Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-021-01947-4

BRIEF REPORT

Does task-irrelevant music affect gaze allocation during real-world scene viewing?

Kristina Krasich[1] · Joanne Kim[2] · Greg Huffman[3] · Annika L. Klaffehn[4] · James R. Brockmole[2]

Accepted: 2 May 2021

Abstract

Gaze control manifests from a dynamic integration of visual and auditory information, with sound providing important cues for how a viewer should behave. Some past research suggests that music, even if entirely irrelevant to the current task demands, may also sway the timing and frequency of fixations. The current work sought to further assess this idea as well as investigate whether task-irrelevant music could also impact how gaze is spatially allocated. In preparation for a later memory test, participants studied pictures of urban scenes in silence or while simultaneously listening to one of two types of music. Eye tracking was recorded, and nine gaze behaviors were measured to characterize the temporal and spatial aspects of gaze control. Findings showed that while these gaze behaviors changed over the course of viewing, music had no impact. Participants in the music conditions, however, did show better memory performance than those who studied in silence. These findings are discussed within theories of multimodal gaze control.

Keywords Gaze control · Music · Visual salience · Semantic informativeness · Eye tracking

Visual processing of the world unfolds across both space and time. Viewing the environment in a piecemeal fashion, we shift our high-fidelity foveal vision to various locations about 3–4 times per second. Spatially, where people look is based on low-level visual information (e.g., Borji & Itti, 2013; Parkhurst et al., 2002), momentary task goals (e.g., Land & Lee, 1994; Yarbus, 1967), scene structure (e.g., Torralba et al., 2006), and scene context (e.g., Shinoda et al., 2001; Võ & Henderson, 2009), which all indicate where important visual information is likely to be located. The time people spend looking at these locations reflects the quality of the available visual information (e.g., Najemnik & Geisler, 2005, 2009), the ease with which objects can be recognized and understood in context (e.g., Hollingworth, 2006), and the relevant goals and strategies of the observer (e.g., Neider & Zelinsky, 2006). Thus, these temporal properties are thought to be linked to the amount of effort that is needed to understand fixated information.

Most studies investigating the parameters of gaze control are unimodal, focusing only on what can be seen. This underrepresents our multimodal environments, in which non-visual information, such as sound, may further guide our gaze behaviors. For example, sound can provide a cue to the location of important information that may improve understanding or modify action (think of a baby crying, a horn blaring, or a dog barking). In some ways, auditory-evoked saccades have different psychophysical properties than visually-evoked saccades: saccades toward auditory targets tend to be longer in duration, have lower peak velocities, and follow curved rather than linear physical trajectories (see Frens et al., 1995). That said, the two types of information can interact to drive gaze. Saccade latencies toward visual-auditory bimodal targets are shorter than those associated with unimodal targets (see Colonius & Arndt, 2001), a result of multisensory integration (Corneil et al., 2002). Further, while fixations can be biased toward parts of a visual scene corresponding to the source of a sound, this bias is dependent on visual scene properties, such as visual salience (Quigley et al., 2008). Thus, the integration of visual and auditory information is possible, with the

✉ Kristina Krasich
kristina.krasich@duke.edu

[1] Center for Cognitive Neuroscience, Duke Institute for Brain Sciences, Duke University, 308 Research Drive Room C03E, LSRC, Durham, NC 27708, USA

[2] University of Notre Dame, Notre Dame, IN, USA

[3] Leidos, Inc., Reston, VA, USA

[4] University of Würzburg, Würzburg, Germany

Published online: 22 June 2021

properties of each stimulus affecting where and when observers move their eyes.

Music is a particularly frequent auditory stimulus that is encountered through much of our daily lives and can, like information in the visual world, vary in its importance with respect to our ongoing tasks and interests. In film, for example, music is used intentionally to complement the storyline and affect viewer experience (see Cohen, 2014, for a review). Indeed, musical soundtracks can influence the emotional impact, interpretation, and remembering of visual information (e.g., Cohen, 2001) and manipulations of film soundtracks can affect both saccade timing and fixation placement. A variety of studies have shown, for example, that saccadic frequency, fixation duration, spatial spread of fixations, and scan paths across a scene are subject to alteration by the accompanying musical soundtrack (Auer et al., 2012; Coutrot et al., 2012; Mera & Stumpf, 2014; Wallengren & Strukelj, 2015; but see Batten & Smith, 2018). Hence, in situations where music is task-related, it can interact with visual content to determine various aspects of gaze allocation.

In many real-world situations, however, the music we hear is task irrelevant. Background music while driving or dining in a restaurant are two common examples. One set of intriguing findings has suggested that even task-irrelevant music may alter gaze control. Specifically, Schäfer and Fachner (2015) presented participants with an image of a house by the sea (45 s) or a short film clip (75 s) of a videotaped road trip on an empty road through an open landscape. Participants freely viewed their assigned stimulus in silence, while listening to an unknown instrumental composition, or while listening to their self-furnished favorite music. For both of the visual stimuli, longer fixations, fewer saccades, and more blinks were observed among participants listening to music relative to those engaged in silent viewing, but no reliable differences in gaze behavior were observed between music types. These findings seem inconsistent with a functional account of music and gaze control, as the music was not specifically orchestrated to guide viewers' gaze. The authors instead hypothesized that listening to music, regardless of music preference, reduced eye movements during scene viewing due to an attentional shift from perceiving the external environment to the processing of inward cognitive and emotional experiences elicited by the music (cf. Fachner, 2011; Herbert, 2011, 2012).

Standing in contrast to Schäfer and Fachner's (2015) demonstration of a relationship between task-irrelevant music and the temporal allocation of overt attention to a scene, Franěk, Šefara, Petružálek, Mlejnek, and van Noorden (2018) have more recently reported a failure to observe effects of music on fixation duration and frequency in a free-viewing task. Our first goal in this report, therefore, was to evaluate prior evidence that the presence of task-irrelevant music can alter various aspects of gaze control timing (i.e., number of fixations, fixation duration, number of saccades, and saccade duration). Franěk and colleagues suggested the resolution of eye tracking equipment or specific experimental procedures may be important factors to address in future research. Hence, we revisited the temporal aspects of gaze control by using eye tracking equipment with high temporal and spatial resolution while employing a more defined memorization task.

Our second goal was to substantially expand upon prior work by considering potential effects of music on overt gaze behaviors that are linked to spatial selection. The temporal and spatial aspects of gaze control are intimately related, so much so that saccade timing alone can successfully predict the locations in a scene where people look (Tatler et al., 2017) and spatial fixation probabilities can be used to predict fixation durations (Einhäuser & Nuthmann, 2016). The systematic changes to saccade timing that have been shown to occur when listening to music may therefore be accompanied by alterations to the selection of fixation points. To evaluate this possibility, we first considered spatial aspects of gaze that can be measured without reference to underlying scene content—namely, saccade amplitude (the subtended distance between two consecutive fixation points) and fixation dispersion (a measure of the spatial extent or "spread" of fixations). We then considered the potential effects of music on spatial aspects of gaze in which the locus of each fixation was considered with respect to the overlapping scene content. For these analyses, we operationalized scene content in terms of objective visual content as well as subjective information value (see Methods for details). This allowed us to evaluate the possibility that the visual system systematically changes *what* visual information is sampled under different auditory conditions and, by doing so, provide the most comprehensive analysis of the effects of task-irrelevant music on gaze to date.

Method

Participants

Sample size was guided by Schäfer and Fachner (2015), who reported relatively large effect sizes of music on gaze (averaged across dependent measures, Cohen's $f = .42$). With three music conditions (described below), an a priori power analysis indicated a sample of approximately 60 participants would provide 80% power to detect similar sized effects of music in our study. We oversampled this estimate and ultimately recruited a group of 75 University of Notre Dame undergraduates to participate in the study for monetary compensation ($5) or course credit (58 self-identified as female; 17 self-identified as male; M_{age} = 19.9 years, SD = 1.6 years). No inclusion or exclusion criteria were established with respect to musical training or experience.[1] Six participants were

[1] Following the experiment, participants in the music conditions were asked to self-report the extent of their formal training in classical music; 60% reported no experience, 22% reported less than 3 years, and 18% reported more than 3 years.

excluded from analysis after completing the study due to errors in calibration and/or data loss from poor-quality eye tracking. Informed consent was obtained from each participant, and the University of Notre Dame Institutional Review Board approved all experimental procedures.

