# The Effect of Music and Editing Style on Subjective Perception of Time When Watching Videos

## An Eye-tracking Study

**Kathryn Nicole Sam and K. Jayasankara Reddy**

***Abstract:*** Arousal, editing style, and eye movements have been implicated in time perception when watching videos. However, little multimodal research has explored how manipulating both the auditory and visual properties of videos affects temporal processing. This study investigated how editing density and music-induced arousal affect viewers' time perception. Thirty-nine participants watched six videos varying in editing density and music while their eye movements were recorded. They estimated the videos' duration and reported their subjective experience of time passage and emotional involvement. Fast-paced editing was associated with the feeling of time passing faster, a relationship mediated by fixation durations. High-arousal background music was also associated with the feeling of time passing faster. The consequences of this study in terms of a possible auditory driving effect are explored.

***Keywords:*** arousal, audiovisual media, duration judgment, editing style, fixations, passage of time judgment, temporal processing, time perception

Audiovisual stimuli are ubiquitous and an essential component of multi-billion-dollar industries such as those of film, video games, and e-learning. They consist of complex features deliberately constructed to deliver a message or elicit an emotion (Grall and Finn 2022). Through their choices of background music, they can often induce varying states of arousal (Ansani et al. 2021; Droit-Volet et al. 2013). They are also accompanied by visual features such as camera angles, movements, cuts, and lighting. These choices, along with those of dialogue, costume, and acting, influence the emotional involvement of the audience, as well as cognitive processes.

*Projections*

doi: 10.3167/proj.2023.170203
Volume 17, Issue 2, Summer 2023: 41–61
ISSN 1934-9688 (Print), ISSN 1934-9696 (Online)

Scientists are only beginning to describe how the decisions made by filmmakers can influence cognitive phenomena. One cognitive domain of interest has been the perception of time. Time perception has typically been measured in terms of duration judgments (DJs) and rate judgments (Brighouse et al. 2014). DJs involve estimating the duration of a time interval (Balzarotti et al. 2021), while rate judgments involve perceiving the rate of change of a stimulus, such as the tempo of a song or how fast images in a montage change (Boltz 2017). Some researchers have also assessed the subjective experiences of how time passes, known as passage of time judgments (PoTJs) (Droit-Volet and Wearden 2016). PoTJs are independent of DJs. They are judgments on how much time seems to be 'flying' or 'dragging.' The present study assesses DJs and PoTJs.

Traditionally, time perception has been viewed through the lens of pacemaker-accumulator models. In these models, pulses are generated by a pacemaker, which is connected to an accumulator. There is also a switch that can block the collection of pulses (Lake et al. 2016). The more pulses accumulated, the longer the duration of time, and this duration can then be transferred to memory and compared to representations of durations stored in long-term memory (ibid.).

Arousal has frequently been associated with time perception (Lake et al. 2016). It is a bodily state, and, as a component of emotion, is associated with interoceptive cues (Wittman 2014). Jessica Lake, Kevin LaBar, and Warren Meck (2016) believe that pacemaker-accumulator models do not adequately represent what is known about the effects of arousal and attention on time perception. They hence proposed a model called the emotion-induced temporal distortions model. According to this model, physiological arousal as well as attention and working memory demands of emotion-relevant stimuli lead to overestimations of durations, while working memory and attention captured by emotion-irrelevant stimuli lead to underestimations. The effects of physiological arousal take place in an earlier time window after stimulus onset compared to the attention and working memory demands placed by emotional stimuli.

Another theory that considers the effect of emotion, while integrating it with the role of interoceptive cues, is the theory of embodied time. The theory of embodied time implicates bodily states in the experience of time. There is evidence that exposure to postures that involve greater movement leads to overestimations in DJs, a process mediated by arousal (Nather et al. 2011). When individuals view sculptures in different postures, they may create an embodied simulation of the posture, leading to changes in arousal and time perception (ibid.). Oculomotor cues are another kind of interoceptive cue that may influence time perception according to this per-

***The theory of embodied time implicates bodily states in the experience of time.***

spective. The increased rate of gaze shifts caused by watching videos with a higher editing density (more cuts) seems to be associated with longer judgments of duration (Balzarotti et al. 2021). This may be explained either by changes in interoceptive cues or changes in visual attention (ibid.).

### Music-induced Arousal and Time Perception

Higher arousal has been associated with both increased DJs (Ansani et al. 2021; Cocenas-Silva et al. 2011) and the experience of time flying faster (PoTJs) (Droit-Volet and Wearden 2016). One study analyzed the factors influencing time perception when listening to Western classical music (Cocenas-Silva et al. 2011). Forty-eight participants listened to sixteen excerpts varying in musical and emotional properties. Using a multidimensional scaling approach, the researchers found that participants overestimated durations when listening to excerpts with high arousal, loudness, and tempo. Similarly, in another study on time perception, forty participants took part in an experiment where the tempo and emotional valence of musical excerpts were manipulated, while keeping other factors constant (Droit-Volet et al. 2013). Faster tempos, associated with greater subjective arousal, led to time overestimations.

***The effect of arousal on time perception varies depending on the stimulus' sensory modality.***

Research on music and time perception has also been conducted in the context of audiovisual media. In one study, participants watched a short film accompanied by one of four background music tracks or with no sound at all. Once more, higher arousal was associated with overestimations of time durations (Ansani et al. 2021). The effect of arousal on time perception varies depending on the stimulus' sensory modality. In a recent study, participants watched twenty-eight film clips presented in three modalities: audio only, video only, or audiovisual (Appelqvist-Dalton et al. 2022). Contrary to what was found by Alessandro Ansani and colleagues (2021), durations were perceived as longer for visual only clips compared to the other two conditions. Additionally, higher arousal, as measured by both self-report and changes in pupil size, was related to longer DJs. This effect was stronger in the audiovisual modality compared to the unimodal conditions. However, neither of these two studies considered visual factors like editing density.

Arousal has also been studied in the context of PoTJs. Using an experience sampling methodology, Sylvie Droit-Volet and John Wearden (2016) found that arousal was associated with PoTJs in both young adults and elderly individuals. This study, however, did not involve audiovisual stimuli.

### Editing Density and Time Perception

While the research in audiovisual media described thus far manipulated background music, the possible contributions of the visual stimulus toward time perception were not assessed. Addressing this gap, researchers compared the effects of slow-paced videos, fast-paced videos, and a master shot (unedited) on DJs, PoTJs, and emotional involvement (Eugeni et al. 2016). They found that fast-paced editing led to increased PoTJs in comparison to the two other conditions, as well as greater time overestimation compared to the master shot. Fast-paced videos (high editing density) were also rated as more emotionally engaging. These findings were echoed in a study comparing edited (using both continuity and discontinuity editing styles) and uncut clips of the film *Le Ballon Rouge* (Albert Lamorisse, 1956). Participants overestimated time durations in the edited clips compared to the uncut clips (Kovarski et al. 2022).

