone in the memory Prompt B2: "favorite country for music:[Chosen Country or International], artist or genre(if given):[Favorite artist or genre], favorite type of music(if given): [Favorite Music], year:[Year of Birth + 20], context: [Short memory description], cutoff year:[Year of the Memory]. This describes myself, my preferred music, the style of songs you need to find songs from, the preferred year for recommendations, the cutoff year for recommended songs, and the context of the memory."

Example of a standardized user prompt for places Prompt B3: "place:[Place], artist or genre(if given):[Artist], country:[Memory Country], year:[Memory Year], context: [Short memory Description]. This describes the place in my memory. Chosen songs should remind me of this country in particular if possible, but you can also add international artists if necessary."

Examples in the context of a memory:

The user chooses their first love encounter from 2023 as a memory. They mention their lover born in 1990, and link them with the artist Billie Eilish. Here is the given prompt from the user, taken from **Prompt B1**, and combined with the system prompt **Prompt A2**: "favorite country for music: International, artist: Billie Eilish, year:2010, context: first encounter with my lover, cutoff year:2023. This describes the person in my memory, their preferred music and country for this music, the (artist/genre) you need to find songs from, the cutoff year for recommended songs, and the context of the memory."

For the place, the user chose the city Paris in France but did not link any artist, which gives **Prompt B1**, later combined with **Prompt A1**: "place: Paris, country: France, year:2023, context: first encounter with my lover. This describes the place in my memory. Chosen songs should remind me of this country in particular if possible, but you can also add international artists if necessary."

Appendix C. Full questionnaire in French, Dutch and English

See Tables C.8–C.10.

Table C.8
Demographic Survey Questions in French, English, and Dutch.

French	English	Dutch
Quel est votre sexe?	What is your gender?	Wat is uw geslacht?
Quel âge avez-vous?	What is your age?	Wat is uw leeftijd?
Quel est le plus haut niveau d'éducation que vous avez terminé ou que vous suivez actuellement?	What is the highest degree or level of school you have completed or that you are currently following?	Wat is het hoogste opleidingsniveau dat u heeft afgerond of waar u nu voor in opleiding bent?
Quelle importance a la musique pour vous?	How important is music to you?	Hoe belangrijk is muziek voor u?
Veuillez indiquer votre préférence de base pour chacun des genres suivants en utilisant l'échelle donnée.	Please indicate your basic preference for each of the following genres using the scale provided.	Geef alstublieft uw basisvoorkeur aan voor elk van de volgende genres met behulp van de gegeven schaal.
À quelle fréquence avez-vous des souvenirs en écoutant une chanson?	How often do you have specific memories while listening to a song?	Hoe vaak heeft u herinneringen terwijl u naar een liedje luistert?
Si vous avez des souvenirs associés à une chanson ou une pièce de musique spécifique, de quelle période de votre vie proviennent-ils généralement?	If you have memories coming up with a song, from which period of your life are they usually?	Als u herinneringen heeft bij een specifiek liedje of muziekstuk, uit welke periode van uw leven komt dit meestal?
À quelle fréquence avez-vous éprouvé des problèmes de mémoire préoccupants au cours de l'année écoulée?	How often in the last year have you been worried about memory problems you experienced?	Hoe vaak heeft u in het afgelopen jaar zorgelijke geheugenproblemen ervaren?
À quelle fréquence avez-vous vécu des périodes de dépression au cours de l'année écoulée?	How often in the last year have you experienced episodes of depressed mood?	Hoe vaak heeft u in het laatste jaar periodes van depressieve stemming meegemaakt?

Table C.9
Survey Questions for Recommendation Rating in French, English, and Dutch.