Stimuli and apparatus

The visual stimuli (see Appendix Fig. 6) consisted of 24 digitized color photographs of real-world urban scenes displayed either in full at a resolution of 800 × 600 pixels or in part by extracting a smaller 200 × 200-pixel portion of the original images (vignettes). The stimuli were presented in 32-bit color on a 20-inch CRT monitor with a screen refresh rate of 85 Hz and resolution of 1,024 × 768 pixels. Urban scenes were chosen as stimuli because they tend to elicit more eye movement activity than natural scenes (Berto et al., 2008; Franěk, Šefara, Petružálek, Cabal, & Myška, 2018; Valtchanov & Ellard, 2015). The auditory stimuli consisted of twenty-four 75-second excerpts of classical and modern-classical music. Twelve of these excerpts were taken from three string quartets (Op. 33 No. 2, Op. 76 No. 3, and Op. 74 No. 4) by Joseph Haydn (1732–1809). The remaining 12 excerpts were drawn from two string quartets (No. 3 and No. 4) by Arnold Schoenberg (1874–1951). The tempos of the Haydn excerpts (M = 147 bpm, SD = 32 bpm) and the Schoenberg excerpts (M = 136 bmp, SD = 16 bpm) did not differ (p = .30). Each excerpt was also characterized by a similar degree of variety in dynamic range (piano to forte). The Haydn quartets were performed by the Kodály Quartet and produced by Naxos Records, and the Schoenberg quartets were performed by the New Vienna String Quartet and produced by Philips Records. A 7-second fade-in and fade-out was used to avoid sudden starts and stops of the music. Auditory stimulus volume was adjusted to provide each participant with a comfortable listening experience.

Participants' eye movements were sampled at a rate of 1000 Hz using an EyeLink 2K tower mounted eye-tracking system (SR Research, Inc.). A chin and forehead rest were used to stabilize the head and to maintain a viewing distance of 80 cm. The eye tracker was calibrated using a 9-point calibration at the beginning of the study. A 1-point calibration was used before each trial to correct for drift in the eye tracker signal relative to spatial position over time. Participants listened to the music excerpts through stereo headphones.

Design and procedure

The study was divided into two phases (see Fig. 1). First, in the *study phase*, participants were shown photographs of 12 real-world scenes for 75 seconds each while their eye movements were recorded. They were instructed to study each picture in preparation for a short test. They were told "the test will quiz your ability to remember whole scenes as well as smaller parts of the scenes, so study each picture carefully and try to remember as much detail as possible" and that the scenes would be presented "for approximately one minute each."

The scenes were viewed individually in a randomized order, separated by a 1-point calibration procedure to correct for subtle drift in the eye tracker signal over time. As a between-subjects manipulation, participants were randomly assigned to 1 of 3 possible between-subjects conditions. In the *classical music condition* (n = 22), participants studied each scene while simultaneously listening to a different excerpt of the Hayden quartets. In the *modern-classical music condition* (n = 23), participants listened to the excerpts from the Schoenberg quartets. Including two different music conditions enabled us to examine the robustness of the possible music effect across different musical structures (in this case, tonal and atonal; cf. Schäfer & Fachner, 2015). In both conditions, the scene–music pairings were randomized across participants, so that the scenes shown and the clips of music played were randomly combined for each participant. In the *no music condition* (n = 24), participants studied the scenes without corresponding music. For consistency in the participant experience, those in the no music condition also wore headphones throughout the study phase, although no sound was emitted.

The study phase was immediately followed by the *test phase*. Participants completed a self-paced old/new forced-choice recognition task. Presented individually and in a random order, participants viewed the 12 studied images, 12 unstudied images (foils), 12 vignettes extracted from the studied images, and 12 vignettes extracted from the foil images. The foil images subjectively matched the studied images in terms of location (e.g., city, architectural style), perspective (e.g., skylines, street views), and content (e.g., shop fronts, crowds, roadways). On each trial participants made a yes/no decision to one of two questions "Is this a picture you studied before?" (when a full scene was presented) or "Is this a piece of a picture you studied before?" (when a vignette was presented). Eye movements were not recorded during the test phase.

Gaze-based dependent variables

Temporal measures When considering the temporal aspects of gaze control, we analyzed the relationship between various measures of saccadic timing and the presence of music (cf. Schäfer & Fachner, 2015). These measures included the average: number of fixations, fixation duration, number of saccades, and saccade duration observed on a trial.[2]

[2] Measures of frequency and duration are not orthogonal. For example, within a given amount of time, an increase in fixation durations should correlate with a decrease in the number of fixations observed. The measures do, however, characterize gaze in different ways and consistency across non-orthogonal measures enables stronger conclusions to be drawn from the data.

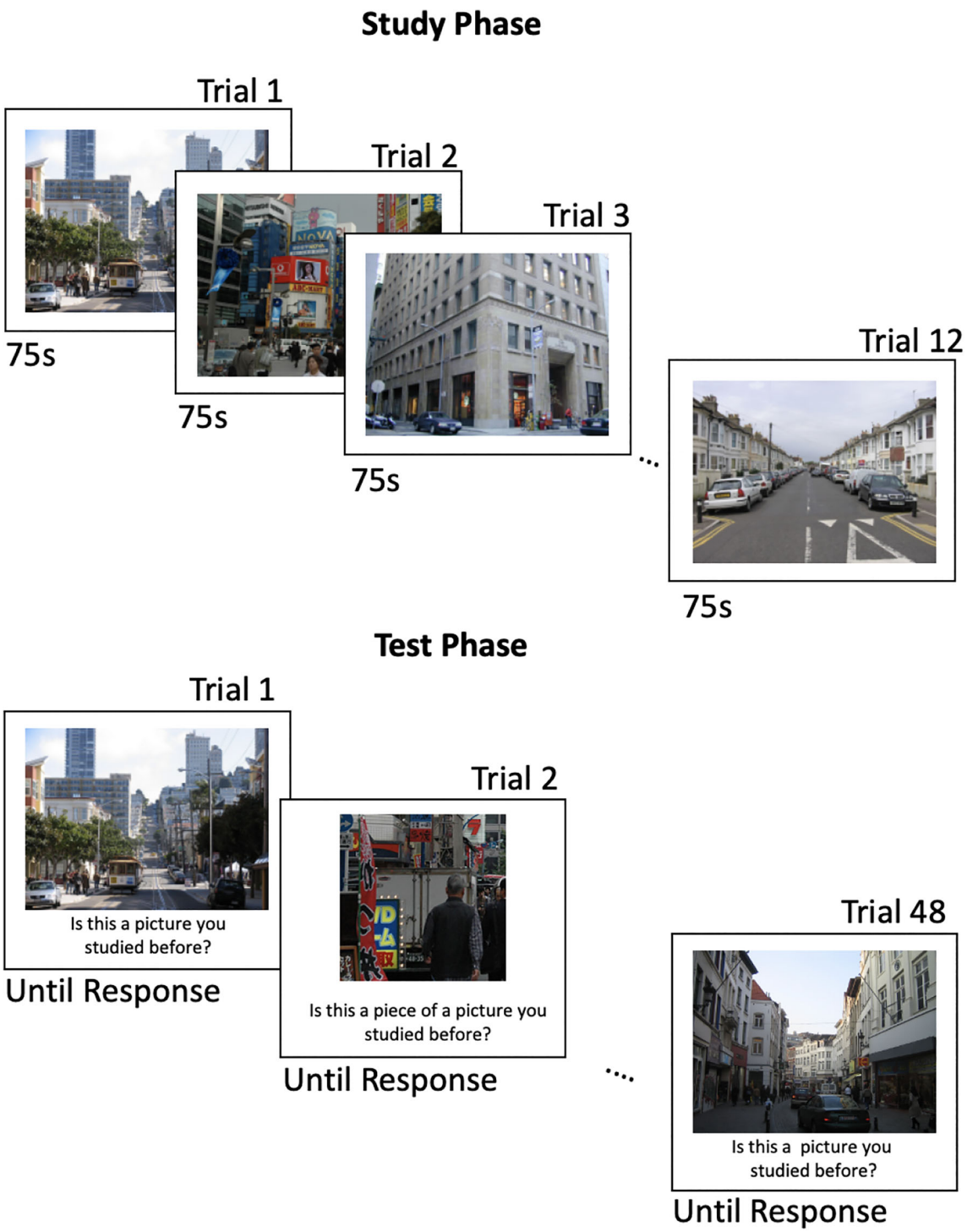

Fig. 1 Example trial sequences in the study phase and test phase. In the study phase, participants viewed each full scene individually for 75 s each. Scenes were presented in a different random order for each participant. In the test phase, participants completed an old/new recognition test by reporting whether or not a given full scene (as in Trial 1 and 48) or a scene vignette (as in Trial 2) was previously seen during the study phase

Spatial measures Content-independent spatial aspects of gaze were measured in terms of saccade amplitude and fixation dispersion. *Saccade amplitude* is the average subtended distance between any two consecutive fixation points. *Fixation dispersion* is a measure of the spatial extent or spread of fixations. It is computed as the root mean square of the Euclidean distance from each fixation to the average position of all fixations.

