Memory

Routledge
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

Music cues impact the emotionality but not richness of episodic memory retrieval

Kelly Jakubowski, Dana Walker & Hongdi Wang

To cite this article: Kelly Jakubowski, Dana Walker & Hongdi Wang (2023): Music cues impact the emotionality but not richness of episodic memory retrieval, Memory, DOI: 10.1080/09658211.2023.2256055

To link to this article: https://doi.org/10.1080/09658211.2023.2256055

Published online: 07 Sep 2023.

MEMORY
https://doi.org/10.1080/09658211.2023.2256055

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Music cues impact the emotionality but not richness of episodic memory retrieval

Kelly Jakubowski [a], Dana Walker[b] and Hongdi Wang[c]

[a]Department of Music, Durham University, Durham, UK; [b]Department of Psychology, Goldsmiths, University of London, London, UK; [c]Department of Psychology, Durham University, Durham, UK

ABSTRACT
Previous studies have found that music evokes more vivid and emotional memories of autobiographical events than various other retrieval cues. However, it is possible such findings can be explained by pre-existing differences between disparate events that are retrieved in response to each cue type. To test whether music exhibits differential effects to other cues even when memory encoding is controlled, we compared music and environmental sounds as cues for memories of the same set of dynamic visual scenes. Following incidental encoding of 14 scenes (7 with music, 7 with sounds), the music and sounds were presented to participants ($N = 56$), who were asked to describe the scenes associated with these cues, and rate various memory properties. Music elicited fewer correct memories and more effortful retrieval than sound cues, and no difference was found in memory detail/vividness between cue types. However, music-evoked memories were rated as more positive and less arousing. These findings provide important critical insights that only partially support the common notion that music differs from other cue types in its effects on episodic memory retrieval.

ARTICLE HISTORY
Received 15 December 2022
Accepted 28 August 2023

KEYWORDS
Episodic memory; music-evoked episodic memory; retrieval cues; encoding; emotion

A range of previous research has demonstrated that different retrieval cues vary both in their efficacy to elicit memories and in terms of the qualities of the memories they evoke. Several such studies have compared cues across sensory modalities revealing, for instance, that olfactory cues evoke significantly more emotional and evocative autobiographical memories than the same cues presented in visual or auditory formats (i.e., campfire, fresh-cut grass, popcorn) (Herz, 2004; Herz & Schooler, 2002). Other work has shown that even cues presented in the same modality can differ in their efficacy; in a study comparing pictorial cues to their visually-presented verbal labels (e.g., "ball", "relaxing on a beach"), the verbal labels evoked significantly more involuntary autobiographical memories than the pictorial cues (Mazzoni et al., 2014).

The aforementioned studies examined differences between autobiographical memories at retrieval. Such paradigms do not assert control over the memory encoding stage, but rather allow participants to retrieve any autobiographical memory from their lives. This means that differences found between the various conditions (e.g., visually versus verbally presented cues) could be due to different autobiographical events being recalled during the different conditions. More recent work has aimed to simulate the encoding and retrieval of everyday events within controlled experiments, which enables the comparison of effects of different cue types on the frequency and content of memories of the *same* event. One notable example is the Simulated Event Paradigm (Congleton et al., 2020; Congleton & Berntsen, 2020) in which first-person perspective videos are presented to participants in an encoding phase, followed by a retrieval phase in which participants describe or rate remembered details from these videos in response to various cue types or prompts (see also a somewhat similar approach used in the trauma film paradigm, (Holmes & Bourne, 2008)).

One cue that has been shown to frequently evoke memories of autobiographical events is music (Jakubowski et al., 2023; Jakubowski & Ghosh, 2021). Music is a stimulus that many people spend a great deal of time and resources listening to (North et al., 2004; Sloboda et al., 2001), and engages listeners cognitively, emotionally, and physically (Janata et al., 2012; Juslin, 2013). Strong experiences with music even activate the brain's reward system (Salimpoor et al., 2011). These emotional and embodied responses to music have been posited as possible explanations for why music-related memory may be preserved longer into older age, even in people with dementia, in comparison to memory for various other everyday stimuli (Stevens, 2015).

To date, research comparing music to other retrieval cue types has been restricted to paradigms invoking retrieval of previously encoded autobiographical memories.

CONTACT Kelly Jakubowski kelly.jakubowski@durham.ac.uk Department of Music, Durham University, Palace Green, Durham, DH1 3RL, UK

Music was found to be less effective, in terms of the number of memories evoked, than various visual cues (words, photographs of famous faces), but elicited more episodically detailed memories than certain (audio)visual cues (photographs of famous faces and TV shows, but not word cues) (Belfi et al., 2016, 2018; Jakubowski et al., 2021; Jakubowski & Eerola, 2022). One set of experiments to date has compared music to environmental sounds as cues for autobiographical memories and found that music evoked fewer but more consistently positive autobiographical memories than the sounds (Jakubowski & Eerola, 2022). A parallel strand of research has focused on comparing music to other cues for autobiographical memory in people with Alzheimer's disease, and found, for instance, a relative preservation of music as a memory cue in comparison to photographs of famous events (Baird et al., 2018) and similar beneficial effects of music and olfactory cues on memory properties (i.e., increased specificity, positive emotions, and mental time travel) compared to a no-cue control condition, with the exception of retrieval times, which were faster for odor- than music-evoked memories (El Haj et al., 2018).

However, to our knowledge, music has not been previously compared against other cues in a paradigm in which the encoding of events is controlled. The present study sought to fill this gap, allowing new insights into whether the aforementioned differences between music and other cues found in natural autobiographical memory retrieval tasks (which typically invoke memories encoded months to years ago) might also be present for newly encoded episodic memories. In addition, controlling the encoding phase enables us to compare music against other cues for memories of the exact *same* events, thereby allowing us to isolate cue-specific from event-specific effects. Our approach thereby allows a novel examination as to whether music cues differentially impact the retrieval of certain details of the same events, or differentially impact memory for the emotionality of these events, in comparison to other cues. Furthermore, such a paradigm allows us to compare music to other cues in terms of *accuracy* of memory recall, which is typically not possible in autobiographical memory retrieval tasks (although see occasional exceptions such as Mace et al., 2011).