French	English	Dutch
Cliquez sur le bouton pour indiquer l'émotion que vous avez ressentie en écoutant la chanson ou le morceau de musique.	Please click on the button to indicate the emotion you felt while listening to the song.	Klik op de knop om de emotie aan te geven die u voelde tijdens het luisteren naar het nummer of muziekstuk.
Dans quelle mesure cette chanson est-elle appropriée pour former une association avec votre souvenir?	How appropriate is this song to form an association with your memory?	Hoe passend is dit nummer om een associatie te vormen met uw herinnering?
Pourquoi avez-vous choisi cette valeur?	Why did you choose this value?	Waarom heeft u deze waarde gekozen?
Dans quelle mesure la chanson correspond-elle au contexte dans lequel le souvenir a eu lieu?	How well does the song fit the context in which your memory occurred?	Hoe goed past het nummer bij de context waarin de herinnering plaatsvond?
Différents aspects de la musique peuvent contribuer à expliquer pourquoi cette chanson correspond bien ou moins bien à votre souvenir. Pouvez-vous indiquer quels facteurs influencent l'association de cette musique avec votre souvenir?	Different aspects of the music can contribute to explaining why this song fits or does not fit with your memory. Can you indicate which factors influence the association of this music with your memory?	Verschillende aspecten van de muziek kunnen bijdragen aan waarom dit nummer goed of juist minder goed bij uw herinnering past. Kunt u aangeven welke factoren de associatie van deze muziek met uw geheugen beïnvloeden?
À quel point êtes-vous familier avec ce style de musique?	How familiar is this style of music?	Hoe bekend bent u met deze muziekstijl?
Combien de fois pensez-vous avoir entendu cette chanson/morceau spécifique avant de l'entendre aujourd'hui?	How many times do you think you have heard this specific song/piece before hearing it today?	Hoe vaak denkt u dat u dit specifieke nummer/stuk heeft gehoord voordat u het vandaag hoorde?
Avez-vous d'autres commentaires sur cette recommandation? Peut-être y a-t-il un autre souvenir que vous y associez davantage? Sinon, indiquez 'non'.	Do you have other comments on this recommendation? Perhaps there is another memory you associate more with it? If not, indicate 'no'.	Heeft u nog andere opmerkingen over deze aanbeveling? Misschien is er een andere herinnering die u er meer mee associeerde? Zo niet, vul dan 'nee' in.

Table C.10
Survey Questions for the final discussion in French, English, and Dutch.

French	English	Dutch
Classez les chansons recommandées de la meilleure à la pire correspondance avec votre souvenir.	Rank the recommended songs from best to worst relative to how much you associate them to your memory.	Rangschik de aanbevolen nummers van beste tot slechtste match met uw herinnering.
Quelle est l'importance pour vous du souvenir que vous avez choisi pour cette expérience?	How important to you was the memory you chose for this experiment?	Hoe belangrijk voor u is de herinnering die u voor dit experiment heeft gekozen?
Pensez-vous que vous auriez dû donner plus d'informations sur vous-même pour obtenir une meilleure recommandation? Si oui, quel type d'informations pensez-vous être pertinent?	Do you think that you should have given more information about yourself to get a better recommendation? And if yes, which kind of information do you expect to be relevant?	Denkt u dat u meer informatie over uzelf had moeten geven om een betere aanbeveling te krijgen? En zo ja, welke soort informatie denkt u dat relevant is?
Cette expérience vous a-t-elle fait réfléchir davantage à votre souvenir?	Did this experiment help you reflect on your memory?	Heeft dit experiment u meer laten nadenken over uw herinnering?
Cette expérience vous a-t-elle aidé à mieux comprendre le type de chanson que vous aimeriez associer à votre souvenir?	Did this experiment help you get a better idea of what kind of song you would want to match to your memory?	Heeft dit experiment u geholpen om een beter idee te krijgen van het soort nummer dat u aan uw herinnering zou willen koppelen?
Si oui, quel type de chanson serait-ce? Pourquoi choisiriez-vous ce type de chanson?	If yes, which kind of song would this be? Why would you choose this kind of song?	Indien ja, wat voor soort nummer zou dit zijn? Waarom zou u voor dit soort nummer kiezen?
Y a-t-il une autre chanson spécifique avec laquelle vous l'associeriez davantage? Si oui, pourquoi choisiriez-vous cette chanson?	Is there another specific song that you would have associated more instead? If yes, why would you choose this?	Is er een ander specifiek nummer waarmee u het meer zou associëren? Indien ja, waarom zou u hiervoor kiezen?

Data availability

The data is shared as a link within the article and the references.

References

Agres, K., Lui, S., Herremans, D., 2019. A novel music-based game with motion capture to support cognitive and motor function in the elderly. In: IEEE Conference on Games (CoG) 2019.