Values are reported on a 0 –1 scale, with higher values indicating greater dispersion of fixations across the scenes.

Content measures

Content-dependent measures of gaze control consider the information available at each fixation point. We considered both objective visual information and subjective semantic content. Objective visual content was measured in terms of *visual salience*. Visual salience computationally denotes the visual distinctiveness of any given location relative to surrounding locations among features such as luminance, contrast, color, and edge orientation. Salience can be topographically represented by generating *saliency maps* that denote the salience of every pixel in the image (see Fig. 2). Salience, however, is not a singular construct and several different approaches have been developed to formalize its computation (e.g., Borji & Itti, 2013; Harel et al., 2007; Hou et al., 2012; Judd et al., 2012; Riche et al., 2013; Walther & Koch, 2006). Here, we chose to use the Adaptive Whitening Saliency model (AWS; Garcia-Diaz et al., 2012).[3] The AWS model is biologically motivated by the idea that the nonlinear neural responses in the visual cortex should be considered as collective neuron populations rather than as single units (decorrelation of neural responses; e.g., Olshausen & Field, 2005). It also assumes that low-level information is carried by high-order statistical structures and adopts a hierarchical approach to statistically whiten low-level features and remove second-order information (i.e., decorrelation and contrast normalization). The AWS model uses L*a*b* color space, which reduces the correlation between color components. Then log-Gabor filters are used to transform luminance into multiscale and multioriented representations, which are subsequently decorrelated using a principal component analysis (PCA). The final saliency map is computed by taking the sum of the squared norm vectors in the final representation and normalizing it to the sum across all pixels of the image. Thus, visual salience in the AWS represents a global decorrelation of the entire image.

Operationalizing the information content available in a scene depends upon the acquisition of data regarding people's subjective interpretation of scene regions. The subjective evaluation of scene content was approached in two ways, each of which characterized the semantic content of the scenes differently. Our first approach employed *meaning maps* (Hayes & Henderson, 2019; Henderson & Hayes, 2017, 2018) which reveal the degree to which locally informative (i.e., recognizable) information is present within small regions (patches) of a scene independently of overall scene context. To generate these maps (see Fig. 2), third-party observers rate how informative or recognizable information is within small patches of the scene, and then patches are interpolated to produce a cohesive map so that each pixel within a scene contains a semantic value. Following procedures described by Hayes and Henderson (2019), we decomposed each of our scenes into partially overlapping circular patches with 3° ("fine" patches) and 7° ("coarse" patches) diameters. The full patch stimulus set consisted of 3,600 unique fine patches and 960 course patches for a total of 4,560 patches. Then, 150 volunteers from Amazon Mechanical Turk (MTurk) each viewed 300 individual and randomly selected patches (for a total of 40,500 ratings) and rated them in terms of how informative or recognizable they were according to a 6-point Likert scale (*very low*, *low*, *somewhat low*, *somewhat high*, *high*, *very high*). Thus, these ratings were made independently of overall scene context or identity. The resulting values for each pixel at each scale (fine and coarse) were averaged to produce a fine and course rating map for each scene, which were then averaged together into a single map. This map was smoothed with a Gaussian filter (using the *imgaussfilt* function in MATLAB; The MathWorks, Natick, MA) and normalized so that the sum meaning score of all pixels within each image was equal to 1.

Our second approach to operationalizing the subjective interpretation of scenes is one which we will refer to as *semantic interest maps* (Tatler et al., 2017). This approach (see Fig. 2) identifies regions within a scene that are judged to be the most semantically informative locations relative to the entire global scene context. Here, third-party observers subjectively select a set of the most semantically informative regions of a scene while viewing the entire image. Following procedures outlined by Tatler et al. (2017), raters viewed each full scene and selected (with a mouse click) the five most semantically informative locations, ignoring visual characteristics such as color and brightness. Semantically informative locations were defined as the locations that were the most "informative about the meaning of the scene." Data were collected from 31 undergraduate students at the University of Notre Dame. Their selections were used to create semantic interest maps—centering Gaussians with full width at half maximum of two degrees around each selected location. This approach computed an arbitrary value for each pixel of the image, with greater values indicating greater semantic informativeness. As with meaning maps, values were normalized so that the sum score of all pixels was equal to 1 within each image.

[3] Although we discuss one model of salience in depth, we also considered a second model to ensure our conclusions were not biased by specific modeling choices. The Graph Based Visual Salience model (GBVS; Harel et al., 2007) computes, and then combines, multiscale feature maps (i.e., intensity, color, and orientation) via linear center-surround computations that mimic human visual receptive fields. The GBVS also promotes higher saliency values in the center of the image to account for observers' tendency to allocate fixations toward the center of static images. The results obtained with both AWS and GBVS were entirely consistent.

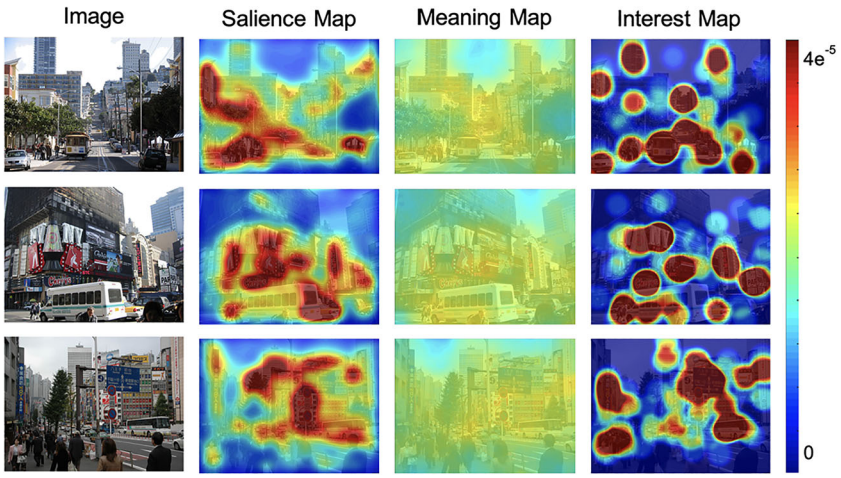

Fig. 2 salience, meaning, and semantic interest maps for three stimuli

Once salience maps, meaning maps, and semantic maps were generated for each scene, (*x, y*) coordinates were extracted for each fixation on each scene and these were overlayed upon each map. An area subtending 2 deg of visual angle (approximating high-fidelity foveal vision) around each fixation location was established and the mean salience and semantic scores among the pixels within each of these areas were calculated. Mean values were then centered and scaled (*z*-scored) using the *scale* function in R (Becker et al., 1988).

Data trims

We established several a priori criteria for including gaze-based data in our analyses. Fixations that occurred outside the scene borders (8.0%), fixation durations under 50 milliseconds or over 2,000 milliseconds (6.2% of fixations), saccade durations over 200 milliseconds (8.9% of saccades), and saccade amplitudes over 20 degrees (<1% of saccades) were excluded. Following these trims, 139,101 fixation samples (87%) and 140,608 saccade samples (88%) were included in the analyses.

Results

Differences in each dependent measure of gaze were considered across both trial viewing time and music conditions. To do so, each trial was divided into 15 5-second time windows. Within each window, average values for each gaze variable were obtained. Gaze variables were then submitted to separate 3 (music condition) × 15 (time interval) mixed model analyses of variance (ANOVA).[4] Corresponding Bayesian ANOVAs were also conducted to characterize the predictive accuracy of the null and alternative hypotheses (e.g., Morey et al., 2016).

Temporal measures of gaze Data are illustrated in Fig. 3. Table 1 summarizes the inferential statistics obtained from the ANOVA analyses. Full model comparisons from the Bayesian ANOVAs are reported in the Appendix. The ANOVA revealed reliable effects of time for each gaze variable. As time progressed into viewing, we observed fewer fixations, longer fixation durations, fewer saccades, and shorter saccade durations. In no case, however, were reliable main effects or interactions involving music observed. Furthermore, with respect to music, Bayes Factors indicated substantial evidence for the null hypothesis for each dependent variable (i.e., $BF_{01} > 3$; see Jeffreys, 1961; Raftery, 1995; Wetzels et al., 2011). Thus, in contrast to prior demonstrations (Schäfer & Fachner, 2015), the presence and type of music played during viewing had no observable impact on temporal measures of gaze behavior.