In the present experiment, we selected music and non-musical environmental sounds (e.g., nature sounds, crowd sounds, mechanical sounds; referred to as "sounds" hereafter for brevity) that were pre-matched on their emotional expression and memorability. The music and sounds were integrated as soundtracks into videos of egocentrically filmed scenes, to simulate the experience of everyday episodic memory encoding (cf. Congleton et al., 2020). Whilst aiming to provide an effective balance between experimental control and ecological validity, we acknowledge that this approach does not fully replicate the experience of autobiographical memory formation; for instance, we focus on memory for visual details rather than multimodal events (although see previous studies indicating that visual imagery is one of the central components of autobiographical memory (Rubin, 2005)) and the videos we used (of 15 s in duration) are less extended in time than some autobiographical events.

Participants in our experiment were presented the set of videos (accompanied by the music/sound clips) in an incidental encoding task, in which they were asked to rate the "fit" of the music/sound with the visual scene. They then heard each of the music and sound clips in a memory test phase. The primary task invoked free recall of the visual scene associated with the music/sound clip. Participants wrote a description of the remembered visual scene, and rated properties of the memory. Following the free recall task, participants also completed a forced-choice recognition task (regardless of whether they were able to retrieve the memory via free recall), in which they were presented and asked to choose between two of the previously seen videos.

The main aim of our experiment was to compare music to environmental sounds as retrieval cues for newly encoded dynamic visual scenes. A secondary aim was to investigate the effects of the emotional valence and arousal of the music/sound cues on memory recall. We compared the cues in terms of (1) efficacy (number of scenes accurately recalled, retrieval effort ratings), (2) emotionality of the memories evoked (valence and arousal ratings), and (3) richness of the memories evoked (vividness and visual imagery ratings, number of correct details recalled). The number of correct details recalled was coded from the written descriptions of the memories, using a modified version of the Experiential Index (Hassabis et al., 2007). In sum, this work provides new critical insights on the degree to which music cues exhibit differential access to (certain aspects of) episodic memories.

Method

Design

The main independent variable of interest was cue type (music/sound), and the secondary independent variable was the emotional expression of the cues. Three levels of emotional expression were used, which corresponded to three of the four quadrants of the two- dimensional, circumplex model of emotions (Posner et al., 2005; Russell, 1980), specifically: positive valence/high arousal, positive valence/low arousal, and negative valence/high arousal cues. Negative valence/low arousal stimuli were not included as the previous study in which these stimuli were validated revealed that environmental sounds were not particularly effective in conveying negative valence/ low arousal emotions (e.g., sadness; Jakubowski & Eerola, 2022). The dependent variables in our experiment were accuracy of recall of the visual scenes (correct/incorrect), and, for scenes that were recalled correctly, we also compared ratings of the effort required to retrieve the memories, ratings of the vividness, visual imagery, valence, and arousal of the memories, and the number of correct details recalled (see coding scheme reported in *Analysis* section below).

Participants

A power analysis was conducted using G*Power software (Faul et al., 2007), which indicated that to detect a medium effect size ($\eta_p^2 = 0.06$) at 80% power, a minimum of 33 participants was needed. This estimated effect size is in line with the results of previous studies comparing music-evoked autobiographical memories to other autobiographical memories and emotional features of music-evoked autobiographical memories (Belfi et al., 2016; Sheldon & Donahue, 2017). In total, 58 participants completed the experiment. One was excluded from subsequent analysis due to failing both attention checks and another was excluded as the visual scenes used in the experiment were familiar to the participant (they had been to the university in Israel where the videos were filmed). The remaining 56 participants were aged 19–33 years ($M = 22.55$, $SD = 3.56$; 39 female, 16 male, 1 preferred not to report a gender). All reported good fluency in English, and most (64%) spoke English as a first language, with 54% born in the United Kingdom. The next most common birth countries were India (9%) and Hong Kong (5%). The majority (75%) were current undergraduate or postgraduate university students. All self-reported normal hearing and the only visual impairment reported was the need for glasses, which relevant participants wore during the experiment. Most categorised themselves as non-musicians (71%), with 27% reporting to be amateur musicians, and 2% (1 person) classifying themselves as a semi-professional musician. 57% had up to 2 years of formal musical training, and 43% had 3 or more years of formal musical training.

Materials/stimuli

Music, sound, and video stimuli

The music and sounds were sourced from a previous study comparing emotional cues for autobiographical memory (Jakubowski & Eerola, 2022). The music was a range of royalty-free music from the Database for Emotional Analysis in Music (DEAM) (Aljanaki et al., 2017). In the present work we used instrumental music (with no lyrics, except one clip that contained non-linguistic vocals, i.e., "ooo"), comprising one piece from each of the following genres: country, classical, hip-hop, standard jazz, experimental jazz, pop, and rock. These pieces of music were validated in previous research to be highly unfamiliar to participants from the population we studied (young adults in the UK), but were from genres of music deemed to be relatively familiar to this group (Jakubowski & Eerola, 2022). The DEAM dataset also includes ratings of the expressed emotional valence and arousal for each piece of music.

The sounds were originally sourced from a royalty-free, professional sound effects database, SoundEffects+ (https://www.soundeffectsplus.com), and all sounds used here were relatively dynamic in nature and often contained multiple sound sources, to equate these to music. These sounds were rated for familiarity, expressed emotional valence, and expressed emotional arousal in previous research (Jakubowski & Eerola, 2022). The individual sound clips were validated as highly unfamiliar, although the clips were from categories of sounds the participants were likely to have heard before (e.g., crowd sounds, machinery, nature sounds), and Jakubowski and Eerola (2022) found that these sound clips did not differ from the music clips in familiarity ratings. Each music clip was closely matched to a sound clip in terms of expressed valence and arousal ratings via a one-to-one matching procedure (Jakubowski & Eerola, 2022; see Figure 1 in their paper). For the present experiment, we selected 2 excerpts from the positive valence/high arousal quadrant, 2 excerpts from the negative valence/high arousal quadrant, and 3 excerpts from the positive valence/low arousal quadrant for both the music and sound stimulus set, for a total of 14 auditory stimuli (7 music, 7 sounds). All music and sounds were cut to 15 s in duration.