Aiken, L.S., West, S.G., 1991. Multiple Regression: Testing and Interpreting Interactions. Sage, Newbury Park, CA.

Alonso-Jiménez, P., Bogdanov, D., Pons, J., Serra, X., 2020. Tensorflow audio models in Essentia. In: International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Andonovski, N., 2022. Episodic representation: a mental models account. Front. Psychol. 13, https://doi.org/10.3389/fpsyg.2022.899371

Arnett, J.J., 2000. Emerging adulthood: a theory of development from the late teens through the twenties. Am. Psychol. 55, 469–480. https://doi.org/10.1037/0003-066x.55.5.469

Baird, A., Brancatisano, O., Gelding, R., Thompson, W.F., 2018. Characterization of music and photograph evoked autobiographical memories in people with alzheimer's disease. J. Alzheimers Dis. 66, 693–706. https://doi.org/10.3233/jad-180627

Baumann, A., Shaw, P., Trotter, L., Clinch, S., Davies, N., 2024. Mnemosyne - supporting reminiscence for individuals with dementia in residential care settings. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/3613904.3642783

Baur, K., Speth, F., Nagle, A., Riener, R., Klamroth-Marganska, V., 2018. Music meets robotics: a prospective randomized study on motivation during robot aided therapy. J. NeuroEng. Rehabil. 15, https://doi.org/10.1186/s12984-018-0413-8
Belfi, A.M., Bai, E., Stroud, A., 2020. Comparing methods for analyzing music-evoked autobiographical memories. Music Percept. 37, 392–402. https://doi.org/10.1525/mp.2020.37.5.392
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300.
Raingeard de la Bletiere, P., Neerincx, M., Schaefer, R., Oertel, C., 2025a. Dataset: A Music Recommender System for Constructed Music Evoked Episodic Memories (CMEEMs)- Non-Personal Data. https://doi.org/10.4121/39b8137d-301e-4fa2-817f-bbd4b791ea30
Raingeard de la Bletiere, P., Neerincx, M., Schaefer, R., Oertel, C., 2025b. Dataset: A Music Recommender System for Constructed Music Evoked Episodic Memories (CMEEMs)-Personal Data. https://doi.org/10.4121/6d820fc9-efcf-4bc8-9fa1-b575e0264f3f
Brinker, B.D., Dinther, R.V., Skowronek, J., 2012. Expressed music mood classification compared with valence and arousal ratings. EURASIP J. Audio Speech Music Process. 2012, https://doi.org/10.1186/1687-4722-2012-24
Broekens, J., Brinkman, W.-P., 2009. Affectbutton: towards a standard for dynamic affective user feedback. In: 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops. IEEE, pp. 1–8, https://doi.org/10.1109/acii.2009.5349347
Broekens, J., Brinkman, W.-P., 2013. Affectbutton: a method for reliable and valid affective self-report. Int. J. Hum.-Comput. Stud. 71, 641–667. https://doi.org/10.1016/j.ijhcs.2013.02.003
Cai, W., Jin, Y., Chen, L., 2021. Critiquing for music exploration in conversational recommender systems. In: Proceedings of the 26th International Conference on Intelligent User Interfaces. Association for Computing Machinery, New York, NY, USA, pp. 480–490, https://doi.org/10.1145/3397481.3450657
Canapa, G., Paternò, F., Santoro, C., 2025. Interactive serious games for cognitive training of older adults: a systematic review. IEEE Trans. Comput. Soc. Syst. 1–24. https://doi.org/10.1109/tcss.2025.3532307
Carraro, D., Bridge, D., 2024. Enhancing recommendation diversity by re-ranking with large language models. ACM Trans. Recomm. Syst. https://doi.org/10.1145/3700604 (just Accepted).
Chater, N., Vitányi, P., 2003. Simplicity: a unifying principle in cognitive science? Trends Cogn. Sci. 7, 19–22. https://doi.org/10.1016/s1364-6613(02)00005-0
Cheng, J., Jiao, C., Luo, Y., Cui, F., 2017. Music induced happy mood suppresses the neural responses to other's pain: evidences from an erp study. Sci. Rep. 7, https://doi.org/10.1038/s41598-017-13386-0
Cohen, J., 2013. Statistical Power Analysis for the Behavioral Sciences. Routledge, https://doi.org/10.4324/9780203771587
Conway, M.A., 2009. Episodic memories. Neuropsychologia 47, 2305–2313. https://doi.org/10.1016/j.neuropsychologia.2009.02.003
Cuan, C., Fisher, E., Okamura, A., Engbersen, T., 2024. Music mode: transforming robot movement into music increases likability and perceived intelligence. J. Hum.-Robot Interact. 14, https://doi.org/10.1145/3686811