The role of eye movements in the relationship between editing and time perception has also been studied (Balzarotti et al. 2021, using the same stimuli as Eugeni et al. 2016). Once more, participants overestimated the duration of videos with a high editing density but perceived that time flew faster while watching them. Importantly, it was found that eye movements mediated the effect of editing density on time perception. Fast-paced videos were associated with shorter fixation durations, which were in turn associated with longer DJs. There was no mediating effect of saccade amplitudes. A tentative explanation was that the interoceptive signals from increased eye movements in the high-density videos led to overestimations of time duration, although attention-based models of time perception could also explain the findings.

***Eye movements mediated the effect of editing density on time perception.***

#### Arousal, Eye Movements, and Visual Stimuli

Another factor that may determine eye movements when watching audiovisual stimuli is music. This could in turn influence time perception by either of these mechanisms—attention or interoceptive cues. Higher arousal induced by auditory cues increases attention in visual search (Asutay and Västfjäll 2017). In the context of films, music varying in arousal can activate schemas, which in turn determine where attention is allocated in a shot, thus influencing eye movements. An anxiety-inducing soundtrack can lead individuals to spend more time looking at minor details in a film (Ansani et al. 2020).

Yet the relationship between arousing music and eye movements is not necessarily straightforward. Ann-Kristin Wallengren and Alexander Strukelj (2015) conducted an exploratory study investigating how action music and soft music affected eye movements when viewing film clips. Action music led to shorter fixation durations (quicker scanning) in one film but longer fixation durations in two others. A possible explanation is that the choice of music in the latter two films created greater ambiguity, prompting closer inspection of the scenes. It follows that the interaction of auditory and visual information is important in the experience of audiovisual stimuli.

Thus, editing density, music-induced arousal, and eye movements have all been linked with the perception of time. However, there is little research into how varying both the auditory and visual features of the same stimulus can impact time perception. Additionally, multimodal studies on another measure of time perception (rate) have found conflicting results about the relative dominance of auditory and visual modalities in temporal processing (Bausenhart et al. 2014; Boltz 2017; Wang et al. 2021). This study aims to explore how both editing density and music-induced arousal influence time perception and emotional involvement when viewing audiovisual stimuli, as well as the role of eye movements.

**Objectives**

- To replicate findings on the effects of arousal and editing density on DJs, PoTJs, and emotional involvement, mediated through eye movements.
- To examine the effects of music-induced arousal on DJs, PoTJs, and emotional involvement in audiovisual stimuli.
- To examine whether there is an interaction between music-induced arousal and editing density on DJs, PoTJs, and emotional involvement.

**Hypotheses**

H1. Participants will perceive the duration of videos with high-arousal music as significantly longer than the duration of videos with low-arousal music.

H2. Participants will experience time as passing faster when watching videos with high-arousal music compared to videos with low-arousal music.

H3. There will be no effect of music-induced arousal on emotional involvement.

H4. Pupil size in high-arousal conditions will be significantly larger than pupil size in low-arousal conditions of the same editing density.

H5. Participants will perceive the duration of videos with a high editing density as significantly longer than the master shot.

H6. Participants will experience time as passing faster when watching videos with a high editing density compared to videos with a low editing density and the master shot.

H7. Participants will rate videos with a high editing density as significantly more emotionally engaging than videos with a low editing density and the master shot.

H8. There will be no significant interaction between the effect of music-induced arousal and editing density on DJs, PoTJs, and emotional involvement.

H9. Fixation durations will mediate the effect of editing density on DJs, but not PoTJs, or emotional involvement.

H10. Saccade amplitudes will not mediate the effect of editing density and arousal on DJs, PoTJs, and emotional involvement.

## Method

### *Sample*

Forty-two participants between 18 and 51 years of age took part in the experiment. They received an Amazon gift coupon worth two hundred rupees for participation. Three participants were excluded due to reported inattention to the videos (*n*=1), too many blinks in the recording (*n*=1), and difficulty seeing the screen (*n*=1). The final sample consisted of 39 participants aged between 18 and 51 years (M =22.4, *SD*=6.27). Thirty-two participants (82.1 percent) were female and 7 (17.9 percent) were male. Due to practical constraints, it was not possible to balance gender in the sample. Fifteen (38.5 percent) had completed a high school education, 22 (56.4) had an undergraduate degree, 1 (2.6 percent) had a master's degree, and 1 (2.6 percent) had completed a doctoral degree. According to the General Health Questionnaire, 24 participants (61.5 percent) were not psychologically distressed, and 15 participants (38.5 percent) were.

Participants were recruited through convenience sampling. They completed an online consent form and pre-experiment questionnaire. Individuals with psychiatric diagnoses were not recruited as psychiatric disorders can influence time perception (e.g., Cáceda et al. 2020; Ptacek et al. 2019; Thoenes and Oberfeld 2017). Those reporting eye infections or allergic reactions leading to watery or irritated eyes were also excluded. The present study was open to individuals between 18 and 55 years of age. This was to increase the diversity of the sample as similar studies in the past have focused only on undergraduate samples. An upper age limit of 55 was selected because differences in saccadic eye movements (Dowiasch et al. 2015) as well as physiological arousal in response to stimuli have been found in older adults (Grühn and Scheibe 2008; Kunzmann et al. 2005; Streubel and Kunzman 2011). It is also likely that time perception changes with aging due to changes in the cortico-thalamic-basal ganglia circuits (Turgeon et al. 2016). From those participants who filled the pre-experiment questionnaire, two participants were screened out due to psychiatric diagnoses, one was screened out due to uncorrected visual impairment, and one due to watery eyes. After screening, forty-two participants were invited to the laboratory for the experimental task.

### *Design*

The study used a 3×2 repeated measures experimental design. The two independent variables were editing density (master shot, low editing density, high editing density), and music-induced arousal (high arousal and low arousal). Mean fixation duration and mean saccade amplitude were mediating variables, and the dependent variables were DJs, PoTJs, and emotional involvement.

### *Tools*

*General Health Questionnaire-12 (GHQ-12)*

The GHQ-12 (Goldberg and Williams 1988) is a 12-item version of the General Health Questionnaire that was developed by David Goldberg (1972). It is a screening tool used to identify individuals who have, or are at risk of developing, psychiatric illnesses (Jackson 2007). It consists of Likert-style questions rated from 0–3. In the present study, a bimodal method of scoring was used. Individuals scoring greater than or equal to 2 were considered "psychologically distressed," as per standard practice in India (Shivakumar et al. 2015).