To ensure the music and sounds were also similarly memorable, we ran an online pilot study with 45 participants, aged 18–35 years ($M = 29.2$, $SD = 4.84$; 32 female, 12 male, 1 other). All were UK residents and spoke English as their first language, and none reported any hearing impairments. Participants were presented 18 stimuli (9 music, 9 sounds) of 15 s in duration and were asked to rate their complexity on a 5-point scale. They then completed a memory test, in which they heard 36 stimuli (18 music, 18 sounds), and were asked whether each had been presented in the previous complexity rating task or not. From this pilot study, we were able to ascertain that the 7 music and 7 sound stimuli we used in the main experiment were well-matched on their memorability (mean accuracy rate for music clips = 78% correct, for sound clips = 74% correct; $t(12) = 0.76$, $p = .46$), and also on ratings of the complexity of the stimuli ($M_{music} = 3.00$, $M_{sound} = 2.63$, $t(12) = 0.93$, $p = .37$).

Videos were obtained from the open source HUJI EgoSeg dataset (https://www.vision.huji.ac.il/egoseg/videos/dataset.html), which contains 122 videos captured from an egocentric (GoPro Hero3 + head-mounted) camera (Poleg et al., 2016). From this set, we selected 14 video segments of 15 s in duration. Video segments were selected to be as distinctive from one another as possible (i.e., not filmed in the same location) and relatively neutral in emotional valence/arousal. All videos contained dynamic elements (e.g., change of location or change of objects within the scene during the 15 s). Most videos (10) included people, but these were all seen in the background or in passing; no people in the videos interacted with the cameraperson.

Figure 1. Estimated marginal means for each of the rated memory features, by cue type (upper panel) and cue emotion (lower panel). Error bars represent ± one standard error of the mean. (neg_high = negative valence/high arousal, pos_high = positive valence/high arousal, pos_low = positive valence/low arousal).

Two versions of each video were created, so that each of the 14 videos was paired with both a music clip and the corresponding sound clip that was matched in emotional valence/arousal. Each participant in the main experiment was presented one version of each video, with the two sets of videos counterbalanced across participants. This ensured that any inherent differences in the memorability of the individual videos were distributed equally between the music and sound cues. Music/sound clips were integrated into the videos using iMovie software, and the audio streams from the original videos (i.e., background sounds) were deleted.

Experiment design, protocol, and rating scales

The main experiment was run via Qualtrics, and all participants were tested individually on the same desktop computer, which ensured a consistent size of presentation of the videos across participants. A sound check was completed prior to the main experiment, in which a sample video with music (from the same dataset, but not used in the main experiment) was played and each participant was able to adjust the system volume to a comfortable level. This same sound level was then maintained throughout the experiment. Sound was played aloud via Genelec speakers.

All 14 video stimuli (including the music/sounds) were presented in an incidental encoding task, in a randomised order, in which participants were asked to rate the fit between the auditory and visual aspects of the video on a 5-point scale (1 = very bad fit, 5 = very good fit). Fit ratings were chosen for this task to ensure participants were attending to both the auditory and visual elements of each video, without alerting them to the subsequent memory task. Participants were not made aware that the focus of the study was on memory or music before completing the experiment; they were simply told "This study explores how people process pairings of everyday scenes with a variety of different types of sounds".

Following this initial rating task, participants were then presented with a memory test. They heard each of the 14 auditory (7 music, 7 sound) stimuli from the encoding task, in a randomised order. For each auditory stimulus, they were asked whether they remembered the visual scene that was presented with that sound in the first part of the experiment, with response options of "Yes", "Maybe", or "No". If they responded "Yes" or "Maybe", they were asked to type a description of the corresponding visual scene, with as many details as they could recall from the scene. They then rated the effort required to retrieve the memory, vividness of the memory, valence of the memory, arousal of the memory, and visual imagery within the memory on 5-point scales. The vividness and visual imagery questions were adapted from the Autobiographical Recollection Test (Berntsen et al., 2019). If they answered "No" to the question of whether they recalled the associated visual scene, they were asked an open question on why they thought they were unable to recall the scene; this question was included in order to roughly equate the required response effort between trials comprising recalled and non-recalled scenes, to ensure participants did not simply answer "No" to all questions to advance through the experiment more quickly. Finally, regardless of whether they were able to freely recall the visual scene, each participant was presented two silent videos and asked to select the video they thought had been paired with the sound (forced-choice recognition task). All questions and rating scales used in the memory task are included in the Appendix.

Procedure

The experiment took place in a quiet laboratory, and all participants completed the study under the supervision of a researcher. After obtaining informed consent, participants completed the sound check, followed by the incidental encoding task ("fit" rating task). They then proceeded directly to the memory task. Following these main tasks, demographic information was collected from each participant, including information on their musical background. Participants took around 30 min to complete the experiment and were compensated £10 for their time.

Analysis

Written descriptions of the remembered visual scenes were scored as correct or incorrect by two research assistants. Any discrepancies or uncertainties were checked and resolved by the first author. We adopted a relatively strict scoring procedure, such that the written description had to contain enough detail to discriminate it from all other scenes in the stimulus set. For instance, descriptions such as "It was a sunny day" or "The inside of a building" were marked as incorrect, since there were multiple scenes that could be referenced by such descriptions. Responses of "No" to the initial question of whether a participant remembered the visual scene that was paired with a music/sound cue were automatically scored as incorrect, since no written description of the scene was provided; such cases accounted for 50% of the incorrect recall responses in the dataset.

