Attention, Perception, & Psychophysics

Time perception in film is modulated by sensory modality and arousal

Journal:	*Attention, Perception, & Psychophysics*
Manuscript ID	PP-ORIG-21-062
Manuscript Type:	Original Manuscript
Date Submitted by the Author:	14-Mar-2021
Complete List of Authors:	Appelqvist Dalton, Mattis; Brown University, Department of Cognitive, Linguistic, and Psychological Sciences Wilmott, James; Brown University, Department of Cognitive, Linguistic, and Psychological Sciences He, Mingjian; Massachusetts Institute of Technology, Harvard-MIT Health Sciences and Technology Simmons, Andrea; Brown University, Department of Cognitive, Linguistic, and Psychological Sciences; Brown University, Carney Institute for Brain Science
Keywords:	attention, Multisensory Processing, music cognition, sound recognition

March 14, 2021

To the Editor:

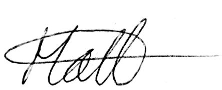
Please consider our manuscript, "Time perception in film is modulated by sensory modality and arousal," for publication in *Attention, Perception, & Psychophysics*. In this paper, we explore how the presence of music affects perception of time in film. Our work is an advance over previous explorations of time perception by using naturalistic film extracts as stimuli, by incorporating behavioral ratings in conjunction with pupillometry, and by separating out the role of sensory modality (visual and auditory) and emotional arousal on time perception.

This work was conducted at Brown University. Stimuli and data have been deposited in the Brown Data Repository https://doi.org/10.26300/ke1d-f930

Thank you for your consideration.

Sincerely,

Mattis Appelqvist Dalton
Mattis_Dalton@alumni.brown.edu



Declarations

Funding: This study was supported by the Independent Concentration Fund (Brown University).

Conflicts of interest/Competing interests: All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Ethics approval: This study was approved by the Institutional Review Board at Brown University.

Consent to participate: Informed consent was obtained from all individual participants included in the study.

Consent for publication: Data were anonymized and participants provided informed consent for publication.

Availability of data and materials: The datasets generated during the current study and stimuli are available in the Brown Digital Repository, https://doi.org/10.26300/ke1d-f930

Code availability: The code generated during the current study is available from the corresponding author on reasonable request.

Statement of significance:

Time perception is central to human psychology and interactions with the environment. Little is known about how we perceive the passing of time in film, a naturalistic multimodal stimulus. Using behavioral ratings and pupillometry, we explore how sensory modalities and arousal affect our perception of time in film extracts. We found that sensory modality (visual, auditory, and auditory-visual) and emotional arousal independently modulate distortions (under-estimates) of time and that the modulating relationship between arousal and time estimation differs between sensory modality. We interpret our findings according to existing biologically-inspired models of time perception. These findings are also relevant to the fields of multimodal perception, music cognition, attention, and arousal.

Time perception in film is modulated by sensory modality and arousal

Mattis Appelqvist-Dalton[a], James P. Wilmott[a],

Mingjian He[c], and Andrea M. Simmons[a,b]

[a]Department of Cognitive, Linguistic, and Psychological Sciences

[b]Carney Institute for Brain Science

Brown University, Providence, RI 02912

[c]Harvard-MIT Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, 02139

Address correspondence to:

Mattis Appelqvist-Dalton

14, Oyez

15130, St Simon

France

Stimuli and data are available in the Brown Digital Repository:

https://doi.org/10.26300/ke1d-f930

Abstract

Time perception is often distorted in cinematic experiences, but little research has used film - a naturalistic multimodal stimulus - to study how we perceive the passing of time. Here, we explore the effect of sensory modality and arousal on our perception of time in film. Using psychometric and psychophysiological methods, we investigate emotional responses to and time estimates of film extracts in three conditions: audiovisual (with music), visual (without music), and auditory (music on its own). Participants were presented with 27 film clips (9 in each condition) from nature documentaries, fiction, animation, and experimental films, and were asked to judge clip duration and to report subjective arousal and valence. Results reveal duration estimates varied with sensory modality. Clip durations were judged to be shorter than actual durations in all conditions, with visual-only clips judged to be longer than auditory-only and audiovisual clips. Consistent with previous research, arousal positively impacted time perception. The effect of arousal on duration perception was strongest when the clips were presented in the audiovisual modality. We discuss the results in relation to multimodal perception, attention, and the known effects of music in cinema.