*Post-video Questionnaire*

After watching each video, participants answered a post-video questionnaire with questions from a previous study on editing and time perception (Eugeni et al. 2016). The questionnaire was translated from Italian, and the accuracy of the translation was confirmed by a psychologist who is fluent in Italian. The questionnaire yields scores for emotional involvement (composite of three 7-point Likert scores, with the total score ranging from 3 to 21), PoTJs (composite of two 9-point Likert scores, with the total score ranging from 2 to 18), and DJs (ranging from 0 to 30 seconds). Participants were also asked to recall the contents of the video. The purpose of this question was to distract the participants from the focus on time perception. The order of this question was changed from the original study. Instead of answering the question immediately after watching the video, participants answered it after answering all other questions to prevent their descriptions of the video from influencing subsequent ratings.

### *Eye-tracking Equipment and Measures*

Eye-tracking data was collected using an EyeLink 1000 Plus eye tracker (1000 Hz) with a desktop mount and a chinrest placed 59 centimeters away from the camera and 75 centimeters away from the screen. Monocular data was recorded from the participant's dominant eye. Saccade amplitudes, fixation durations, and pupil size data were recorded during each video clip. The thresholds used for distinguishing saccades and fixations were a velocity threshold of 30°/s and an acceleration threshold of $8000°/s^2$. Samples below the threshold were classified as fixations while those above the threshold were classified as saccades. Stimuli and post-video questionnaires were presented through the Experiment Builder Software (SR Research Ltd. 2020).

### *Stimuli*

The stimuli consisted of one practice video 12 seconds in duration, six target videos of 13.5s each, and nine fillers between 6s and 25s long. The target clips varied on two parameters—music (high arousal and low arousal) and editing density (master shot, low editing density (four cuts), and high editing density (twelve cuts)). The videos were previously used in studies on editing density and time perception (Balzarotti et al. 2021; Eugeni et al. 2016). Each clip depicts a man pouring water into a glass and then drinking the water.

The high-arousal background music was an excerpt from "Unter Donner and Blitz" by Johann Strauss. The low-arousal background music clip was an excerpt from "Intermezzo" from the Carmen Suite by Georges Bizet. Both clips have a positive valence and were identified as being high and low in arousal respectively by Gillian Sandstrom and Frank Russo (2010). The loudness of both audio tracks was matched to -23 LUFS using Adobe Audition 2021.

When piloting the experiment using only the target videos, participants thought all the videos had the same duration. Filler videos of varying durations were hence included in the task to reduce the sense that videos of the same duration were being presented multiple times. The filler videos included a woman blowing a balloon, two women gardening, a cake being decorated, and swirls of paint. Each type of video had either two or three versions varying in duration and editing density.

### *Procedure*

The study was approved by the Institutional Review Board of CHRIST (Deemed to be University). Before coming to the laboratory, the participants completed an online form where they provided consent and were screened. They were told that the purpose of the experiment was to understand the factors that make a video entertaining.

Participants were told that they would watch a series of videos and answer six questions at the end of each video. They were asked to pay attention as they would have to describe the contents of the video later. They were also instructed not to count during the video.

Participants were first shown a practice video of a man moving his hands in a circular motion on a table to control for a novelty effect (as in Balzarotti et al. 2021). The fifteen remaining videos were presented in a pseudorandom order (such that two target videos were not presented consecutively). The background music was delivered through Audio-Technica ATH-M20x headphones. Before each video, a scrambled isoluminant image of the first frame of the video was presented for 3000ms to establish a baseline pupil size for the trial. After the experiment, participants were debriefed, and

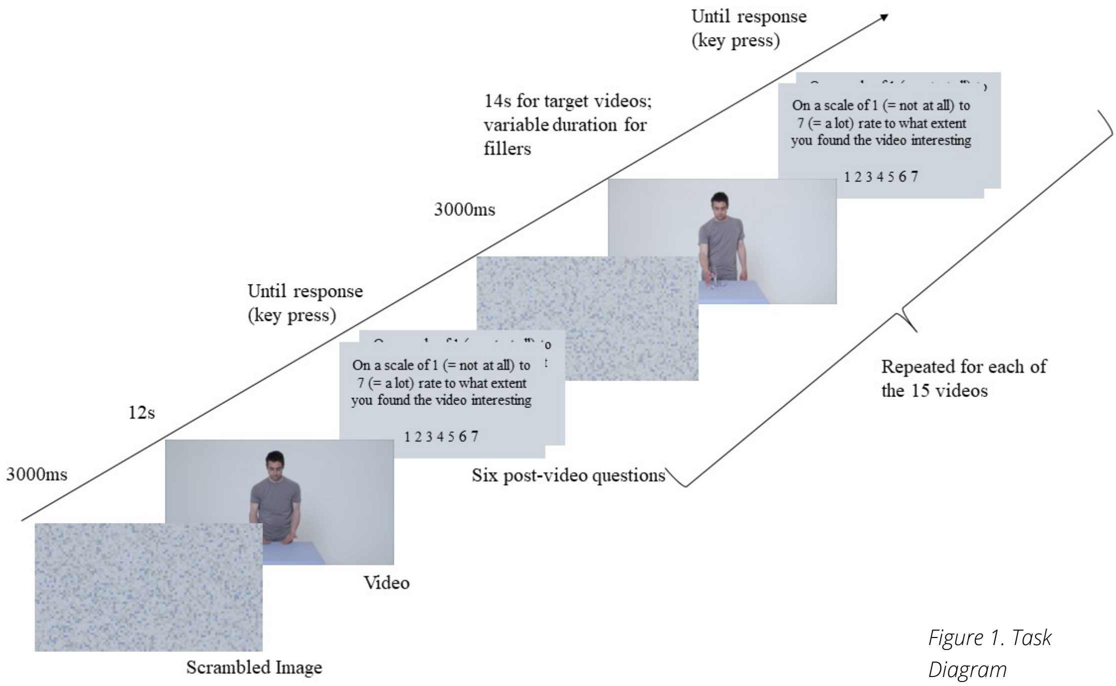


*Figure 1. Task Diagram*

the true purpose of the experiment was disclosed. Figure one depicts the experimental task.

***Data Analysis***

Data from thirty-nine participants was used for behavioral analysis. To test hypotheses one through eight, multilevel models were created using Generalized Linear Mixed Models (GLMMs) in IBM SPSS version 21. The data structure was specified with participant identifiers as the subject, and arousal and editing density as repeated measures. Hypotheses nine and ten were tested by constructing a multilevel mediation model using the MLMED macro for SPSS.

To test whether higher arousal background music was associated with larger pupil sizes (hypothesis four), a GLMM was run with pupil size as the dependent variable, arousal as a fixed effect and participant and video type as random effects with random slopes. Video type was added as a random effect to account for differences in luminance across the videos. Baseline pupil size measured 3000ms before each trial was used as a covariate. Simple contrasts were computed using estimated marginal means.

Hypotheses one and five concerning DJs were tested by constructing a GLMM with editing density, arousal, and GHQ score as fixed effects, participants as random effects with random slopes, and DJs as the predicted variable. As per standard practice, for DJs, an accuracy score was created by dividing the participant's response by the actual duration of the video (Balzarotti et al. 2021).