In addition, for correctly recalled scenes, we coded the number of details recalled according to the four categories of the Experiential Index (Hassabis et al., 2007): Spatial Reference, Entity Present, Sensory Description, and Thought/Emotion/Action. Specifically, following the protocol for the Experiential Index, each written description was segmented into a set of statements/details, and each of these was classified into one of these four categories as relevant. For example, "it was on the left" was coded as a Spatial Reference and "there are a few people" was coded as an Entity Present. Repetitions of the same information (e.g., referring to the same car in multiple statements within the memory description) were only coded once. We then summed the values across each of these four categories to create a composite score representing the number of details recalled. We also initially included categories for External Details (e.g., repeated statements or details not directly related to the scene contents such as "I think this was the first video I saw") and Incorrect Details (e.g., stating there were "a couple people" in the scene when there was only one). However, it was found that responses falling into these categories were relatively rare (83% of memories had no External Details, 97% of memories had no Incorrect Details); as such, they were not considered in the main analysis.

Two raters were initially trained on the coding protocols and categorised all correct memory descriptions for three videos (104 memory descriptions, 22% of all correctly recalled memories). Intraclass correlation coefficients (ICC) were calculated using two-way mixed effects models (with random effects for both rater and participant) via the R package "irr" (Gamer et al., 2019). ICC values for each of the three videos were 0.85, 0.83, and 0.93, indicating a good to excellent level of reliability between the two raters that is in line with previous studies using similar memory coding schemes (e.g., Belfi et al., 2016, 2022). As such, the remaining memories were coded by one of the two raters.

For the main analyses we used mixed effects models, with "Participant" included as a random effect, given the multiple observations per participant in the dataset. "Video" was also included as a random effect, to control for any potential differences between the 14 videos used. The fixed effects included in the models were cue type (two categories: music, sound), cue emotion (three categories: positive valence/high arousal, negative valence/high arousal, positive valence/low arousal), and the interaction of cue type and cue emotion. Binomial mixed effects models were used to investigate the effects of these factors on whether each trial was recalled correctly or incorrectly in the free recall task and forced-choice recognition task. For correctly recalled trials, we also used a binomial mixed effects model to predict whether participants responded "Yes" or "Maybe" that they were able to recall the scene (as an index of confidence of recall). Linear mixed effects models were used for all other dependent variables. Models were fitted using the "lme4" package in R (Bates et al., 2015) and the statistical significance of the fixed effects was assessed with Wald χ^2 tests using the Anova() function of the "car" package (Fox & Weisberg, 2019). The Wald test is a classic approach to null hypothesis significance testing, similar to the likelihood ratio test, and tests whether the estimated value of each coefficient in the model is significantly different from zero by comparing the coefficient's estimated value with the estimated standard error for the coefficient. Post hoc contrasts were performed using the "emmeans" R package (Lenth, 2022). All collected data are available at https://osf.io/96pyu/.

Results

Recall accuracy

In total, 49% of the visual scenes cued by music were recalled correctly in the free recall task, and 73% of the sound-cued scenes were recalled correctly. 66% of the scenes evoked by positive valence/low arousal cues were recalled correctly; this figure was 63% for negative valence/high arousal cues and 52% for positive valence/high arousal cues. A binomial mixed effects model revealed a significant effect of cue type, with visual scenes that were cued by sounds more likely to be correctly retrieved than those cued by music ($\chi^2(1) = 50.48$, $p < .001$). No significant effect of cue emotion ($\chi^2(2) = 3.38$, $p = .18$) or interaction of cue type and cue emotion ($\chi^2(2) = 5.24$, $p = .07$) were found.

In the forced-choice recognition task, performance was very high, suggesting the video and music/sound pairings were well encoded overall. Total scores ranged from 11 to 14 trials (out of 14) correct, with a mean score of 13.36 (*SD* = 0.92). Even on this relatively easier task, the sound cues elicited better memory performance than the music cues ($M_{music} = 6.52$, $M_{sound} = 6.84$), although this difference was not statistically significant in a binomial mixed effects model ($\chi^2(1) = 2.76$, $p = .10$). No significant effect of cue emotion ($\chi^2(2) = 1.10$, $p = .58$) or interaction of cue type and cue emotion ($\chi^2(2) = 0.09$, $p = .95$) were found on recognition task performance.

For trials where the scene was recalled correctly, we also fitted a binomial mixed effects model predicting whether participants responded "Yes" or "Maybe" to the question of whether they were able to recall the scene. A significant effect of cue type ($\chi^2(1) = 14.20$, $p < .001$) revealed that "Yes" responses were more likely in response to sound than music cues. A significant effect of cue emotion was also found ($\chi^2(1) = 8.12$, $p = .02$), with post hoc, Bonferroni-correct pairwise contrasts revealing that negative valence/high arousal cues elicited more "Yes" responses than positive valence/high arousal cues ($p = .03$, all other $ps > .07$). No significant interaction of cue type and cue emotion ($\chi^2(2) = 0.91$, $p = .63$) was found.

Memory features

Table 1 shows the results from the linear mixed effects models predicting ratings of the memory features, and Figure 1 shows the estimated marginal means from these models by cue type and cue emotion. In the model predicting ratings of retrieval effort, the only statistically significant predictor was cue type (see Table 1), with memories of visual scenes cued by sound cues being rated as requiring less retrieval effort than those cued by music (see Figure 1). No effects of cue type, cue emotion, or interaction were found on ratings of the vividness or amount of visual imagery within the remembered scenes.

Valence ratings of the remembered scenes were significantly affected by both cue type and cue emotion. Scenes retrieved in response to music cues were rated as more positive than scenes retrieved in response to sound cues. Although the means indicated a trend toward memories evoked by negatively valenced cues being rated as more negative, there were no statistically significant differences between the three emotion categories for the cues in post hoc, pairwise contrasts with Bonferroni correction ($ps > .08$).

Table 1. Main effects and interaction of the fixed predictors (cue type, cue emotion) from the linear mixed effects models predicting ratings of retrieval effort, vividness, visual imagery, valence, and arousal for the remembered scenes.