Introduction

The manner in which we perceive time has fascinated physicists, philosophers, psychologists, and artists for centuries. We now realize that the subjective experience of time is not isomorphic to physical time, and indeed can be altered by internal and external factors. For example, judgments of the duration of events are modulated by age, gender, trait indices (anxiety and fearfulness), psychopathology, cognitive load, and mood (Block, Hancock, & Zakay, 2010; Droit-Vollet, Fayolle, & Gil, 2011; Lake, 2016). In laboratory experiments, several stimulus properties influence perceived time by producing "time distortions" in which subjective estimates are either shortened or lengthened. Non-salient, familiar, and high-probability content lead to time compression, where events are judged shorter compared to a baseline. On the other hand, increased stimulus complexity, magnitude, tempo, size, and intensity can lengthen judgments of duration relative to a subjective baseline (Tse, Intriligator, Rivest, & Cavanagh, 2004; Eagleman, 2008; van Wassenhove, Buonomano, Shimojo, & Shams, 2008; Droit-Volet, Ramos, Bueno, & Bigand, 2013; Cai & Wang, 2014; Matthews, 2015).

Time perception is also influenced by emotions, and arousal is key in this process (Merchant, Harrington, & Meck, 2013). The effect of emotions on time perception is often conceptualized by dimensional models such as the affective circumplex model (Russel, 1980). This model posits that emotions (affect) can be mapped on two central dimensions: valence (from positive to negative or from pleasant to unpleasant) and arousal (from low or calm to high or stimulated). The effect of arousal on time perception has been investigated using visual stimuli (Gil & Droit-Volet, 2012) as well as auditory stimuli including music (Droit-Volet et al., 2013) and emotional sounds such as laughter, sobs, and erotic sounds (Noulhiane, Mella, Samson, Ragot & Pouthas, 2007; Mella, Conty, & Pouthas, 2011). In all these instances, increased arousal lengthens estimates of subjective time. The effect of arousal on duration perception has been attributed to the activation of the locus coeruleus norepinephrine system, and many of its known effects are in accordance with the recorded phenomenology of temporal distortions such as altered attention, heightened perception, or “chronostasis” - the experience of time slowing down or coming to a halt (Arstila, 2012).

Several scholars have asserted that the perception of time is also influenced by sensory modality. Paul Fraisse, in *Psychologie du Temps,* posited that the duration of an interval depends on the *‘nature’* of its limits (Fraisse, 1957, pp. 136). For instance, in the case of *empty* intervals (a duration determined by two stimuli), if the limits are auditory the duration will seem shorter than if the limits are visual (Meumann, 1894, as cited in Fraisse, 1957). On the other hand, for *full* intervals (intervals of time filled with multiple events or continuous events), auditory durations are estimated as longer than visual durations by approximately 20% (Behar & Bevan, 1961). More recent research has confirmed that auditory signals are generally perceived as lasting longer than visual signals of the same objective duration (Droit-Volet, Meck, & Penney, 2007; Burr, Banks, & Morrone, 2009; Merchant et al., 2013; Ortega, Guzman-Martinez, Grabowecky, & Suzuki, 2014).

Much of our understanding of how sensory modality influences perception of time is based on experiments in which non-naturalistic stimuli are presented in an artificial context. Film is an ecologically-valid way of studying the influence of sensory modality on time perception.

Indeed, the manipulation of time is central to film-making: *"You will always recognise the editing of Bergman, Bresson, Kurusawa or Antonioni; none of them could ever be confused with anyone else, because of each one's perception of time [...]"* (Tarkovski, 1988, pp. 121). To date, there is very little research on time distortions in a multimodal context, such as in film, where auditory and visual cues are present simultaneously and can interact. Droit-Volet et al. (2011) used films as a mood-eliciting technique to evaluate the effect of mood on duration perception. They recorded participants' subjective mood ratings and performance on a temporal bisection task where participants rated a neutral probe (blue circle) as *long* or *short* based on memorized durations. Participants' subjective time estimates in this task increased after watching fearful, but not sad or neutral, films. Wöllner, Hammerschmidt, and Albrecht (2018) examined the impact of visually-stretched time (also known as 'slow-motion') on duration estimates and judgments of emotional affect in film clips. Overall, the film clips were rated as shorter than actual durations, and slow-motion scenes produced underestimates of duration compared to scenes in real time. When a musical score was added to the film extracts (in a multimodal condition), participants judged the duration of these clips as longer than in a visual-only (unimodal) condition. The addition of music also led to increases in physiological arousal as indexed by changes in galvanic skin responses, heart rate, respiration rate, and pupil diameter. The authors did not examine whether physiological arousal impacted perceived duration nor did they investigate time estimates or physiological measures in an auditory-only (unimodal) condition.