Davies, C., Page, B., Driesener, C., Anesbury, Z., Yang, S., Bruwer, J., 2022. The Power of Nostalgia: Age and Preference for Popular Music. 33, 681–692. https://doi.org/10.1007/s11002-022-09626-7.
Deezer SA, 2024. Deezer. https://www.deezer.com. Music streaming service.
Dudzik, B., Hung, H., Neerincx, M., Broekens, J., 2020. Investigating the influence of personal memories on video-induced emotions. In: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization. Association for Computing Machinery, New York, NY, USA, pp. 53–61, https://doi.org/10.1145/3340631.3394842
Farkas, T., Denisova, A., Wiseman, S., Fiebrink, R., 2022. The effects of a soundtrack on board game player experience. In: CHI Conference on Human Factors in Computing Systems. ACM, pp. 1–13, https://doi.org/10.1145/3491102.3502110
Faul, L., LaBar, K.S., 2023. Mood-congruent memory revisited. Psychol. Rev. 130, 1421–1456. https://doi.org/10.1037/rev0000394
Fernández-Pérez, D., Ros, L., Latorre, J.M., 2023. The role of the personal relevance of images in retrieving autobiographical memories for emotion regulation: a randomized controlled trial study. Curr. Psychol. 43, 3523–3537. https://doi.org/10.1007/s12144-023-04582-5
Festini, S.B., McDonough, I.M., 2024. Impact of individual differences in cognitive reserve, stress, and busyness on episodic memory: an FMRI analysis of the Alabama brain study on risk for dementia. Cogn. Affect. Behav. Neurosci. 25, 63–88. https://doi.org/10.3758/s13415-024-01246-0
Fraile, E., Bernon, D., Rouch, I., Pongan, E., Tillmann, B., Lévêque, Y., 2019. The effect of learning an individualized song on autobiographical memory recall in individuals with alzheimer's disease: a pilot study. J. Clin. Exp. Neuropsychol. 41, 760–768. https://doi.org/10.1080/13803395.2019.1617837
Goodall, G., Taraldsen, K., Granbo, R., Serrano, J.A., 2021. Towards personalized dementia care through meaningful activities supported by technology: a multisite qualitative study with care professionals. BMC Geriatrics 21, https://doi.org/10.1186/s12877-021-02408-2
Greene, C.M., Bahri, P., Soto, D., 2010. Interplay between affect and arousal in recognition memory. PLoS ONE 5, e11739. https://doi.org/10.1371/journal.pone.0011739
Greene, N.R., Naveh-Benjamin, M., 2023. Adult age-related changes in the specificity of episodic memory representations: a review and theoretical framework. Psychol. Aging 38, 67–86. https://doi.org/10.1037/pag0000724
Jakubowski, K., Belfi, A.M., Kvavilashvili, L., Ely, A., Gill, M., Herbert, G., 2023. Comparing music- and food-evoked autobiographical memories in young and older adults: a diary study. Br. J. Psychol. 114, 580–604. https://doi.org/10.1111/bjop.12639
Jakubowski, K., Francini, E., 2022. Differential effects of familiarity and emotional expression of musical cues on autobiographical memory properties. Q. J. Exp. Psychol. 76, 2001–2016. https://doi.org/10.1177/17470218221129793
Jakubowski, K., Lee, E., Bai, E., Belfi, A.M., 2024. Individual differences in music-evoked autobiographical memories. Musicae Sci. https://doi.org/10.1177/10298649241288173
Jallais, C., Gilet, A.-L., 2010. Inducing changes in arousal and valence: comparison of two mood induction procedures. Behav. Res. Methods 42, 318–325. https://doi.org/10.3758/brm.42.1.318
Janata, P., Tomic, S.T., Rakowski, S.K., 2007. Characterisation of music-evoked autobiographical memories. Memory 15, 845–860. https://doi.org/10.1080/09658210701734593
Jäncke, L., 2008. Music, memory and emotion. J. Biol. 7, 21. https://doi.org/10.1186/jbiol82
Jeunehomme, O., Heinen, R., Stawarczyk, D., Axmacher, N., D'Argembeau, A., 2022. Representational dynamics of memories for real-life events. iScience 25, 105391. https://doi.org/10.1016/j.isci.2022.105391
Jin, Y., Htun, N.N., Tintarev, N., Verbert, K., 2019. Contextplay: evaluating user control for context-aware music recommendation. In: Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization. Association for Computing Machinery, New York, NY, USA, pp. 294–302, https://doi.org/10.1145/3320435.3320445
Kaiser, A.P., Berntsen, D., 2022. The cognitive characteristics of music-evoked autobiographical memories: evidence from a systematic review of clinical investigations. WIREs Cogn. Sci. 14, https://doi.org/10.1002/wcs.1627