Hypotheses two and six were tested by constructing a similar GLMM, but with PoTJs as the predicted variable. Similarly, hypotheses three and seven were tested by constructing a GLMM with emotional involvement as the target variable. All models also had an interaction term (arousal * editing density) to test hypothesis eight. Since the frequency distributions for pupil sizes, DJs, and emotional involvement were positively skewed, a gamma distribution with a log link function was used for the models predicting these variables. A normal distribution with an identity link function was used for the model predicting PoTJs.

For the mediation analyses, three participants were excluded due to inaccurate recordings, resulting in a sample of thirty-six. Multilevel mediation has been identified as an appropriate strategy for the analysis of within-subjects experiments in cognitive psychology that involve mediating variables (Vuorre and Bolger 2018). The MLMED macro for multilevel mediation was used in IBM SPSS 21. Following the procedure of Balzarotti et al. (2021), editing density was coded as a binary variable for this analysis, with 0 representing unedited videos (the master shot) and 1 representing edited videos (low and high editing density). Editing density was added as the independent variable, audio as a level-1 covariate, and fixation durations and saccade amplitude were added as mediators. Participant identifiers were entered as random effects. Random slopes were not included as doing so resulted in non-convergence of the model. As the residuals for emotional involvement and duration judgment accuracy scores were not normally distributed, a log transformation was performed on these two variables for the mediation analysis, although the residuals for DJs were still not normally distributed after the transformation.

## Results

The reliability of the General Health Questionnaire -12 (GHQ-12) in this sample was acceptable (Cronbach's $\alpha$=.79). Table 1 depicts the means and standard deviations of the study variables in each condition.

### *Pupil Size*

The main effect of arousal on pupil size was significant ($F$ (2, 213) = 123.6, $p$<.001). Pupil size was larger in the high-arousal music conditions compared to the low-arousal conditions (Mean difference = 53.6).

### *Behavioral Data*

Table 2 depicts the fixed effects results of the multilevel models used to analyze the behavioral data. Table 3 depicts the pairwise comparisons of

*Table 1. Descriptive statistics of study variables in each condition*

| Video type | Master shot | | Low editing density | | High editing density | |
|---|---|---|---|---|---|---|
| Arousal level | Low | High | Low | High | Low | High |
| Pupil size (mm) | 3.78 (0.61) | 3.85 (0.62) | 3.76 (0.57) | 3.83 (0.60) | 3.70 (0.52) | 3.82 (0.56) |
| Average Fixation Duration (ms) | 503.1 (108.7) | 538.8 (198.3) | 478.1 (199.6) | 534.9 (268.2) | 366.7 (90.3) | 362.5 (61.5) |
| Average Saccade Amplitude | 3.01 (0.61) | 3.18 (0.76) | 2.49 (0.56) | 2.64 (0.68) | 3.35 (0.54) | 3.27 (0.51) |
| Emotional Involvement | 2.4 (1.3) | 3.0 (1.5) | 2.5 (1.2) | 3.0 (1.7) | 2.5 (1.3) | 3.2 (1.6) |
| Passage of Time Judgment | 4.2 (1.8) | 5.4 (1.6) | 4.3 (2.0) | 5.3 (1.9) | 4.9 (1.8) | 5.8 (1.8) |
| Duration Judgment | 0.64 (0.31) | 0.61 (0.24) | 0.64 (0.30) | 0.60 (0.26) | 0.63 (0.29) | 0.65 (0.28) |

*Note.* Mean pupil sizes, fixation durations, and saccade amplitudes were reported for ***n***= 36 participants. Mean emotional involvement, passage of time judgment, and duration judgment are reported for ***n***=39 participants. The duration judgment is an accuracy index where values below 1 indicate underestimation and values above 1 indicate overestimation.

*Table 2. Fixed effects of editing density, arousal, and GHQ on passage of time judgments, emotional involvement, and duration judgments*

| Dependent Variable | Predictor | *F* | *df1* | *df2* | *p* |
|---|---|---|---|---|---|
| Passage of Time Judgment | Corrected model | 8.32 | 6 | 66 | <.001 |
| | Editing Density | 5.13 | 2 | 103 | .008 |
| | Arousal | 32.17 | 1 | 160 | <.001 |
| | GHQ | 0.76 | 1 | 34 | .391 |
| | Editing ✻ Arousal | 0.09 | 2 | 103 | .912 |
| Emotional Involvement | Corrected model | 2.11 | 6 | 66 | .063 |
| | Editing Density | 1.00 | 2 | 106 | .372 |
| | Arousal | 9.23 | 1 | 167 | .003 |
| | GHQ | 0.01 | 1 | 33 | .923 |
| | Editing ✻ Arousal | 0.04 | 2 | 106 | .964 |
| Duration Judgment | Corrected model | 0.62 | 6 | 61 | .710 |
| | Editing Density | 0.49 | 2 | 94 | .613 |
| | Arousal | 0.24 | 1 | 147 | .623 |
| | GHQ | 0.45 | 1 | 34 | .506 |
| | Editing ✻ Arousal | 1.20 | 2 | 94 | .306 |

**Note.* Satterthwaite method for degrees of freedom.

*Table 3. Bonferroni-adjusted pairwise comparisons for the effects of editing density and audio on passage of time judgments*

| Comparison | | Contrast estimate | *SE* | *t* | *df* | *p* |
|---|---|---|---|---|---|---|
| Video | | | | | | |
| Master shot | – Low editing density | –0.10 | 0.22 | –0.49 | 101 | .629 |
| Master shot | – High editing density | –0.67 | 0.21 | –3.13 | 105 | .007 |
| Low editing density | – High editing density | –0.56 | 0.24 | –2.32 | 103 | .045 |
| Audio | | | | | | |
| High arousal | – Low arousal | 1.04 | 0.18 | 5.67 | 160 | <.001 |

significant effects. There was a significant effect of both editing density and music-induced arousal on PoTJs. Estimated marginal means and pairwise contrasts showed that participants experienced time as passing significantly faster ($F$ (2, 103) = 5.13, $p$ =.008) for the videos with a high editing density ($M$ = 5.35) compared to both the low editing density ($M$ = 4.79) and master shot videos ($M$ = 4.68). However, there was no difference between the low editing density video and the master shot. Participants also experienced time passing faster ($F$ (1,160) = 32.17, $p$<.001) when listening to high-arousal ($M$ = 5.46) compared to low-arousal music ($M$ = 4.42).

The corrected fixed effects models predicting DJs ($F$ (6, 61) = 0.06, $p$ =.710) was not fit. Likewise, the fixed effects model predicting emotional involvement ($F$ (1, 167) = 9.22, $p$ = 0.06) was not fit, although differences in estimated means were in the expected direction. Participants' level of psychological distress measured by the General Health Questionnaire did not predict any of the dependent variables. Table 4 illustrates the fixed coefficients and variance components for the empty and tested multilevel models.