Spatial measures of gaze. Data are illustrated in Fig. 4. Table 2 summarizes the inferential statistics obtained from the ANOVA analyses, and full model comparisons from the Bayesian ANOVAs are again reported in the Appendix. The ANOVA revealed several reliable effects of time. As the trial progressed, saccade amplitudes increased, and both the visual salience and semantic informativeness associated with fixated scene regions decreased. There was no main effect of time for fixation dispersions. Further, as with the analysis of the temporal measures of gaze behavior, there were no reliable main effects or interactions observed for music. Additionally, with the exception of saccade amplitude (where evidence for the null hypothesis was modest), Bayes factors indicated substantial evidence for the null hypothesis across all dependent measures (i.e., $BF_{01} > 3$). Thus, the presence and type of music played during viewing had no observable impact on the selection of fixated scene regions.

[4] Collapsing the classical and modern-classical conditions into a single group and conducting a 2 (no music vs. music) × 15 (time intervals) ANOVA yielded consistent results.

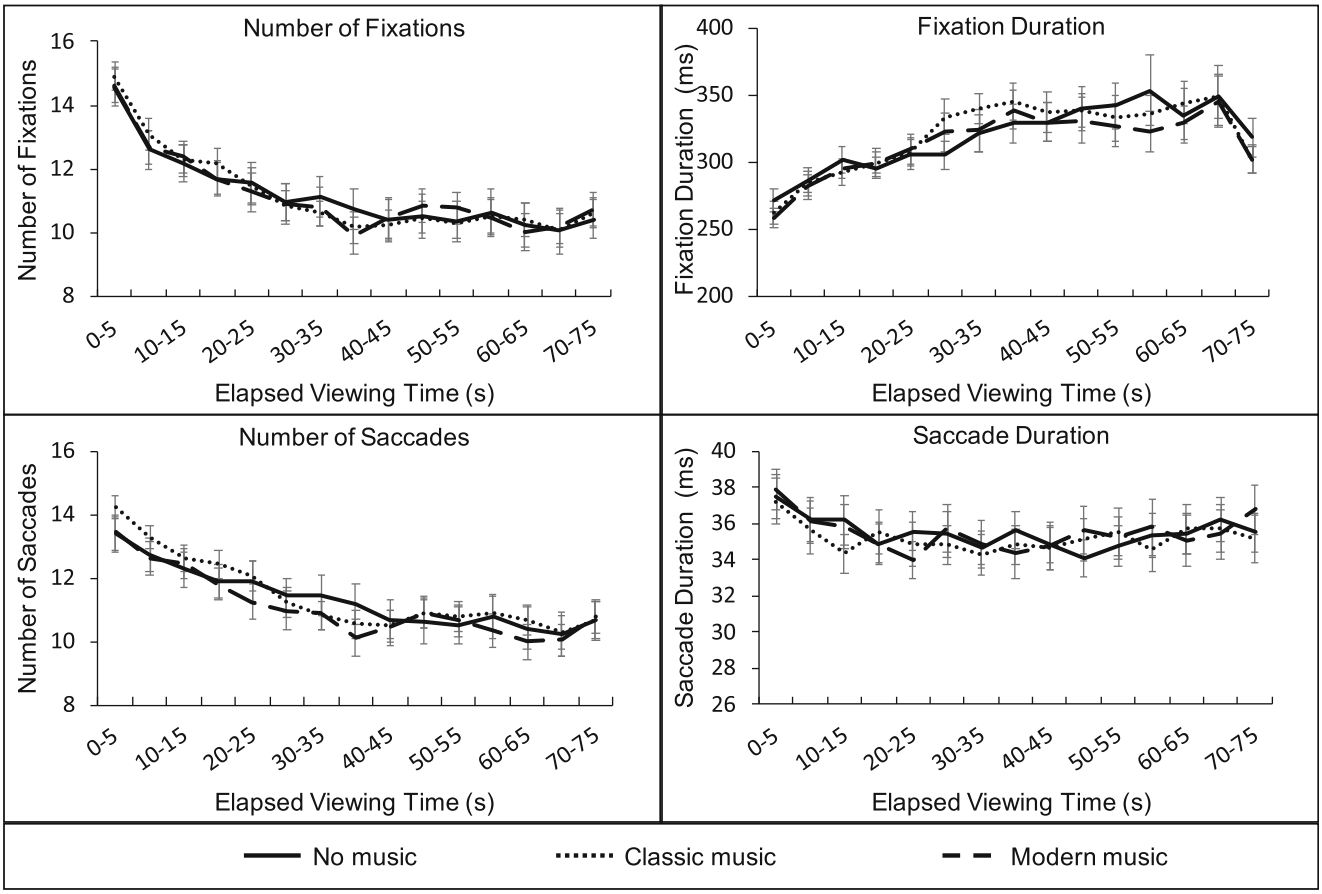

Fig. 3 Temporal aspects of gaze control by music condition plotted over time. Error bars depict +/- 1 standard error of the mean

Table 1 Summary of main effects and interactions for ANOVA analyses of temporal gaze measures, with music as a between-subjects factor and time as a within-subjects factor

Gaze variable	Effect	Statistics (F, p, η_p^2)		
Number of Fixations	Music	$F(2, 66) = .001$	$p = .999$	$\eta_p^2 = .000$
	Time	$\mathbf{F(14, 924) = 74.8}$	$\mathbf{p < .001}$	$\boldsymbol{\eta_p^2 = .531}$
Music × Time	$F(28, 924) = .765$	$p = .806$	$\eta_p^2 = .023$	
Fixation Duration	Music	$F(2, 66) = .073$	$p = .923$	$\eta_p^2 = .002$
	Time	$\mathbf{F(14, 924) = 21.6}$	$\mathbf{p < .001}$	$\boldsymbol{\eta_p^2 = .246}$
	Music × Time	$F(28, 924) = .728$	$p = .848$	$\eta_p^2 = .022$
Number of Saccades	Music	$F(2, 66) = .135$	$p = .874$	$\eta_p^2 = .004$
	Time	$\mathbf{F(14, 924) = 46.4}$	$\mathbf{p < .001}$	$\boldsymbol{\eta_p^2 = .413}$
	Music × Time	$F(28, 924) = .995$	$p = .473$	$\eta_p^2 = .029$
Saccade Duration	Music	$F(2, 66) = .023$	$p = .977$	$\eta_p^2 = .001$
	Time	$\mathbf{F(14, 924) = 4.95}$	$\mathbf{p < .001}$	$\boldsymbol{\eta_p^2 = .070}$
	Music × Time	$F(28, 924) = 1.02$	$p = .444$	$\eta_p^2 = .030$