Kamehkhosh, I., Jannach, D., 2017. User perception of next-track music recommendations. In: Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization. ACM, pp. 113–121, https://doi.org/10.1145/3079628.3079668
Kathios, N., Bloom, P.A., Singh, A., Bartlett, E., Algharazi, S., Siegelman, M., Shen, F., Beresford, L., DiMaggio-Potter, M.E., Bennett, S., Natarajan, N., Ou, Y., Loui, P., Aly, M., Tottenham, N., 2024. On the role of familiarity and developmental exposure in music-evoked autobiographical memories. Memory 33, 178–192. https://doi.org/10.1080/09658211.2024.2420973
Kim, S., Kang, H., Choi, S., Kim, D., Yang, M., Park, C., 2024. Large language models meet collaborative filtering: an efficient all-round llm-based recommender system. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. ACM, pp. 1395–1406, https://doi.org/10.1145/3637528.3671931
Kinugawa, K., Schumm, S., Pollina, M., Depre, M., Jungbluth, C., Doulazmi, M., Sebban, C., Zlomuzica, A., Pietrowsky, R., Pause, B., Mariani, J., Dere, E., 2013. Aging-related episodic memory decline: are emotions the key? Front. Behav. Neurosci. 7, https://doi.org/10.3389/fnbeh.2013.00002
Lachman, M.E., Teshale, S., Agrigoroaei, S., 2014. Midlife as a pivotal period in the life course: balancing growth and decline at the crossroads of youth and old age. Int. J. Behav. Dev. 39, 20–31. https://doi.org/10.1177/0165025414533223
Lee, M.L., Dey, A.K., 2007. Providing good memory cues for people with episodic memory impairment. In: Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility. ACM, https://doi.org/10.1145/1296843.1296867
Lewis, P.A., Critchley, H.D., Smith, A.P., Dolan, R.J., 2005. Brain mechanisms for mood congruent memory facilitation. NeuroImage 25, 1214–1223. https://doi.org/10.1016/j.neuroimage.2004.11.053
Li, F., Hu, X., 2023. Background Music for Studying: A Naturalistic Experiment on Music Characteristics and User Perception. 30, 62–72. https://doi.org/10.1109/MMUL.2023.3243209.
Lim, G.H., Hong, S.W., Lee, I., Suh, I.H., Beetz, M., 2013. Robot recommender system using affection-based episode ontology for personalization. In: 2013 IEEE RO-MAN, pp. 155–160, https://doi.org/10.1109/ROMAN.2013.6628437
Barrault, L., Chung, Y., M.C. Meglioli, et al., 2023. Seamless: Multilingual Expressive and Streaming Speech Translation.
Lorah, J., 2018. Effect size measures for multilevel models: definition, interpretation, and timss example. Large-scale Assess. Educ. 6, https://doi.org/10.1186/s40536-018-0061-2
de l'Etoile, S.K., 2002. The effect of a musical mood induction procedure on mood state-dependent word retrieval. J. Music Ther. 39, 145–160. https://doi.org/10.1093/jmt/39.2.145
Means, B., Loftus, E.F., 1991. When personal history repeats itself: decomposing memories for recurring events. Appl. Cogn. Psychol. 5, 297–318. https://doi.org/10.1002/acp.2350050402
Mehl, K., Reschke-Hernandez, A.E., Hanson, J., Linhardt, L., Frame, J., Dew, M., Kickbusch, E., Johnson, C., Bai, E., Belfi, A.M., 2024. Music-evoked autobiographical memories are associated with negative affect in younger and older adults. Exp. Aging Res. 1–18. https://doi.org/10.1080/0361073x.2024.2302785
Mirandola, C., Toffalini, E., 2016. Arousal—but not valence—reduces false memories at retrieval. PLOS ONE 11, e0148716. https://doi.org/10.1371/journal.pone.0148716
Morales-Calva, F., Leal, S.L., 2024. Tell me why: the missing w in episodic memory's what, where, and when. Cogn. Affect. Behav. Neurosci. 25, 6–24. https://doi.org/10.3758/s13415-024-01234-4
Mou, L., Li, J., Li, J., Gao, F., Jain, R., Yin, B., 2021. Memomusic: a personalized music recommendation framework based on emotion and memory. In: 2021 IEEE 4th International Conference on Multimedia Information Processing and Retrieval (MIPR). IEEE, https://doi.org/10.1109/mipr51284.2021.00064
Neerincx, M.A., van Vught, W., Blanson Henkemans, O., Oleari, E., Broekens, J., Peters, R., Kaptein, F., Demiris, Y., Kiefer, B., Fumagalli, D., Bierman, B., 2019. Socio-cognitive engineering of a robotic partner for child's diabetes self-management. Front. Robot. AI 6, https://doi.org/10.3389/frobt.2019.00118
Neo4j, Inc, 2024. Neo4j. https://neo4j.com. Version 5.0.