### *Mediation Analysis*

Mediation analysis was used to test hypotheses nine and ten. The results of the mediation analysis are depicted in Figure 2. Fixation duration mediated the effect of editing density on PoTJ. Edited videos were associated with smaller fixation durations, which were in turn associated with faster passage of time. Fixation durations did not mediate the relationship between editing density and DJs or emotional involvement. Saccade amplitudes did not mediate the relationships between editing and arousal and time perception.

## Discussion

The present study investigated how varying a video's pace of editing as well as its background music influences perceptions of time and emotional involvement, mediated by eye movements. Participants watched six target

*Table 4. Fixed coefficients and variance components for passage of time judgments, duration judgments, and emotional involvement*

| | Passage of Time Judgment | | Emotional Involvement | | Duration Judgment | |
|---|---|---|---|---|---|---|
| Measure/ Parameter | Empty Model Est (SE) | Tested Model Est (SE) | Empty Model Est (SE) | Tested Model Est (SE) | Empty Model Est (SE) | Tested Model Est (SE) |
| Fixed Effects | | | | | | |
| Intercept | 5.07** (0.22) | 5.62** (0.41) | 0.91** (0.07) | 1.07** (0.13) | –0.54** (0.07) | –0.55** (0.11) |
| Audio | | | | | | |
| Low arousal (A) | – | –0.93* (0.34) | – | –0.20 (0.04) | | –0.05 (0.05) |
| Video | | | | | | |
| Mastershot (1) | – | –0.61* (0.27) | – | –0.10 (0.10) | | -0.06 (0.05) |
| Low Editing Density (2) | – | –0.46 (0.30) | – | –0.04 (0.12) | | -0.09 (0.05) |
| GHQ – Not distressed | – | 0.40 (0.46) | – | –0.01 (0.14) | | 0.09 (0.05) |
| Video * Audio | | | | | | |
| Low arousal * Mastershot | - | –0.11 (0.43) | – | 0.01 (0.14) | | 0.08 (0.08) |
| Low arousal * Low editing density | - | –0.21 (0.49) | – | 0.04 (0.15) | | 0.10 (0.07) |
| Variance Components | | | | | | |
| Video clip 1A | – | 1.22** (0.37) | – | 0.14** (0.04) | | 0.09** (0.02) |
| Video clip 1B | – | 1.12** (0.34) | – | 0.17** (0.05) | | 0.03** (0.01) |
| Video clip 2A | – | 2.58** (0.69) | – | 0.14** (0.04) | | 0.05** (0.01) |
| Video clip 2B | – | 1.73** (0.50) | – | 0.26** (0.07) | | 0.04** (0.01) |
| Video clip 3A | – | 2.68** (0.71) | – | 0.24** (0.07) | | 0.02* (0.01) |
| Video clip 3B | – | 1.51** (0.44) | – | 0.15** (0.04) | | 0.07** (0.01) |
| Intercept | – | 0.68 (0.45) | – | 0.06 (0.04) | | 0.08* (0.04) |
| AIC | 875.72 | 838.22 | 339.07 | 334.73 | 67.86 | 85.40 |

*Note.* *$p<0.05$; **$p<0.01$. The estimates reported in this table were obtained using high arousal (Audio B), high editing density (Video 3), and GHQ scores above the cutoff (psychologically distressed) as reference conditions. 1 = Mastershot, 2 = Low Editing Density, 3= High Editing Density, A= Low Arousal, B= High Arousal, AIC = Akaike Corrected Information Criterion.

videos that had three levels of editing density and two levels of arousal. Music-induced arousal was associated with the subjective sense of the passage of time, but not emotional involvement. Editing density predicted fixation durations, which in turn had a weak relationship with PoTJs. Neither editing nor arousal predicted judgments of duration. There was no interaction between editing and arousal on time perception or emotional involvement.

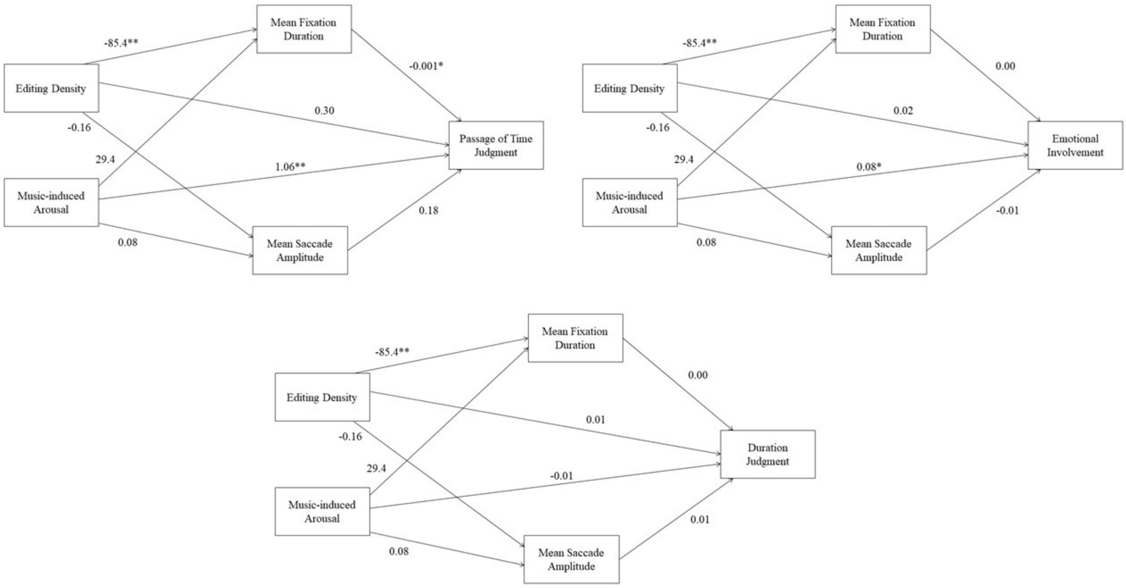


*Figure 2. Within person-level effects of editing density and music-induced arousal on PoTJs, duration judgments, and emotional involvement via fixation duration and saccade amplitude.*
*Note*. Unstandardized estimates are reported. *$p$<.05, **$p$<.01.

Arousal predicted pupil size across all video types when accounting for baseline pupil size, supporting hypothesis four. This provides physiological evidence that the arousal manipulation was successful. Arousal also predicted PoTJs, with participants experiencing time as passing faster when listening to high-arousal music compared to low-arousal music, in support of hypothesis two. This finding adds to evidence by Droit-Volet and Wearden (2016) who found a positive relationship between arousal and PoTJs in young adults.