Note. Statistically reliable effects ($p < .05$) are bolded

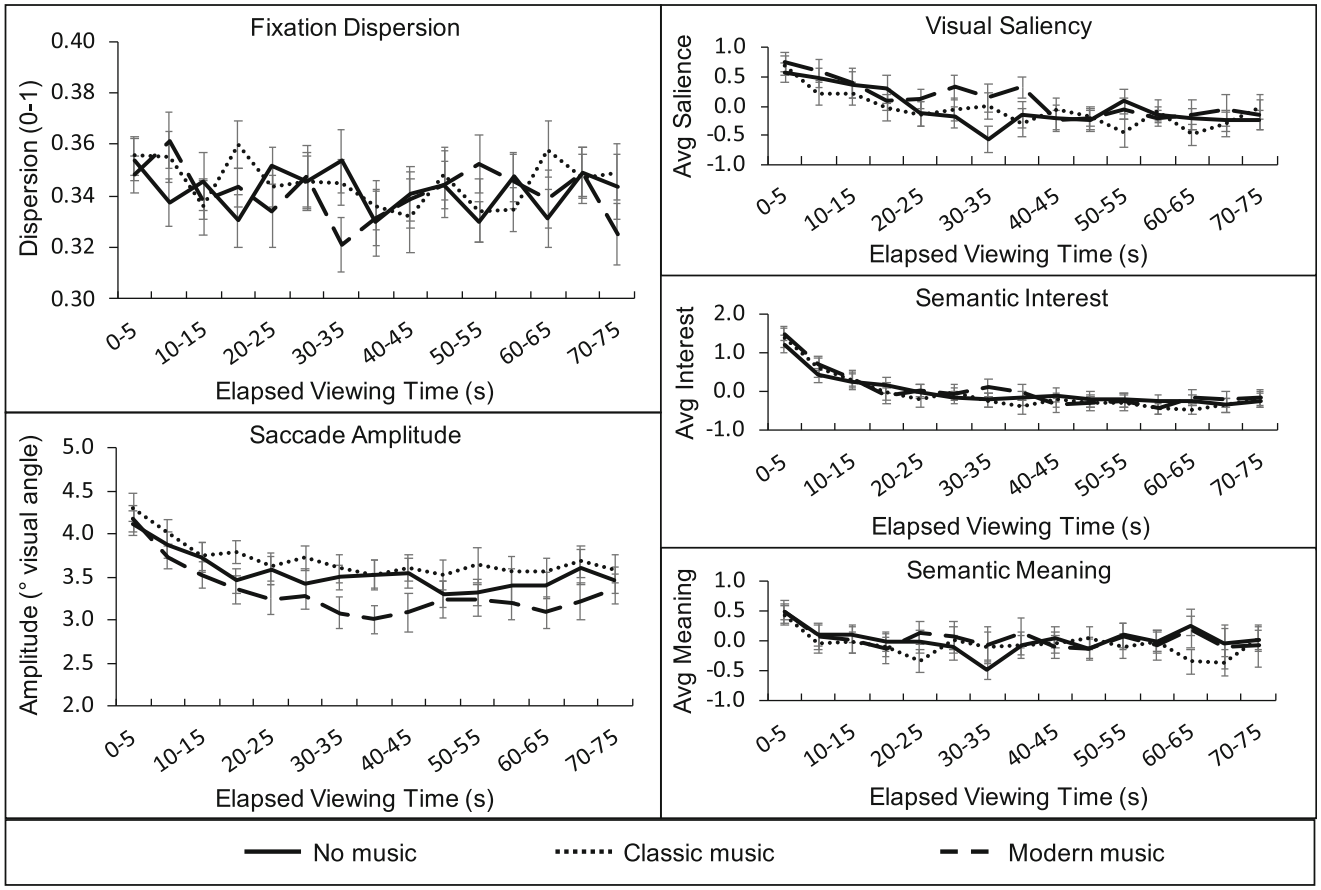

Fig. 4 Spatial aspects of gaze control. Solid, dashed, and dotted lines correspond to the no-music, classical music, and modern-classical music conditions, respectively. Error bars depict +/- 1 standard error of the mean

Memory task performance

While the above analyses focused on gaze behavior during the study phase, we also considered the potential effects of music on memory performance on the old/new recognition task used in the test phase of the study. Using a signal detection theory framework to characterize memory, participants' responses were characterized in terms of hits (defined here as correctly categorizing a studied stimulus as 'old') and false alarms (defined here as incorrectly categorizing a novel stimulus as 'old'). From the resulting hit and false alarm rates, we calculated a', a measure of participants' ability to discriminate between the studied and novel stimuli (in such analyses, an a' value of .5 indicates chance performance while perfect discrimination results in an a' value of 1).

In separate one-way ANOVAs with music condition as a between-subjects factor, we considered participants' ability to discriminate studied and novel scenes presented in their entirety and their ability to discriminate small vignettes extracted from both studied and novel scenes. With respect to whole scenes, we observed a small but reliable effect of condition, $F(2, 66) = 3.31$, $p = .043$, $\eta_p^2 = .09$. While memory was excellent in all conditions, planned comparisons indicated that, together, a small but significant scene memory advantage was observed in the classical ($a' = .989$) and modern-classical ($a' = .990$) music conditions relative to the no music condition ($a' = .968$), $t(66) = 2.57$, $p = .012$, Cohen's $d = 0.58$. The classical and modern-classical music conditions did not differ, $t(66) = 1.09$, $p = .914$, Cohen's $d = 0.05$. With respect to vignettes, a similar numerical pattern was observed among the no music ($a' = .80$), classical music ($a' = .82$), and modern-classical music ($a' = .83$) conditions, but these differences were not statistically reliable, $F(2, 66) = .48$, $p = .619$, $\eta_p^2 = .01$. Hence, our data provides some evidence that listening to music during the study phase may facilitate global scene memory.

Table 2 Summary of main effects and interactions for ANOVA analyses of spatial gaze measures, with music as a between-subjects factor and time as a within-subjects factor

Gaze variable	Effect	Statistics (F, p, η_p^2)		
Fixation Dispersion	Music	$F(2, 66) = .287$	$p = .751$	$\eta_p^2 = .009$
	Time	$F(14, 924) = .883$	$p = .577$	$\eta_p^2 = .013$
	Music × Time	$F(28, 924) = .945$	$p = .548$	$\eta_p^2 = .028$
Saccade Amplitude	Music	$F(2, 66) = 1.61$	$p = .208$	$\eta_p^2 = .046$
	Time	$\mathbf{F(14, 924) = 22.1}$	$\mathbf{p < .001}$	$\boldsymbol{\eta_p^2 = .251}$
	Music × Time	$F(28, 924) = 1.23$	$p = .194$	$\eta_p^2 = .036$
Visual Salience	Music	$F(2, 66) = .476$	$p = .624$	$\eta_p^2 = .014$
	Time	$\mathbf{F(14, 924) = 9.15}$	$\mathbf{p < .001}$	$\boldsymbol{\eta_p^2 = .122}$
	Music × Time	$F(28, 924) = 1.30$	$p = .136$	$\eta_p^2 = .038$
Semantic Interest	Music	$F(2, 66) = .118$	$p = .889$	$\eta_p^2 = .004$
	Time	$\mathbf{F(14, 924) = 36.7}$	$\mathbf{p < .001}$	$\boldsymbol{\eta_p^2 = .358}$
	Music × Time	$F(28, 924) = .824$	$p = .728$	$\eta_p^2 = .024$
Semantic Meaning	Music	$F(2, 66) = .106$	$p = .899$	$\eta_p^2 = .003$
	Time	$\mathbf{F(14, 924) = 3.58}$	$\mathbf{p < .001}$	$\boldsymbol{\eta_p^2 = .051}$
	Music × Time	$F(28, 924) = 1.11$	$p = .317$	$\eta_p^2 = .033$

Note. Statistically reliable effects ($p < .05$) are bolded

Discussion

Our goal in this project was to first conceptually replicate a prior demonstration that task-irrelevant music alters temporal aspects of gaze control during natural scene viewing and to then consider novel effects of music on the spatial allocation of gaze. Observers memorized images of urban scenes in silence or while listening to one of two types of instrumental music. Eye movements were recorded to measure temporal aspects of gaze control—including the number of fixations, fixation duration, number of saccades, and saccade duration—as well as spatial aspects of gaze control—including saccade amplitude, fixation dispersion, and the visual salience and semantic content of fixated locations. Our findings showed that, although the presence of music improved scene memory, there was no evidence that music presence or type affected any of these measures of gaze control.

The first obvious contrast with our study and prior work is our failure to replicate effects of music on some temporal aspects of gaze control (Valtchanov & Ellard, 2015). Our experimental design was sensitive enough to detect relatively subtle changes in gaze measures across time, and our music manipulation was strong enough to have an observable effect on memory performance. Despite this, we could not detect music-related changes in any of the temporal aspects of gaze control we

Table 3 Summary of descriptive (means and standard deviations) and inferential statistics related to postexperiment questionnaire items

Item	Mean rating no music	Mean rating classical music	Meat rating contemp. music	Statistic
This activity did not hold my attention at all.	3.04 (1.65)	2.91 (.921)	2.74 (.964)	$F(2, 66) = .352$, $p = .705$
I felt like I had control over my thoughts	4.71 (1.54)	4.95 (1.40)	4.48 (.846)	$F(2, 66) = .754$, $p = .475$
I allowed my thoughts to wander on purpose.	2.92 (1.82)	2.14 (1.17)	2.43 (1.27)	$F(2, 66) = 1.69$, $p = .193$
I found my thoughts wandering spontaneously.	4.54 (1.56)	4.22 (1.27)	4.09 (1.35)	$F(2, 66) = .650$, $p = .525$

Note. For each item, participants were asked to provide a rating on a 7-point Likert scale. Higher values correspond to stronger endorsements. Data were analyzed with separate one-way ANOVAs