Norman, G.R., Sloan, J.A., Wyrwich, K.W., 2003. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med. Care. 41, 582–592. https://doi.org/10.1097/01.mlr.0000062554.74615.4c

Panda, R., Malheiro, R., Paiva, R.P., 2023. Audio Features for Music Emotion Recognition: A Survey. 14, pp. 68–88. https://doi.org/10.1109/TAFFC.2020.3032373.

Panda, R., Redinho, H., Gonçalves, C., Malheiro, R., Paiva, R.P., 2021. How does the Spotify API compare to the music emotion recognition state-of-the-art? https://doi.org/10.5281/ZENODO.5045100 https://zenodo.org/record/5045100.

Posner, J., Russel, J.A., Peterson, B.S., 2005. The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev. Psychopathol. 17, https://doi.org/10.1017/s0954579405050340

Raingeard De La Bletiere, P., Neerincx, M.A., Schaefer, R., Bierbach, C.O.G., 2024. Mixed methods analysis of a music recommender system based on episodic memories. https://doi.org/10.17605/OSF.IO/VWFAM. https://osf.io/vwfam/.

Rao, C.B., Peatfield, J.C., McAdam, K.P.W.J., Nunn, A.J., Georgieva, D.P., 2021. A focus on the reminiscence bump to personalize music playlists for dementia. J. Multidiscip. Healthc. 14, 2195–2204. https://doi.org/10.2147/jmdh.s312725

Rentfrow, P.J., Gosling, S.D., 2003. The do re mi's of everyday life: the structure and personality correlates of music preferences. J. Pers. Soc. Psychol. 84, 1236–1256. https://doi.org/10.1037/0022-3514.84.6.1236

Ribeiro, F.S., Santos, F.H., Albuquerque, P.B., Oliveira-Silva, P., 2019. Emotional induction through music: measuring cardiac and electrodermal responses of emotional states and their persistence. Front. Psychol. 10, https://doi.org/10.3389/fpsyg.2019.00451

Samson, L., Carcreff, L., Noublanche, F., Noublanche, S., Vermersch-Leiber, H., Annweiler, C., 2025. User experience of a semi-immersive musical serious game to stimulate cognitive functions in hospitalized older patients: questionnaire study. JMIR Serious Games 13, e57030. https://doi.org/10.2196/57030