Dependent measure	Effect of cue type	Effect of cue emotion	Cue type × cue emotion interaction
Effort	$\chi^2(1) = 4.51$, $p = .034$*	$\chi^2(2) = 0.52$, $p = .77$	$\chi^2(2) = 1.66$, $p = .44$
Vividness	$\chi^2(1) = 1.44$, $p = .23$	$\chi^2(2) = 0.11$, $p = .95$	$\chi^2(2) = 0.02$, $p = .99$
Visual Imagery	$\chi^2(1) = 1.05$, $p = .30$	$\chi^2(2) = 0.26$, $p = .88$	$\chi^2(2) = 2.96$, $p = .23$
Valence	$\chi^2(1) = 42.12$, $p < .001$***	$\chi^2(2) = 6.16$, $p = .046$*	$\chi^2(2) = 2.09$, $p = .35$
Arousal	$\chi^2(1) = 19.60$, $p < .001$***	$\chi^2(2) = 39.53$, $p < .001$***	$\chi^2(2) = 4.05$, $p = .13$

Note: * = $p < .05$, *** = $p < .001$; "Participant" and "Video" were included as random effects in all models.

Arousal ratings were also significantly impacted by both cue type and cue emotion. Scenes recalled in response to music were rated as more relaxing (less arousing) than scenes recalled in response to sounds. Negative valence/high arousal cues evoked memories rated as more arousing than both positive valence/high arousal cues ($p = .004$) and positive valence/low arousal cues ($p < .001$) in post hoc, pairwise contrasts with Bonferroni correction. No difference was found in memory arousal ratings between the positive valence/high arousal cues and positive valence/low arousal cues ($p = .47$).

Memory descriptions

On average, participants wrote 29 words when describing the recalled visual scenes (range = 3–127 words, *SD* = 18). An initial analysis via a linear mixed effects model revealed that the total word count of the written descriptions did not systematically vary according to cue type ($\chi^2(1) = 0.02$, $p = .90$), cue emotion ($\chi^2(2) = 0.39$, $p = .82$), or their interaction ($\chi^2(2) = 0.75$, $p = .69$). The number of correct details recalled within the written descriptions (composite score from the Experiential Index coding process) ranged from 1 to 26 ($M = 8$, $SD = 4$). A linear mixed effects model predicting the number correct of details recalled revealed no significant effect of cue type ($\chi^2(1) = 0.06$, $p = .81$), cue emotion ($\chi^2(2) = 2.52$, $p = .28$), or their interaction ($\chi^2(2) = 0.05$, $p = .98$).

Fit between visual and auditory pairings

One potential explanation for the difference we found in recall accuracy between the music- and sound-cued memories is that there was some difference in how well the music/sounds complemented the visual scenes they accompanied, which may have affected the binding of the visual and auditory elements in memory. To probe this idea further, we examined the fit ratings between the visual and auditory elements that we collected during the encoding task. We found that, overall, the music cues ($M = 2.88$, $SD = 0.49$) were rated higher in their fit to the visual scenes than the sounds ($M = 1.97$, $SD = 0.49$), via a paired-samples t-test, $t(55) = 10.11$, $p < .001$.

We also reran the binomial mixed effects model predicting accuracy of recall of each scene with fit ratings included as an additional predictor. We found that fit ratings were a significant positive predictor ($\beta = 0.32$, $SE = 0.09$, $\chi^2(1) = 13.62$, $p < .001$), with scenes rated higher in fit between the visual and auditory elements more likely to be recalled correctly. In addition, cue type was still a statistically significant predictor of recall accuracy ($\chi^2(1) = 57.45$, $p < .001$), with sounds more likely to cue a correctly recalled scene than music. Furthermore, there was a significant interaction between cue type and fit ratings ($\chi^2(1) = 6.37$, $p = .01$), which revealed that increasing fit ratings had more impact on increasing recall accuracy for music-evoked memories than sound-evoked memories (estimated slope from the model for music cues = 0.54, for sound cues = 0.10). Taken together, these results indicate that the overall difference in recall accuracy of memories elicited by music versus sound cues cannot be explained solely by how well they were perceived to fit with the visual scene.

Discussion

We developed a novel paradigm for comparing music and environmental sounds as cues for incidentally encoded memories of dynamic, everyday visual scenes. Music differed from other sound cues in terms of the number of memories evoked and emotion ratings of the remembered scenes, with no difference between the two cue types in the vividness or episodic richness of the memories. These findings extend previous research by demonstrating that different retrieval cue types differentially impact the accessibility and emotionality of episodic memories (cf. Congleton et al., 2020; Herz, 2004; Herz & Schooler, 2002; Mazzoni et al., 2014). These results occurred despite the fact that both cue types were presented in the same (auditory) modality and were matched on key features such as their emotional expression, memorability, and complexity. Previous research has demonstrated that differences in the presentation of highly similar cues in the same modality (e.g., visually presented verbal versus pictorial cues; Mazzoni et al., 2014) significantly impact the number of autobiographical memories evoked; our findings extend such results to memory for the same set of newly encoded events.

The sound cues evoked a significantly greater number of correct memories than the music in a free recall task. In addition, ratings of retrieval effort were significantly lower (i.e., less effort required) and a greater proportion of "Yes" rather than "Maybe" responses were given to the free recall question (indicating greater confidence in recall) for sound-evoked than music-evoked memories. This aligns with findings from a study that used these same stimuli as cues in an autobiographical memory retrieval task (Jakubowski & Eerola, 2022), suggesting music is a less effective cue than emotionally matched environmental sounds for both memories freely selected from one's entire store of lifetime events *and* those encoded within a controlled experiment. More broadly, other studies using autobiographical memory retrieval paradigms have found similar results, in that music cued fewer autobiographical memories than photographs of famous faces (Belfi et al., 2016) and famous events (Baird et al., 2018) in healthy adults. One potential explanation for the lower efficacy of music as a memory cue may be the less referential nature of music in comparison to other cue types. For example, other sounds, famous faces, and events may be more easily associated with verbal labels, which may facilitate memory retrieval, especially if such labels are relatively concrete and imageable (Rubin & Schulkind, 1997; Uzer et al., 2012; Williams et al., 1999). Future research could investigate whether music with more concretely referential structures (e.g., music composed to sound like a river or traffic jam) might be a more effective memory cue than other types of music.