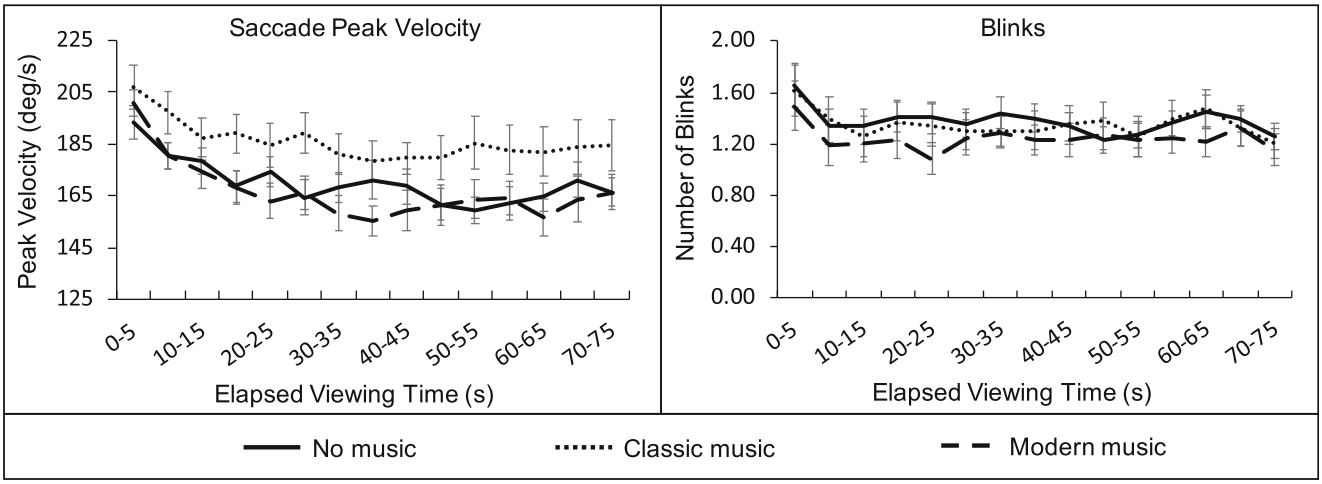

Fig. 5 Oculomotor aspects of gaze control associated with cognitive effort. Solid, dashed, and dotted lines correspond to the no-music, classical music, and modern-classical music conditions, respectively. Error bars depict +/- 1 standard error of the mean

measured. Our null findings specifically related to fixation duration and saccade frequency seem to confirm another recent failure to observe an effect of task-irrelevant music on these aspects of gaze (Franěk, Šefara, Petružálek, Mlejnek, & van Noorden, 2018). The reasons for the divergent findings presented by Schäfer and Fachner (2015) on the one hand, and Franěk, Šefara, Petružálek, Mlejnek, and van Noorden (2018) and the current study on the other hand are not clear. Franěk, Šefara, Petružálek, Mlejnek, and van Noorden (2018) postulated that their replication failure may have stemmed from limitations of their eye tracking equipment or unknown variations in musical experience or preferences. The first of those possibilities seems unlikely in light of our study which brought faster temporal sampling (1000 Hz vs. 60 Hz) and spatial accuracy (±.15 deg vs. ±.4 deg) to the recording of eye movements, but the others remain viable options. We think variations in listeners' instructions or task goals might be additional factors, but our study provides consistency of findings across a memorization task and a free-viewing task (Franěk, Šefara, Petružálek, Mlejnek, & van Noorden, 2018).

Another intriguing possibility to consider is that task-irrelevant music may not directly impact temporal aspects of gaze control. Schäfer and Fachner (2015) proposed that music may encourage attention to perceptually decouple from external inputs and turn inward toward internal experiences elicited by the music (cf. Fachner, 2011; Herbert, 2011, 2012). This deprioritization of external information may in turn impact how gaze is allocated (Faber et al., 2020). Interestingly, the findings from Schäfer and Fachner (2015) are entirely consistent with those that have been observed in studies of mind wandering—another perceptually decoupled state of attention (e.g., Kam & Handy, 2013; Schooler et al., 2011; Smallwood, 2013). That is, mind wandering has been associated with fewer and longer fixations as well as more frequent eyeblinks (Krasich et al., 2018; Reichle et al., 2010; Smilek et al., 2010; Uzzaman & Joordens, 2011; Zhang et al., 2020). Further, it is possible that music actually encourages mind wandering (cf. Koelsch et al., 2019; Taruffi et al., 2017), which would account for the correlational relationship between music and gaze observed in Schäfer and Fachner (2015). We have no way of testing this possibility here, but we suspect higher rates of mind wandering in Schäfer and Fachner's (2015) study because they used a very simple stimulus set and a free-viewing task. Comparatively, our task involved multiple stimuli and was more cognitively demanding, as participants had to memorize visually complex scenes for a later test, and the rate of mind wandering is shown to be inversely related to ongoing task demands (Smallwood & Schooler, 2006). In a postexperiment questionnaire, we did ask our participants to report on their attentiveness during the study phase and observed no differences across conditions (see Table 3). While this is an admittedly insensitive approach to quantifying mind wandering (Murray et al., 2020), these results do suggest that the level of attentiveness was similar across our groups, possibly explaining our failure to replicate

the relationship between task-irrelevant music and temporal aspects of gaze control.

While boundary conditions on the effect task-irrelevant music on gaze have yet to be fully elucidated, our findings suggest that music does not affect overt visual attention automatically and the effects of task-irrelevant music on gaze control may not be as robust as when music (or an auditory stimulus more generally) is task-relevant. Beyond the temporal aspects of gaze, our findings further showed no observable influence on the spatial allocation of gaze within scenes. This suggests that gaze control mechanisms do not "rebalance" salient and semantically informative information when task-irrelevant music is present. This contrasts with situations in which music is integral to the viewing experience, as it is when watching films where music has been shown to influence the content that is viewed (Auer et al., 2012; Coutrot et al., 2012; Mera & Stumpf, 2014; Wallengren & Strukelj, 2015). Thus, our null findings are consistent with a functional account of music on gaze control: Task-relevant music can direct where observers should look (Cohen, 2014), whereas task-irrelevant music provides no such environmental cues. Hence, while models of gaze control require a better integration of sound and vision, our results suggest it would be more fruitful to concentrate this effort on auditory stimulation that is relevant to an observer's task. Our findings also suggest that the effects of task-irrelevant music on cognitive processes, such as time estimation (Cassidy & MacDonald, 2010), skilled motor performance (e.g., Ünal et al., 2012), and decision-making (e.g., Day et al., 2009), may be driven by mechanisms related to motivation or emotion rather than visual information acquisition.

Despite the clear absence of an effect of music on any temporal or spatial components of gaze, we did observe that participants who listened to music performed slightly better on the memory test than those who studied the scenes in silence. While the effects of background music on memory are certainly equivocal (see Kämpfe et al., 2011, for a meta-analysis), this difference in our study might suggest that the participants in the two music conditions exerted more cognitive effort while studying than those who studied in silence. Therefore, post hoc, we explored the possibility that the presence of music affected two oculomotor components of gaze that have been linked to cognitive effort and/or load—namely, blink rate (e.g., Ranti et al., 2020; Shultz et al., 2011) and saccade peak velocity (e.g., Di Stasi et al., 2013). Illustrated in Fig. 5, our findings showed that blinks decreased over viewing time, $F(14, 924) = 6.17$, $p < .001$, but no effect of music was observed, $F(14, 924) < 1$, nor did these factors interact, $F(14, 924) < 1$. Peak velocity also decreased over time, $F(14, 924) = 19.2$, $p < .001$, and a marginal effect of music was observed, $F(2, 66) = 2.54$, $p = .085$, along with a marginal interaction of these factors, $F(14, 924) = 1.37$, $p = .096$. However, Bayesian ANOVAs did not show strong evidence in favor of these effects (i.e., $BF_{01} > 3$). Hence, we conclude that, as with temporal and spatial aspects of gaze, task-irrelevant music failed to meaningfully alter oculomotor components of gaze related to cognitive processing. Thus, the mechanism(s) underlying the memory benefit associated with music in our study appears to be independent of the gaze control system.[5]

In summary, studies of gaze control often fail to account for the multimodal environments. While task-relevant auditory stimuli, such as music, can be orchestrated to guide viewers' behaviors, task-irrelevant music may also impact gaze (Schäfer & Fachner, 2015). The current work, however, found no evidence that task-irrelevant ("background") music modified the temporal, spatial, or oculomotor aspects of gaze control related to cognition. These findings are consistent with functional accounts of music on gaze control, but further work is needed to establish the robustness of music's effect on gaze as well as possible underlying mechanisms that may better characterize this relationship.

Author note Portions of this work constituted a senior thesis by J.K. K.K., J.K., and J.R.B. conceived the study. J.K., G.H., and A.L.K. programmed the study and/or contributed methodological tools. J.K. and A.L.K. collected the data. K.K., J.K., G.H., and J.R.B. analyzed the data. K.K., J.K., and J.R.B. wrote the paper. Address correspondence concerning this manuscript to Kristina Krasich, Center for Cognitive Neuroscience, Duke Institute for Brain Sciences, 308 Research Drive Room C03E, LSRC, Durham, NC 27708, or via email to kristina.krasich@duke.edu

[5] In follow-up questionnaires, participants in the no music rated the task as more "boring" (4.29 on a 7-point Likert scale where higher ratings indicate higher levels of boredom) than participants in the classical music (3.41) and modern-classical music (3.65) conditions, $F(2, 66) = 3.14$, $p = .050$. While this difference was not associated with gaze, it is possible that differing levels of experienced boredom underlies the differences in memory performance.