Sanner, S., Balog, K., Radlinski, F., Wedin, B., Dixon, L., 2023. Large language models are competitive near cold-start recommenders for language- and item-based preferences. In: Proceedings of the 17th ACM Conference on Recommender Systems. ACM, pp. 890–896, https://doi.org/10.1145/3604915.3608845

Schacter, D.L., Addis, D.R., 2007. The cognitive Neuroscience of constructive memory: remembering the past and imagining the future. Philos. Trans. R. Soc. B Biol. Sci. 362, 773–786. https://doi.org/10.1098/rstb.2007.2087

Sharples, M., Jeffery, N., du Boulay, J.B.H., Teather, D., Teather, B., du Boulay, G.H., 2002. Socio-cognitive engineering: a methodology for the design of human-centred technology. Eur. J. Oper. Res. 136, 310–323. https://doi.org/10.1016/s0377-2217(01)00118-7

Sion, Y., Diaz Reyes, C., Lamas, D., Mokhalled, M., 2023. Vibmory - mapping episodic memories to vibrotactile patterns. In: Proceedings of the Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction. ACM, https://doi.org/10.1145/3569009.3572747

Spotify Ltd, 2024. Spotify. https://www.spotify.com. Music streaming service.

Stephan, Y., Sutin, A.R., Luchetti, M., Terracciano, A., 2020. Personality and memory performance over twenty years: findings from three prospective studies. J. Psychosom. Res. 128, 109885. https://doi.org/10.1016/j.jpsychores.2019.109885

Talarico, J.M., LaBar, K.S., Rubin, D.C., 2004. Emotional intensity predicts autobiographical memory experience. Mem. Cogn. 32, 1118–1132. https://doi.org/10.3758/bf03196886

Tanner, A., Urech, A., Schulze, H., Manser, T., 2023. Older adults' engagement and mood during robot-assisted group activities in nursing homes: development and observational pilot study. JMIR Rehabil. Assist. Technol. 10, e48031. https://doi.org/10.2196/48031

Tikkanen, R., Iivari, N., 2011. The role of music in the design process with children. Springer Berlin Heidelberg, pp. 288–305. https://doi.org/10.1007/978-3-642-23765-2_21

Tulving, E., 1984. Précis of elements of episodic memory. Behav. Brain Sci. 7, 223–238. https://doi.org/10.1017/s0140525x0004440x

Tulving, E., Thomson, D.M., 1973. Encoding specificity and retrieval processes in episodic memory. Psychol. Rev. 80, 352–373. https://doi.org/10.1037/h0020071

Tz-Han, L., Wan-Ru, W., I-Hui, C., Hui-Chuan, H., 2023. Reminiscence music intervention on cognitive, depressive, and behavioral symptoms in older adults with dementia. Geriatr. Nurs. 49, 127–132. https://doi.org/10.1016/j.gerinurse.2022.11.014

Västfjäll, D., 2001. Emotion induction through music: a review of the musical mood induction procedure. Musicae Sci. 5, 173–211. https://doi.org/10.1177/10298649020050s107

Völker, J., 2019. Personalising music for more effective mood induction: exploring activation, underlying mechanisms, emotional intelligence, and motives in mood regulation. Musicae Sci. 25, 380–398. https://doi.org/10.1177/1029864919876315

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E.H., Le, Q.V., Zhou, D., 2024. Chain-of-thought prompting elicits reasoning in large language models. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook, NY, USA.

World Health Organization, 2015. World Report on Ageing and Health. World Health Organization, Geneva.

Wu, J., Chang, C.-C., Yu, T., He, Z., Wang, J., Hou, Y., McAuley, J., 2024. Coral: collaborative retrieval-augmented large language models improve long-tail recommendation. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. ACM, pp. 3391–3401, https://doi.org/10.1145/3637528.3671901

Xi, Y., Liu, W., Lin, J., Chen, B., Tang, R., Zhang, W., Yu, Y., 2024. MemoCRS: memory-enhanced sequential conversational recommender systems with large language models. In: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. Association for Computing Machinery, New York, NY, USA, pp. 2585–2595, https://doi.org/10.1145/3627673.3679599

Zimprich, D., 2018. Individual differences in the reminiscence bump of very long-term memory for popular songs in old age: a non-linear mixed model approach. Psychol. Music 48, 547–563. https://doi.org/10.1177/0305735618812199