The music and sounds were paired at encoding with the same (relatively neutral) everyday scenes across different participants, and these two cue types were carefully matched on their emotional expression. Despite this, music-evoked memories were rated as significantly more positive and relaxing (less arousing) than sound-evoked memories. This also partially aligns with the findings of Jakubowski and Eerola (2022), who found that music consistently evoked relatively positive autobiographical memories regardless of its emotional valence, while negatively valenced sound and word cues evoked more negative memories. Here, we did not find a significant interaction as in this previous work, but simply that scenes recalled in response to music cues were rated as more positive at retrieval. Music-evoked memories were also rated as more relaxing. This result bears resemblance to the findings of a study of individuals with mild Alzheimer's disease, who showed a significant reduction in anxiety in an autobiographical recall task during music than a silent control condition (Irish et al., 2006). However, our results go beyond these findings to indicate that music cannot only decrease felt anxiety, but can actually change how an event is remembered. Taken together, these findings indicate that music may bias our memories of the emotionality of events. These results provide important support for therapeutic and wellbeing-related uses of music, since they indicate that music cues may be somewhat unique in terms of how they shape the emotional tone of our memories.

Negative valence/high arousal cues also evoked memories with greater confidence (i.e., greater proportion of "Yes" over "Maybe" responses to the free recall question) than positive valence/high arousal cues. This result broadly aligns with previous findings that negative emotions enhance recall of central details of events (Berntsen, 2002) and that, when arousal differences are controlled for, negative information is remembered more accurately and confidently than positive information (Dewhurst & Parry, 2000; Kensinger et al., 2007). In addition, negative valence/high arousal cues evoked memories rated as more arousing. Previous studies using music as cues for autobiographical memories have found that both high arousal and *positive* valence are significant predictors of memory arousal (e.g., Jakubowski & Francini, 2022), suggesting a slight difference between the findings produced by our task and tasks invoking free retrieval of any autobiographical memory.

No significant effects of cue type (music/sound) were found on ratings of vividness, ratings of visual imagery, the number of words in the memory descriptions, or the number of correct details recalled (as coded via the Experiential Index). Taken together, these results suggest music is not different from other sounds in terms of the detail or richness of memories evoked when comparing these as cues for newly encoded everyday scenes. As several studies have found that music-evoked autobiographical memories are more vivid and episodically detailed than other types of autobiographical memories (Belfi et al., 2016, 2018, 2022; Jakubowski et al., 2021), our results suggest that such previous findings may be explained by inherent differences between the remembered events. For instance, music may be more frequently coupled to autobiographical events that are encoded in more detail in the first place than various other common cues, perhaps because such events are particularly emotional, salient, or self-defining (Talarico et al., 2004, 2009; Walker et al., 2003).

When scenes were rated higher in fit between the visual and auditory elements, they were more likely to be recalled correctly. Previous research has shown that events comprising multiple elements are remembered better when these elements are congruous or semantically related (Craik & Tulving, 1975; Schulman, 1974; Staresina et al., 2009). In our study, this was effect was more pronounced for the music than sound cues, which indicates that increasing fit within the context of an event could be one potential means of improving the relatively poorer music-evoked memory performance evidenced here.

More broadly, this research provides a necessary critical insight into the common notion that music may be more effective, or cue more vivid and emotional memories, than other episodic memory retrieval cues (c.f., Halpern et al., 2018), by controlling the memory encoding phase and events that are coupled to the cues. The paradigm introduced here could be used to further contrast music against other cue types (including other aesthetic stimuli, word cues, etc.), to provide more comprehensive insights into whether the present results replicate across different domains. To further increase the comparability to naturalistic autobiographical memory tasks, this paradigm could also be adapted to study memory recall over longer time delays, for instance days or months after initial encoding. Furthermore, it would be useful to obtain emotion ratings of the auditory/visual stimulus pairings *at encoding* as well as at retrieval, to test whether music also differentially impacts emotional responses to the scenes on first exposure. Finally, the everyday scenes we used here were relatively neutral and mundane; future research should thereby probe the extent to which pairing music/sounds to *emotional* events (including those congruent/incongruent with the emotionality of the cues) affects performance on these tasks.

In conclusion, we found that music was a less effective retrieval cue than other (emotionally matched) sounds for newly encoded dynamic visual scenes, both in terms of retrieval accuracy and the perceived effort required to retrieve a memory. Scenes recalled in response to music cues were rated as more positive and relaxing than sound-evoked memories of such scenes, and the two cue types did not differ in relation to the vividness/amount of detail recalled. Thus, when the encoding stage is controlled, music does not appear to prevail over other cues in the number or richness of memories it evokes, but music does seem to differentially affect the emotionality of these memories. Such work has key implications for a range of applications in which music is used as a memory cue, from reminiscence therapies to nostalgia-inducing advertisements.

Open Scholarship

This article has earned the Center for Open Science badge for Open Data. The data are openly accessible at https://osf.io/96pyu/ .

Acknowledgements

Thanks to Tuomas Eerola (Durham University) for advice on study design and stimulus selection, and to Akvilė Jadzgevičiūtė (Durham University) for assistance with data collection.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by a Leverhulme Trust Early Career Fellowship [grant number ECF2018-209] awarded to the first author. Ethical approval was granted by the Durham University Music Department Ethics Committee. All collected data are available at https://osf.io/96pyu/.