Hypothesis three was supported, as the overall model predicting emotional involvement was not fit, although differences were in the expected direction (with slightly higher emotional involvement ratings for high-arousal music). This is inconsistent with previous work in performing arts (Latulipe et al. 2011) where self-reported audience engagement ratings were correlated with arousal as measured by galvanic skin response. Some have even used physiological arousal as a measure of engagement (Juvrud et al. 2022). Further research may be necessary on situations in which physiological arousal may not always be associated with self-reported interest, as was the case in the present study.

The fixed effects model for DJs was not fit. Hence, hypotheses one and five were rejected. This may be because of minimal intraindividual variability in the DJ responses. Many participants seemed to realize that all the target videos were of the same length, despite the use of fillers. The repetitiveness of the videos could explain this—participants watched six videos of the same action type in contrast to three videos used in previous studies (Balzarotti et al. 2021; Eugeni et al. 2016). It has been postulated that using videos of the same duration may bias participants' responses in timing tasks (Kovarski et al. 2022). Additionally, the presence of music might have made it easier to judge the duration of videos. Participants might have inferred from the start and end points of the background music that the videos were of the same duration, although participants did not mention any such strategy during the debriefing. Alternatively, the mere presence of music might facilitate more accurate discrimination abilities between time intervals regardless of an explicit cognitive strategy.

It is noteworthy that, in contrast to previous research where, on average, participants overestimated video durations (Balzarotti et al. 2021), in the present study, participants consistently underestimated video durations, similarly to Ansani and colleagues (2021) (mean duration judgment accuracy values were lower than 1 in all conditions). Since the previous study did not use background music, the present findings may reflect the finding that duration estimates are longer when watching video-only clips compared to clips that have both audio and video (Appelqvist-Dalton et al. 2022). The presence of the background music might have led to underestimations of the video duration. It should, however, be noted the earlier researchers (Balzarotti et al. 2021) used slightly shorter video clips of two other actions in addition to the videos about drinking water, which could have contributed to this difference.

Hypothesis six was supported. Participants rated time as passing significantly faster when watching high editing density videos compared to both the master shot and low editing density videos. This adds to previous research on editing style and subjective experience of time passage (Balzarotti et al. 2021; Eugeni et al. 2016).

In the present study, music-induced arousal had a stronger effect on PoTJs than editing density[1]. The greater influence of auditory compared to visual information might reflect an auditory driving effect. Marilyn Boltz (2017) found that, when presented with both auditory and visual information, auditory information exerted a greater influence on perceived rates of change in the stimulus. Hence, the findings of this study may reflect a heavier reliance on auditory information when perceiving time—music 'drives' the experience of time more than visual information like pace of editing.

***The greater influence of auditory compared to visual information might reflect an auditory driving effect.***

Editing density did not predict emotional involvement, in contrast with existing research (Eugeni et al. 2016). It could be that the presence of background music acted to minimize differences in emotional involvement that may have arisen in a video-only modality. There was also no interaction between editing density and arousal, consistent with hypothesis eight.

The results of the mediation analysis refuted hypothesis nine. While previous findings suggest that fixation durations mediate the relationship between editing density and DJs, but not PoTJs (Balzarotti et al. 2021), the opposite was found in this case. Once more, the lack of significant findings regarding DJs is likely due to the low variability in responses. Had the study used stimuli and audio clips of varying durations, more valid results might have been obtained. The weak relationship between fixation durations and PoTJs might reflect that there is indeed an effect of oculomotor cues on PoTJs which was too small to be detected in previous research.

Even though differing results were found regarding fixation durations, the results related to saccade amplitudes were consistent with previous research. There was no mediating effect of saccade amplitude on any of the dependent variables, in support of hypothesis ten. This suggests that the frequency of eye movements or the duration of fixations may be more important in timing than the length of each saccade. There was also no mediating effect of eye movements on emotional involvement, which is consistent with the expectation that eye movements are unrelated to emotional involvement.

The finding that arousal did not influence either of the eye-tracking variables could be due to the nature of the stimuli. The video clips depicted an unambiguous action performed by a single actor with a stark mise-en-scène. The action itself did not have any narrative meaning. Hence, arousal levels may not have facilitated any differing scanning patterns, to resolve ambiguity for example (Ansani et al. 2020; Wallengren and Strukelj 2015). This may also explain the absence of a significant interaction between editing and music-induced arousal.

Finally, while some research has indicated that clinical conditions such as depression (Thönes and Oberfeld 2015) influence time perception, psychological distress as measured by the General Health Questionnaire was not associated with time perception in the present study. It could be that time perception is only affected when psychological distress is very severe, a distinction that was not captured by the dichotomous classification. Furthermore, individuals with psychiatric disorders were excluded from taking part in the study.

### Limitations

The present study has certain limitations. Although participants were not told the purpose of the experiment, they likely understood that it was about time perception due to the repetitive nature of the questions. Thus, the study was not a true retrospective timing task (Block et al. 2018). It has been suggested that retrospective timing tasks can be carried out on the same participant if activities are conducted between stimulus presentation and the time estimation (Block et al. 2018). In the present study, participants completed three questions related to emotional involvement before answering the questions about timing.

Due to practical constraints, the sample size was small compared to previous studies on editing density and time perception. As a result, it had less statistical power and might not have detected some small effects. Data was preprocessed to remove smooth pursuit eye movements that were misclassified as saccades. This was done using visual estimation based on a temporal graph of the eye coordinates. Since the identification of true fixations was a manual process, it is possible that some human error affected the mean fixation duration values.

## Conclusion

The present study found that music-induced arousal was associated with subjective PoTJs when viewing audiovisual media. There was a positive relationship between editing density and PoTJs, mediated by fixation durations. This adds to the research implicating arousal and editing style as important variables in temporal processing.

There is now a growing body of literature about PoTJs as a measure of time perception that is independent of DJs. Future studies could focus on incorporating this facet into existing models of temporal processing and examining the neural mechanisms involved. Research on flow states, which involve optimal levels of arousal (Peifer et al. 2014) and a feeling that time is flying (van der Linden et al. 2020) could provide insights into this phenomenon. The results of the present study could be useful in the creation of educational videos, where it might be of interest to manipulate subjective feelings of the passage of time (Ansani et al. 2021) and facilitate flow (Ibáñez et al. 2014) to improve learning outcomes.

Future studies on time perception in audiovisual media could also use clips and soundtracks of varying durations such that participants cannot infer the duration of a video based on salient aspects of the music. The present findings may be explained through the role of attention or interoceptive signals from eye movements. Hence, further research can be conducted to explore the most appropriate explanation.

## Acknowledgments

We would like to acknowledge the guidance and support of Mr. Adhiraj Chowdhury and Dr. Madhavi Rangaswamy, NeuroCognition Lab, CHRIST (Deemed to be University) for instrument use training and experiment building for eye-tracking experiments in this study. We would also like to acknowledge Mr. Sumer Vaid and Dr. Tony Sam George for guidance with statistical analyses.