Appendix 1

Fig. 6 The complete stimulus set. White boxes indicate the vignettes taken from each of the full scene photographs

Appendix 2

All analyses were conducted with the software package JASP (Version 0.14.1) using default priors. The models are ordered by their predictive performance in reference to the best model.

Fixation count					
Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Time	.20	.784	14.54	1.00	
Time + Condition	.20	.216	1.10	3.64	8.14
Time + Condition + Time× Condition	.20	2.75 e −4	.00	2847.14	48.51
Null model (incl. subject)	.20	3.30 e−140	1.32 e−139	2.38 e+139	0.75
Condition	.20	6.25 e−141	2.50 e−140	1.25 e+140	3.58

Note. All models include subject; P(M) = prior model probability, P(M|data) = posterior model probability; BF_M = change from prior to posterior model odds

Fixation duration					
Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Time	.20	.87	26.20	1.00	
Time + Condition	.20	.132	.61	6.56	6.25
Time + Condition + Time× Condition	.20	1.64 e−4	6.56 e −4	5289.12	38.03
Null model (incl. subject)	.20	5.91 e−46	2.37 e−45	1.47 e+45	0.34
Condition	.20	8.49 e−47	3.39 e−46	1.02 e+46	3.47

Note. All models include subject; P(M) = prior model probability, P(M|data) = posterior model probability; BF_M = change from prior to posterior model odds

Saccade count					
Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Time	.20	0.78	13.881	1.000	
Time + Condition	.20	0.22	1.146	3.485	7.334
Time + Condition + Time× Condition	.20	9.636 e−4	0.004	805.622	41.868
Null model (incl. subject)	.20	2.361e−94	9.443e−94	3.288e+93	0.266
Condition	.20	5.157e−95	2.063e−94	1.505e+94	3.498

Note. All models include subject; P(M) = prior model probability, P(M|data) = posterior model probability; BF_M = change from prior to posterior model odds

Saccade duration					
Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Time	.20	.781	14.29	1.00	
Time + Condition	.20	.218	1.11	3.59	8.95
Time + Condition + Time× Condition	.20	8.96 e−4	.00	871.89	48.94
Null model (incl. subject)	.20	1.11 e−6	4.45 e−6	702396.67	.20
Condition	.20	3.06 e−7	1.22 e−6	2.56 e+6	3.65

Note. All models include subject; P(M) = prior model probability, P(M|data) = posterior model probability; BF_M = change from prior to posterior model odds

Saccade amplitude					
Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Time	.20	.60	5.93	1.00	
Time + Condition	.20	.40	2.63	1.51	4.57
Time + Condition + Time× Condition	.20	.01	.03	93.24	1.64
Null model (incl. subject)	.20	4.48 e−46	1.79 e−45	1.33 e+45	.30
Condition	.20	2.58 e−46	1.03 e−45	2.32 e+45	.97

Note. All models include subject; P(M) = prior model probability, P(M|data) = posterior model probability; BF_M = change from prior to posterior model odds

Fixation dispersion					
Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Null model (incl. subject)	.20	.96	100.54	1.00	
Condition	.20	.04	.16	25.45	1.14
Time	.20	4.55 e−4	.00	2114.83	.26
Time + Condition	.20	1.77 e−5	7.09 e−5	54290.23	.79
Time + Condition + Time× Condition	.20	5.72 e−8	2.29 e−7	1.68 e+7	.90

Note. All models include subject; P(M) = prior model probability, P(M|data) = posterior model probability; BF_M = change from prior to posterior model odds

Visual salience					
Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Time	.20	.87	27.52	1.00	
Time + Condition	.20	.12	0.56	7.08	.55
Time + Condition + Time× Condition	.20	.00	0.01	249.65	.59
Null model (incl. subject)	.20	9.79 e−17	3.91 e−16	8.92 e+15	.23
Condition	.20	1.27 e−17	5.08 e−17	6.87 e+16	.49

Note. All models include subject; P(M) = prior model probability, P(M|data) = posterior model probability; BF_M = change from prior to posterior model odds

Semantic interest

Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Time	.20	.89	24.07	1.00	
Time + Condition	.20	.14	.66	6.02	6.67
Time + Condition + Time× Condition	.20	1.10 e−4	4.40 e −4	7793.12	49.80
Null model (incl. subject)	.20	4.44 e−77	1.78 e−76	1.9 e+76	.28
Condition	.20	5.19 e−78	2.08 e−77	1.65 e+77	1.03

Note. All models include subject; P(M) = prior model probability, P(M|data) = posterior model probability; BF_M = change from prior to posterior model odds

Semantic meaning

Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Time	.20	.87	27.29	1.00	
Time + Condition	.20	.13	0.57	6.99	10.29
Time + Condition + Time× Condition	.20	.00	.01	539.69	0.24
Null model (incl. subject)	.20	.00	.01	730.05	31.86
Condition	.20	2.45 e−4	9.81 e−4	3555.88	3.50

Note. All models include subject; P(M) = prior model probability, P(M|data) = posterior model probability; BF_M = change from prior to posterior model odds

Blink rate

Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Time	.20	.72	10.29	1.00	
Time + Condition	.20	.28	1.55	2.58	7.48
Time + Condition + Time× Condition	.20	3.14 e −4	.00	2296.95	41.97
Null model (incl. subject)	.20	7.81 e−10	3.12 e−9	9.23 e+8	.19
Condition	.20	2.70 e−10	1.08 e−9	2.67 e+9	3.60

Note. All models include subject; P(M) = prior model probability, P(M|data) = posterior model probability; BF_M = change from prior to posterior model odds

Saccade peak velocity

Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Time	.20	.526	4.44	1.00	
Time + Condition	.20	.450	3.28	1.17	6.75
Time + Condition + Time× Condition	.20	.023	.10	22.50	7.29
Null model (incl. subject)	.20	1.58 e−39	6.31 e−39	3.34 e+38	7.25
Condition	.20	1.40 e−39	5.59 e−39	3.77 e+38	6.75

Note. All models include subject; P(M) = prior model probability, P(M|data) = posterior model probability; BF_M = change from prior to posterior model odds

References

Auer, K., Vitouch, O., Koreimann, S., Pesjak, G., Leitner, G., & Hitz, M. (2012). When music drives vision: Influences of film music on viewers' eye movements. *Proceedings of the 12th International Conference on Music Perception and Cognition*, 73–76.

Batten, J., & Smith, T. J. (2018). Looking at sound: Sound design and the audiovisual influences on gaze. In T. Dwyer, C. Perkins, S. Redmond, & J. Sita (Eds.), *Seeing into screens: Eye tracking and the moving image.* Bloomsbury.

Becker, R. A., Chambers, J. M., & Wilks, A. R. (1988) *The new S language*. Wadsworth & Brooks/Cole.

Berto, R., Massaccesi, S., & Pasini, M. (2008). Do eye movements measured across high and low fascination photographs differ? Addressing Kaplan's fascination hypothesis. *Journal of Environmental Psychology, 28,* 185–191.

Borji, A., & Itti, L. (2013). State-of-the-Art in Visual Attention Modeling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*, 185–207.

Cassidy, G. G., & MacDonald, R. A. R. (2010). The effects of music on time perception and performance of a driving game. *Scandinavian Journal of Psychology, 51,* 455–464.

Cohen, A. J. (2001). Music as a source of emotion in film. In P. Juslin & J. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 249–272).

Cohen, A.J. (2014). Film music from the perspective of cognitive science. In D. Neumeyer (Ed.), *The Oxford handbook of film music studies*. Oxford University Press.

Colonius, H., & Arndt, P. (2001). A two-stage model for visual-auditory interaction in saccadic latencies. *Perception & Psychophysics*, *63*, 126–147.

Corneil, B. D., Van Wanrooij, M., Munoz, D. P., & Van Opstal, A. J. (2002). Auditory- visual interactions subserving goal-directed saccades in a complex scene. *Journal of Neurophysiology*, *88*, 438–454.

Coutrot, A., Guyander, N., Ionescu, G., & Caplier, A. (2012). Influence of soundtrack on eye movements during video exploration. *Journal of Eye Movement Research, 5*(2), 1–10.