ORCID

Kelly Jakubowski http://orcid.org/0000-0002-4954-7117

References

Aljanaki, A., Yang, Y. H., & Soleymani, M. (2017). Developing a benchmark for emotional analysis of music. *PLoS ONE*, *12*(3), e0173392. https://doi.org/10.1371/journal.pone.0173392

Baird, A., Brancatisano, O., Gelding, R., & Thompson, W. F. (2018). Characterization of music and photograph evoked autobiographical memories in people with Alzheimer's disease. *Journal of Alzheimer's Disease*, *66*(2), 693–706. https://doi.org/10.3233/JAD-180627

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Belfi, A. M., Bai, E., Stroud, A., Twohy, R., & Beadle, J. N. (2022). Investigating the role of involuntary retrieval in music-evoked autobiographical memories. *Consciousness and Cognition*, *100*, 103305. https://doi.org/10.1016/j.concog.2022.103305

Belfi, A. M., Karlan, B., & Tranel, D. (2016). Music evokes vivid autobiographical memories. *Memory*, *24*(7), 979–989. https://doi.org/10.1080/09658211.2015.1061012

Belfi, A. M., Karlan, B., & Tranel, D. (2018). Damage to the medial prefrontal cortex impairs music-evoked autobiographical memories. *Psychomusicology: Music, Mind, and Brain*, *28*(4), 201–208. https://doi.org/10.1037/pmu0000222

Berntsen, D. (2002). Tunnel memories for autobiographical events: Central details are remembered more frequently from shocking than from happy experiences. *Memory & Cognition*, *30*(7), 1010–1020. https://doi.org/10.3758/BF03194319

Berntsen, D., Hoyle, R. H., & Rubin, D. C. (2019). The autobiographical recollection test (ART): A measure of individual differences in autobiographical memory. *Journal of Applied Research in Memory and Cognition*, *8*(3), 305–318. https://doi.org/10.1037/h0101839

Congleton, A. R., & Berntsen, D. (2020). The devil is in the details: Investigating the influence of emotion on event memory using a simulated event. *Psychological Research*, *84*(8), 2339–2353. https://doi.org/10.1007/s00426-019-01215-0

Congleton, A. R., Nielsen, N. P., & Berntsen, D. (2020). Through the gateway of the senses: Investigating the influence of sensory modality-specific retrieval cues on involuntary episodic memory. *Psychological Research*, *85*, 1292–1306. https://doi.org/10.1007/s00426-020-01304-5

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Dewhurst, S. A., & Parry, L. A. (2000). Emotionality, distinctiveness, and recollective experience. *European Journal of Cognitive Psychology*, *12*(4), 541–551. https://doi.org/10.1080/095414400750050222

El Haj, M., Gandolphe, M. C., Gallouj, K., Kapogiannis, D., & Antoine, P. (2018). From nose to memory: The involuntary nature of odor-evoked autobiographical memories in Alzheimer's disease. *Chemical Senses*, *43*(1), 27–34. https://doi.org/10.1093/chemse/bjx064

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage.

Gamer, M., Lemon, J., & Fellows Puspendra Singh, I. (2019). *irr: Various coefficients of interrater reliability and agreement*. R package version 0.84.1.

Halpern, A. R., Talarico, J. M., Gouda, N., & Williamson, V. J. (2018). Are musical autobiographical memories special? It ain't necessarily so. *Music Perception*, *35*(5), 561–572. https://doi.org/10.1525/mp.2018.35.5.561

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences*, *104*(5), 1726–1731. https://doi.org/10.1073/pnas.0610561104

Herz, R. S. (2004). A naturalistic analysis of autobiographical memories triggered by olfactory visual and auditory stimuli. *Chemical Senses*, *29*(3), 217–224. https://doi.org/10.1093/chemse/bjh025

Herz, R. S., & Schooler, J. W. (2002). A naturalistic study of autobiographical memories evoked by olfactory and visual cues: Testing the Proustian hypothesis. *The American Journal of Psychology*, *115*(1), 21–32. https://doi.org/10.2307/1423672

Holmes, E. A., & Bourne, C. (2008). Inducing and modulating intrusive emotional memories: A review of the trauma film paradigm. *Acta Psychologica*, *127*(3), 553–566. https://doi.org/10.1016/j.actpsy.2007.11.002

Irish, M., Cunningham, C. J., Walsh, J. B., Coakley, D., Lawlor, B. A., Robertson, I. H., & Coen, R. F. (2006). Investigating the enhancing effect of music on autobiographical memory in mild Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *22*(1), 108–120. https://doi.org/10.1159/000093487

Jakubowski, K., Belfi, A. M., & Eerola, T. (2021). Phenomenological differences in music- and television-evoked autobiographical memories. *Music Perception*, *38*(5), 435–455. https://doi.org/10.1525/mp.2021.38.5.435

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. *British Journal of Psychology*, *114*, 580–604. https://doi.org/10.1111/bjop.12639

Jakubowski, K., & Eerola, T. (2022). Music evokes fewer but more positive autobiographical memories than emotionally matched sound and word cues. *Journal of Applied Research in Memory and Cognition*, *11*(2), 272–288. https://doi.org/10.1016/j.jarmac.2021.09.002

Jakubowski, K., & Francini, E. (2022). Differential effects of familiarity and emotional expression of musical cues on autobiographical memory properties. *Quarterly Journal of Experimental Psychology*, *76*, 2001–2016. https://doi.org/10.1177/17470218221129793

Jakubowski, K., & Ghosh, A. (2021). Music-evoked autobiographical memories in everyday life. *Psychology of Music*, *49*(3), 649–666. https://doi.org/10.1177/0305735619888803

Janata, P., Tomic, S. T., & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *Journal of Experimental Psychology: General*, *141*(1), 54–75. https://doi.org/10.1037/a0024208

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of Life Reviews*, *10*(3), 235–266. https://doi.org/10.1016/j.plrev.2013.05.008

Kensinger, E. A., O'Brien, J. L., Swanberg, K., Garoff-Eaton, R. J., & Schacter, D. L. (2007). The effects of emotional content on reality-monitoring performance in young and older adults. *Psychology and Aging*, *22*(4), 752–764. https://doi.org/10.1037/0882-7974.22.4.752

Lenth, R. V. (2022). *Emmeans: Estimated marginal means, aka least-squares means*. R Package Version 1.7.2. https://CRAN.R-project.org/package=emmeans

Mace, J. H., Atkinson, E., Moeckel, C. H., & Torres, V. (2011). Accuracy and perspective in involuntary autobiographical memory. *Applied Cognitive Psychology*, *25*(1), 20–28. https://doi.org/10.1002/acp.1634