**Kathryn Nicole Sam** has a master's degree in neuropsychology from CHRIST (Deemed to be University), Bengaluru. She is interested in the experimental study of narratives. Specifically, she is interested in the cognitive and neural processes associated with viewing films.

**Dr K. Jayasankara Reddy** is an Associate Dean of Social Sciences, professor of psychology, and coordinator for the Centre for Research at CHRIST (Deemed to be University), Bengaluru. He is also in charge of the Cognitive Lab. He is an active teacher, researcher, and practitioner in the fields of health, cognitive neuropsychology, and neuroscience.

## Notes

[1] There is no agreed-upon way of computing standard effect sizes in multilevel models (e.g., Rights and Sterba 2019). Unstandardized effect sizes (mean differences) are hence reported in this study, consistent with recommendations on reporting effect sizes (Pek and Flora 2018). In order to compare standard effect sizes of editing and arousal, a two-way repeated measures ANOVA with editing, arousal, and editing*arousal as factors was performed. Arousal ($\eta_p^2$ =.369) had a greater within-subjects effect on PoTJs compared to editing ($\eta_p^2$ =.117).

## References

Ansani, Alessandro, Marco Marini, Francesca D'Errico, and Isabella Poggi. 2020. "How Soundtracks Shape What We See: Analyzing the Influence of Music on Visual Scenes Through Self-Assessment, Eye Tracking, and Pupillometry." *Frontiers in Psychology* 11 (October). https://doi.org/10.3389/fpsyg.2020.02242.

Ansani, Alessandro, Marco Marini, Luca Mallia, and Isabella Poggi. 2021. "Music and Time Perception in Audiovisuals: Arousing Soundtracks Lead to Time Overestimation No Matter Their Emotional Valence." *Multimodal Technologies and Interaction* 5 (11): 68. https://doi.org/10.3390/mti5110068.

Appelqvist-Dalton, Mattis, James P. Wilmott, Mingjian He, and Andrea Megela Simmons. 2022. "Time Perception in Film Is Modulated by Sensory Modality and Arousal." *Attention, Perception, & Psychophysics* 84 (March): 926–942. https://doi.org/10.3758/s13414-022-02464-9.

Asutay, Erkin, and Daniel Västfjäll. 2017. "Exposure to Arousal-Inducing Sounds Facilitates Visual Search." *Scientific Reports* 7 (1): 10363. https://doi.org/10.1038/s41598-017-09975-8.

Balzarotti, Stefania, Federica Cavaletti, Adriano D'Aloia, Barbara Colombo, Elisa Cardani, Maria Rita Ciceri, Alessandro Antonietti, and Ruggero Eugeni. 2021. "The Editing Density of Moving Images Influences Viewers' Time Perception: The Mediating Role of Eye Movements." *Cognitive Science* 45 (4): e12969. https://doi.org/10.1111/cogs.12969.

Bausenhart, Karin M., Maria Dolores de la Rosa, and Rolf Ulrich. 2014. "Multimodal Integration of Time: Visual and Auditory Contributions to Perceived Duration and Sensitivity." *Experimental Psychology* 61 (4): 310–322. https://doi.org/10.1027/1618-3169/a000249.

Block, Richard A., Simon Grondin, and Dan Zakay. 2018. "Prospective and Retrospective Timing Processes: Theories, Methods, and Findings." In *Timing and Time Perception: Procedures, Methods & Applications*, ed. Argiro Vatakis, Fuat Balcı, Massimiliano Di Luca, and Ángel Correa, 32–51. Leiden, The Netherlands: Brill.

Boltz, Marilyn G. 2017. "Auditory Driving in Cinematic Art." *Music Perception* 35 (1): 77–93. https://doi.org/10.1525/mp.2017.35.1.77.

Brighouse, Carolyn, Jess Hartcher-O'Brien, and Carmel A. Levitan. 2014. "Encoding of Duration and Rate by an Integrative Model of Temporal Processing." *Timing & Time Perception Reviews* 1 (3): 1–11. https://doi.org/10.1163/24054496-00101003.

Cáceda, Ricardo, Jessica M. Carbajal, Ronald M. Salomon, Jordan E. Moore, Greg Perlman, Prasad R. Padala, Abdullah Hasan, and Pedro L. Delgado. 2020. "Slower Perception of Time in Depressed and Suicidal Patients." *European Neuropsychopharmacology* 40 (November): 4–16. https://doi.org/10.1016/j.euroneuro.2020.09.004.

Cocenas-Silva, Raquel, José Lino Oliveira Bueno, Paul Molin, and Emmanuel Bigand. 2011. "Multidimensional Scaling of Musical Time Estimations." *Perceptual and Motor Skills* 112 (3): 737–748. https://doi.org/10.2466/11.24.PMS.112.3.737-748.

Dowiasch, Stefan, Svenja Marx, Wolfgang Einhäuser, and Frank Bremmer. 2015. "Effects of Aging on Eye Movements in the Real World." *Frontiers in Human Neuroscience* 9 (February). https://doi.org/10.3389/fnhum.2015.00046.

Droit-Volet, Sylvie, and John Wearden. 2016. "Passage of Time Judgments Are Not Duration Judgments: Evidence from a Study Using Experience Sampling Methodology." *Frontiers in Psychology* 7 (February). https://doi.org/10.3389/fpsyg.2016.00176.

Droit-Volet, Sylvie, Danilo Ramos, José L. O. Bueno, and Emmanuel Bigand. 2013. "Music, Emotion, and Time Perception: The Influence of Subjective Emotional Valence and Arousal?" *Frontiers in Psychology* 4 (July). https://doi.org/10.3389/fpsyg.2013.00417.

Eugeni, Ruggero, Stefania Balzarotti, Frederica Cavaletti, and Adriano D'Aloia. 2016. "It Doesn't Seem It, But It Is: A Neurofilmological Approach to the Subjective Experience of Moving-Image Time." In *The Extended Theory of Cognitive Creativity: Interdisciplinary Approaches to Performativity*, ed. Antonino Pennisi and Alessandra Falzone, 243–265. Cham, Switzerland: Springer Nature.

Goldberg, David P. 1972. *The Detection of Psychiatric Illness by Questionnaire: A Technique for the Identification and Assessment of Non-psychotic Psychiatric Illness*. Oxford: Oxford University Press.

Goldberg, David P., and Paul Williams. 1988. *A User's Guide to the General Health Questionnaire*. Berkshire: NFER-Nelson.

Grall, Clare, and Emily S. Finn. 2022. "Leveraging the Power of Media to Drive Cognition: A Media-Informed Approach to Naturalistic Neuroscience." *Social Cognitive and Affective Neuroscience* 17 (6): 598–608. https://doi.org/10.1093/scan/nsac019.

Grühn, Daniel, and Susanne Scheibe. 2008. "Age-Related Differences in Valence and Arousal Ratings of Pictures from the International Affective Picture System (IAPS): Do Ratings Become More Extreme with Age?" *Behavior Research Methods* 40 (2): 512–521. https://doi.org/10.3758/BRM.40.2.512.