Day, R.-F., Lin, C.-H., Huang, W.-H., & Chuang, S.-H. (2009). Effects of music tempo and task difficulty on multi-attribute decision-making: An eye-tracking approach. *Computers in Human Behavior, 25,* 120–143.

Di Stasi, L. L., Marchitto, M., Antoli, A., & Canas, J. J. (2013). Saccade peak velocity as an alternative index of operator attention: Short review. *European Review of Applied Psychology, 63,* 335–343.

Einhäuser, W., & Nuthmann, A. (2016). Salient in space, salient in time: Fixation probability predicts fixation duration during natural scene viewing. *Journal of Vision*, *16*, 13–13.

Faber, M., Krasich, K., Bixler, R. E., Brockmole, J. R., & D'Mello, S. K. (2020). The eye–mind wandering link: Identifying gaze indices of mind wandering across tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *46*, 1201.

Fachner, J. (2011). Time is the key - music and altered states of consciousness. In E. Cardenas, M. Winkelmann, C. Tart, & S. Krippner (Eds.), *Altering consciousness: A multidisciplinary perspective. Vol. 1: History, culture and the humanities* (pp. 355–376). Praeger.

Franěk, M., Šefara, D., Petružálek, J., Cabal, J., & Myška, K. (2018). Differences in eye movements while viewing images with various levels of restorativeness. *Journal of Environmental Psychology, 57,* 10–16.

Franěk, M., Šefara. D., Petružálek, J., Mlejnek, R., & van Noorden, L. (2018). Eye movements in scene perception while listening to slow and fast music. *Journal of Eye Movement Research, 11*(2), 8.

Frens, M. A., Van Opstal, A. J., & Van der Willigen, R. F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Perception & Psychophysics*, *57*, 802–816.

Garcia-Diaz, A., Leboran, V., Fdez-Vidal, X. R., & Pardo, X. M. (2012). On the relationship between optical variability, visual saliency, and eye fixations: A computational approach. *Journal of Vision*, *12*, 17–17.

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. In B. Schölkopf, J. Platt, & T. Hofmann (Eds.), *Advances in neural information processing systems* (pp. 545–552). MIT Press. https://doi.org/10.7551/mitpress/7503.001.0001

Hayes, T. R., & Henderson, J. M. (2019). Scene semantics involuntarily guide attention during visual search. *Psychonomic Bulletin & Review*, *26*, 1683–1689.

Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour*, *1*, 743.

Henderson, J. M., & Hayes, T. R. (2018). Meaning guides attention in real-world scene images: Evidence from eye movements and meaning maps. *Journal of Vision*, *18*, 10–10.

Herbert, R. (2011). *Music listening: Absorption, dissociation and trancing*. Ashgate.

Herbert, R. (2012). Musical and non-musical involvement in daily life: The case of absorption. *Musicae Scientiae, 16,* 41–66.

Hollingworth A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 58–69.

Hou, X., Harel, J., & Koch, C. (2012). Image signature: Highlighting sparse salient regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*, 194–201.

Jeffreys, H. (1961). Theory of Probability (3rd Ed.). Oxford University Press.

Judd, T., Durand, F., & Torralba, A. (2012). *A benchmark of computational models of saliency to predict human fixations*. (Tech. Rep. No. MITCSAIL-TR-2012-001). Cambridge, MA: MIT Computer Science and Artificial Intelligence Laboratory.

Kam, J. W., & Handy, T. C. (2013). The neurocognitive consequences of the wandering mind: a mechanistic account of sensory-motor decoupling. *Frontiers in Psychology*, *4*, 725.

Kämpfe, J., Sedlmeier, P., & Renkewitz, F. (2011). The impact of background music on adult listeners: A meta-analysis. *Psychology of Music*, *39*, 424–448.

Koelsch, S., Bashevkin, T., Kristensen, J., Tvedt, J., & Jenstschke, S. (2019). Heroic music stimulates empowering thoughts during mind-wandering. *Scientific Reports, 9,* Article 10317.

Krasich, K., McManus, R., Hutt, S., Faber, M., D'Mello, S. K., & Brockmole, J. R. (2018). Gaze-based signatures of mind wandering during real-world scene processing. *Journal of Experimental Psychology: General*, *147*, 1111–1124.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, *369* (6483), 742–744.

Mera, M., & Stumpf, S. (2014). Eye-tracking film music. *Music and the Moving Image, 7,* 3–23.

Morey, R. D., Romeijn, J. W., & Rouder, J. N. (2016). The philosophy of Bayes factors and the quantification of statistical evidence. *Journal of Mathematical Psychology*, *72*, 6–18.

Murray, S., Krasich, K., Schooler, J. W., & Seli, P. (2020). What's in a task? Complications in the study of the task-unrelated-thought variety of mind wandering. *Perspectives on Psychological Science, 15*(3), 572–588.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 43*, 387–391. *Vision Research, 10,* 1286–1294.

Najemnik, J., & Geisler, W. S. (2009). Simple summation rule for optimal fixation selection in visual search. *Vision Research, 10,* 1286–1294.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research. 46,* 614–621.

Olshausen, B. A., & Field, D. J. (2005). How close are we to understanding V1?. *Neural Computation*, *17*, 1665–1699.

Parkhurst, D. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*, 107–123.

Quigley, C., Onat, S., Harding, S., Cooke, M., & Konig, P. (2008). Audio-visual integration during overt visual attention. *Journal of Eye Movement Research, 1*(4)*, 1–17*.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological Methodology 1995* (pp. 111–196). Blackwell.

Ranti, C., Jones, W., Klin, A., & Shultz, S. (2020). Blink rate patterns provide a reliable measure of individual engagement with scene content. *Scientific Reports, 10,* Article 8267.

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science, 21,* 1300 – 1310.

Riche, N., Mancas, M., Duvinage, M., Mibulumukini, M., Gosselin, B., & Dutoit, T. (2013). RARE2012: A multi-scale rarity-based saliency detection with its comparative statistical analysis. *Signal Processing: Image Communication*, *28*, 642–658.

Schäfer, T., & Fachner, J. (2015). Listening to music reduces eye movements. *Attention, Perception, & Psychophysics, 77,* 551–559.

Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in cognitive sciences*, *15*, 319–326.

Shinoda, H., Hayhoe, M. M., & Shrivastava, A. (2001). What controls attention in natural environments? *Vision Research*, *41*, 3535–3545.

Shultz, S., Klin, A., & Jones, W. (2011). Inhibition of eye blinking reveals subjective perceptions of stimulus salience. *Proceedings of the National Academy of Sciences, 108,* 21270–21275.

Smallwood, J. (2013). Distinguishing how from why the mind wanders: a process–occurrence framework for self-generated mental activity. *Psychological bulletin*, *139*, 519.

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*, 946.

Smilek, D., Carriere, J. S., & Cheyne, J. A. (2010). Out of mind, out of sight: Eye blinking as indicator and embodiment of mind wandering. *Psychological Science, 21,* 786 –789

Taruffi, L., Pehrs, C., Skouras, S., & Koelsch, S. (2017). Effects of sad and happy music on mind-wandering and the default mode network. *Scientific Reports, 7,* Article 14396.

Tatler, B. W., Brockmole, J. R., & Carpenter, R. H. S. (2017). LATEST: A Model of Saccadic Decisions in Space and Time. *Psychological Review*, *124,* 267–300.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, *113,* 766.

Ünal, A. B., Steg, L., & Epstude, K. (2012). The influence of music on mental effort and driving performance. *Accident Analysis & Prevention*, *48*, 271–278.

Uzzaman, S., & Joordens, S. (2011). The eyes know what you are thinking: Eye movements as an objective measure of mind wandering. *Consciousness and Cognition, 20,* 1882–1886.

Valtchanov, D., & Ellard, C. G., (2015). Cognitive and affective responses to natural scenes: Effects of low level visual properties on preference, cognitive load, and eye-movements. *Journal of Environmental Psychology, 43,* 184–195.

Võ, M. L. H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, *9*, 24–24.

Wallengren, A.-K., & Strukelj, A. (2015). Film music and visual attention: A pilot experiment using eye-tracking. *Music and the Moving Image, 8,* 69–80.

Walther, D. & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, *19,* 1395–1407.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 *t* tests. *Perspectives on Psychological Science, 6,* 291–298.

Yarbus, A. L. (1967). *Eye movements during perception of complex objects*. Plenum.

Zhang, H., Anderson, N. C., & Miller, K. F. (2020). Refixation patterns of mind-wandering during real-world scene perception. *Journal of Experimental Psychology: Human Perception and Performance, 47*(1), 36–52. https://doi.org/10.1037/xhp0000877