Mazzoni, G., Vannucci, M., & Batool, I. (2014). Manipulating cues in involuntary autobiographical memory: Verbal cues are more effective than pictorial cues. *Memory & Cognition*, *42*(7), 1076–1085. https://doi.org/10.3758/s13421-014-0420-3

North, A. C., Hargreaves, D. J., & Hargreaves, J. J. (2004). Uses of music in everyday life uses of music in everyday life. *Music Perception: An Interdisciplinary Journal*, *22*(1), 41–77. https://doi.org/10.1525/mp.2004.22.1.41

Poleg, Y., Ephrat, A., Peleg, S., & Arora, C. (2016). Compact CNN for indexing egocentric videos. *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–9. IEEE. https://doi.org/10.1109/WACV.2016.7477708

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(3), 715–734. https://doi.org/10.1017/S0954579405050340

Rubin, D. C. (2005). A basic-systems approach to autobiographical memory. *Current Directions in Psychological Science*, *14*(2), 79–83. https://doi.org/10.1111/j.0963-7214.2005.00339.x

Rubin, D. C., & Schulkind, M. D. (1997). Properties of word cues for autobiographical memory. *Psychological Reports*, *81*(1), 47–50. https://doi.org/10.2466/pr0.1997.81.1.47

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience*, *14*(2), 257–262. https://doi.org/10.1038/nn.2726

Schulman, A. I. (1974). Memory for words recently classified. *Memory & Cognition*, *2*(1), 47–52. https://doi.org/10.3758/BF03197491

Sheldon, S., & Donahue, J. (2017). More than a feeling: Emotional cues impact the access and experience of autobiographical memories. *Memory & Cognition*, *45*(5), 731–744. https://doi.org/10.3758/s13421-017-0691-6

Sloboda, J. A., O'Neill, S. A., & Ivaldi, A. (2001). Functions of music in everyday life: An exploratory study using the experience sampling method. *Musicae Scientiae*, *5*(1), 9–32. https://doi.org/10.1177/102986490100500102

Staresina, B. P., Gray, J. C., & Davachi, L. (2009). Event congruency enhances episodic memory encoding through semantic elaboration and relational binding. *Cerebral Cortex*, *19*(5), 1198–1207. https://doi.org/10.1093/cercor/bhn165

Stevens, C. J. (2015). Is memory for music special? *Memory Studies*, *8*(3), 263–266. https://doi.org/10.1177/1750698015584873

Talarico, J. M., Berntsen, D., & Rubin, D. C. (2009). Positive emotions enhance recall of peripheral details. *Cognition & Emotion*, *23*(2), 380–398. https://doi.org/10.1080/02699930801993999

Talarico, J. M., Labar, K. S., & Rubin, D. C. (2004). Emotional intensity predicts autobiographical memory experience. *Memory & Cognition*, *32*(7), 1118–1132. https://doi.org/10.3758/BF03196886

Uzer, T., Lee, P. J., & Brown, N. R. (2012). On the prevalence of directly retrieved autobiographical memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1296–1308. https://doi.org/10.1037/a0028142

Walker, W. R., Skowronski, J. J., & Thompson, C. P. (2003). Life is pleasant—and memory helps to keep it that way!. *Review of General Psychology*, *7*(2), 203–210. https://doi.org/10.1037/1089-2680.7.2.203

Williams, J. M. G., Healy, H. G., & Ellis, N. C. (1999). The effect of imageability and predicability of cues in autobiographical memory. *The Quarterly Journal of Experimental Psychology Section A*, *52*, 555–579. https://doi.org/10.1080/713755828

Appendix: Main Experiment Questions

Memory feature questions (response options in italics)

Do you remember the visual scene that was presented with this sound?
Yes; Maybe; No

If a response of "Yes' or "Maybe" was given to the first question, the following questions were displayed:

Please provide a written description of the visual scene you remembered. Please provide as many details as you are able to recall from the scene.
Open response

How much effort did it take to recall this visual scene?
*1 (*No effort at all*); 2; 3; 4; 5 (*A lot of effort*)*

My memory of this visual scene has lots of details.
*1 (*Strongly disagree*); 2; 3; 4; 5 (*Strongly agree*)*

The feelings I experience as I remember the visual scene are …
*1 (*Very negative*); 2; 3 (*Neither negative nor positive*); 4; 5 (*Very positive*)*

When remembering the visual scene, I feel …
*1 (*Very relaxed*); 2; 3 (*Neither relaxed nor alert*); 4; 5 (*Very alert*)*

While remembering the scene, I can see it in my mind.
*1 (*Strongly disagree*); 2; 3; 4; 5 (*Strongly agree*)*

If a response of "No" was given to the first question, the following question was displayed:

Why do you think you were unable to recall a visual scene for this particular sound?
Open response

The following question was displayed to all participants, regardless of their response to the first question:

Which of these two visual scenes do you think was paired with this sound?
Video 1; Video 2 (both played as silent videos)

Demographic questions (response options in italics)

Please enter your gender
Male; Female; Other; Prefer not to say

In which country were you born?
Dropdown menu

What is your first (native) language?
Open response

What is the highest educational qualification you have attained?
Primary school; High school/GCSE; A-Levels; Currently pursuing undergraduate degree; Undergraduate degree completed; Currently pursuing postgraduate degree; Postgraduate degree completed

Do you currently suffer from any hearing impairments?
Yes; No

[If "yes" response to above] Please provide a short description of the hearing impairment, and any measures you are currently taking to correct it (wearing a hearing aid, etc.).
Open response

Do you currently suffer from any visual impairments?
Yes; No

[If "yes" response to above] Please provide a short description of the visual impairment, and any measures you are currently taking to correct it (wearing glasses, etc.).
Open response

Musicianship questions (response options in italics)

Which of the following best describes you?
Non-musician; music-loving non-musician; amateur musician; serious amateur musician; semi-professional musician; professional musician

I have had __ years of formal training on a musical instrument (including voice) during my lifetime.
0; 0.5; 1; 2; 3–5; 6–9; 10 or more