Ibáñez, María Blanca, Ángela Di Serio, Diego Villarán, and Carlos Delgado Kloos. 2014. "Experimenting with Electromagnetism Using Augmented Reality: Impact on Flow Student Experience and Educational Effectiveness." *Computers & Education* 71 (February): 1–13. https://doi.org/10.1016/j.compedu.2013.09.004.

Jackson, C. 2007. "The General Health Questionnaire." *Occupational Medicine* 57 (1): 79. https://doi.org/10.1093/occmed/kql169.

Juvrud, Joshua, Gabriel Ansgariusson, Patrik Selleby, and Magnus Johansson. 2022. “Game or Watch: The Effect of Interactivity on Arousal and Engagement in Video Game Media.” *IEEE Transactions on Games* 14 (2): 308–317. https://doi.org/10.1109/TG.2021.3073084.

Kovarski, Klara, Joanna Dos Reis, Claire Chevais, Anaïs Hamel, Dominique Makowski, and Marco Sperduti. 2022. “Movie Editing Influences Spectators’ Time Perception.” *Scientific Reports* 12 (November). https://doi.org/10.1038/s41598-022-23992-2.

Kunzmann, Ute, Cenita S. Kupperbusch, and Robert W. Levenson. 2005. “Behavioral Inhibition and Amplification During Emotional Arousal: A Comparison of Two Age Groups.” *Psychology and Aging* 20 (1): 144–158. https://doi.org/10.1037/0882-7974.20.1.144.

Lake, Jessica I., Kevin S. LaBar, and Warren H. Meck. 2016. “Emotional Modulation of Interval Timing and Time Perception.” *Neuroscience & Biobehavioral Reviews* 64 (May): 403–420. https://doi.org/10.1016/j.neubiorev.2016.03.003.

Latulipe, Celine, Erin A Carroll, and Danielle Lottridge. 2011. “Love, Hate, Arousal and Engagement: Exploring Audience Responses to Performing Arts.” In *CHI ‘11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1845–1854. New York: Association for Computing Machinery. https://doi.org/10.1145/1978942.1979210.

Nather, Francisco C., José L. O. Bueno, Emmanuel Bigand, and Sylvie Droit-Volet. 2011. “Time Changes with the Embodiment of Another’s Body Posture.” *PLoS ONE* 6 (5): e19818. https://doi.org/10.1371/journal.pone.0019818.

Peifer, Corinna, André Schulz, Hartmut Schächinger, Nicola Baumann, and Conny H. Antoni. 2014. “The Relation of Flow-Experience and Physiological Arousal under Stress — Can u Shape It?” *Journal of Experimental Social Psychology* 53 (July): 62–69. https://doi.org/10.1016/j.jesp.2014.01.009.

Pek, Jolynn, and David B. Flora. 2018. “Reporting Effect Sizes in Original Psychological Research: A Discussion and Tutorial.” *Psychological Methods* 23 (2): 208–225. https://doi.org/10.1037/met0000126.

Ptacek, Radek, Simon Weissenberger, Ellen Braaten, Martina Klicperova-Baker, Michal Goetz, Jiri Raboch, Martina Vnukova, and George B. Stefano. 2019. “Clinical Implications of the Perception of Time in Attention Deficit Hyperactivity Disorder (ADHD): A Review.” *Medical Science Monitor* 25 (May): 3918–3924. https://doi.org/10.12659/MSM.914225.

Rights, Jason D., and Sonya K. Sterba. 2019. “Quantifying Explained Variance in Multilevel Models: An Integrative Framework for Defining R-Squared Measures.” *Psychological Methods* 24 (3): 309–338. https://doi.org/10.1037/met0000184.

Sandstrom, Gillian M., and Frank A. Russo. 2010. “Music Hath Charms: The Effects of Valence and Arousal on Recovery Following an Acute Stressor.” *Music and Medicine* 2 (3): 137–143. https://doi.org/10.1177/1943862110371486.

Shivakumar, Prafulla, Shilpa Sadanand, Srikala Bharath, N Girish, and Mathew Varghese. 2015. “Identifying Psychological Distress in Elderly Seeking Health Care.” *Indian Journal of Public Health* 59 (1): 18. https://doi.org/10.4103/0019-557X.152849.

SR Research. *Experiment Builder*. V. 2.2.299. SR Research. 2020.

Streubel, Berit, and Ute Kunzmann. 2011. “Age Differences in Emotional Reactions: Arousal and Age-Relevance Count.” *Psychology and Aging* 26 (4): 966–978. https://doi.org/10.1037/a0023424.

Thönes, Sven, and Daniel Oberfeld. 2015. “Time Perception in Depression: A Meta-Analysis.” *Journal of Affective Disorders* 175 (1): 359–372. https://doi.org/10.1016/j.jad.2014.12.057.

Thoenes, Sven, and Daniel Oberfeld. 2017. "Meta-Analysis of Time Perception and Temporal Processing in Schizophrenia: Differential Effects on Precision and Accuracy." *Clinical Psychology Review* 54 (June): 44–64. https://doi.org/10.1016/j.cpr.2017.03.007.

Turgeon, Martine, Cindy Lustig, and Warren H. Meck. 2016. "Cognitive Aging and Time Perception: Roles of Bayesian Optimization and Degeneracy." *Frontiers in Aging Neuroscience* 8 (May). https://doi.org/10.3389/fnagi.2016.00102.

van der Linden, Dimitri, Mattie Tops, and Arnold B. Bakker. 2021. "Go with the Flow: A Neuroscientific View on Being Fully Engaged." *European Journal of Neuroscience* 53 (4): 947–63. https://doi.org/10.1111/ejn.15014.

Vuorre, Matti, and Niall Bolger. 2018. "Within-Subject Mediation Analysis for Experimental Data in Cognitive Psychology and Neuroscience." *Behavior Research Methods* 50 (October): 2125–2143. https://doi.org/10.3758/s13428-017-0980-9.

Wallengren, Ann-Kristin, and Alexander Strukelj. 2015. "Film Music and Visual Attention: A Pilot Experiment Using Eye-Tracking." *Music and the Moving Image* 8 (2): 69–80. https://doi.org/10.5406/musimoviimag.8.2.0069.

Wang, Xinyue, Clemens Wöllner, and Zhuanghua Shi. 2021. "Perceiving Tempo in Incongruent Audiovisual Presentations of Human Motion: Evidence for a Visual Driving Effect." *Timing & Time Perception* 10 (1): 75–95. https://doi.org/10.1163/22134468-bja10036.

Wittmann, Marc. 2014 "Embodied Time: The Experience of Time, the Body, and the Self." In *Subjective Time: The Philosophy, Psychology, and Neuroscience of Temporality*, ed. Valtteri Arstila and Dan Lloyd, 507–523. Cambridge: MIT Press.