The presence of music has become an expected feature of modern films. In fact, the use of music in film is so ubiquitous that its absence will often be perceived (Beeman, 1981). The different parameters of music can affect our perception of time. For instance, Jones and Boltz

(1989) suggest that the effect of music on time perception is driven by the perceptual expectancies that listeners build when they hear a piece of music. Harmonic patterns and rhythmic accentuations lead subjects to anticipate particular musical events, and when these events happen earlier than expected, time will be judged shorter - and vice versa. Moreover, music is often used in supermarkets and waiting rooms to reduce the subjective experience of time, in order to encourage consumerist behaviors (e.g. Oakes, 2003). Music is therefore an important feature of film that should be taken into consideration when studying time perception in film.

In this paper, we ask how the added presence of the auditory modality (which always included music) to a visual scene influences subjective estimates of duration of that scene. Participants estimated the duration of film extracts presented unimodally (auditory only, visual only) or multimodally (audiovisual) and rated their subjectively experienced degree of arousal and valence in response to each clip. We hypothesized based on the findings of Wöllner et al. (2018) that after controlling for arousal and valence, duration estimates would be lengthened in the multimodal condition as compared to either of the unimodal conditions. In addition, we predicted that arousal would modulate these duration estimates. Arousal was measured by pupillometry, an appropriate index of modulation in physiological arousal levels (Aston-Jones & Cohen, 2005), paired with self-reports of subjective arousal. Finally, we assessed whether the strength of the modulating relationship between arousal and time estimation differed between modality conditions. We hypothesized that variations in arousal would influence subjective estimates of time, and that the effect of arousal would be strongest in the multimodal condition.

Methods

Participants

This study was approved by the Institutional Review Board at Brown University and all participants provided written informed consent. A total of 30 participants (14 male, 16 female) from Brown University (mean age: 20.7, SD = 1.5) participated in the study. One male participant's responses were removed from data analysis because of missing pupil data due to equipment malfunction, and one female participant's data were removed because of exceedingly high standard deviations in subjective ratings in one condition, resulting in a total N of 28 participants. Participants were tested individually in a single 40 min long session. They were compensated $10 for their participation or received course credit as compensation.

None of the participants had severe visual impairment and none reported chronic fatigue, both of which have been shown to influence pupillometry measurement in the context of auditory perception (Zekveld, Koelewijn, & Kramer, 2018). One participant had a bone-anchored device but had high indices of music listening and instrument practice (which we suggest indicates that the device did not affect the participant's music perception or propensity to be aroused by musical stimuli), so the data from this participant were preserved in the analysis.

Stimulus clips

Stimuli for this study consisted of film clips from the LIRIS-ACCEDE database (Baveye, Dellandrea, Chamaret, & Chen, 2015). This database is composed of 9800 clips shared under Creative Commons licenses from films that are often little known, which limits the potential effect of familiarity on subjective ratings and physiological measurements. These films had been annotated previously through online surveys along the two affective dimensions of the affective circumplex model (Russell, 1980). For our experiment, we chose a subset of 27 clips, equally distributed among three valence settings (low, neutral, and high) and three arousal levels (low, neutral, and high), as annotated in the LIRIS database. We excluded clips from the horror genre or any other clips that might cause discomfort in the viewer, and kept only clips with a musical soundtrack. The clips varied between 5 and 13 sec in length, durations the authors of the database considered to be long enough to make the viewer feel strong emotions but short enough to elicit only one emotion per excerpt (Baveye et al., 2015). Examples of selected clips are shown in **Fig. 1A**.

For each of the clips, we then generated three viewing conditions that differed in their perceptual modality using Adobe Premiere CC. These three conditions are audiovisual (AV), visual-only (VO; here the music track was deleted), and auditory-only (AO; here a black screen replaced the visual scene). In the AV and AO conditions, the audio gain on the soundtrack was normalized using the Adobe Media Encoder CC peak normalizer. The pool of participants was divided into three groups of 10. Each group viewed clips in all three conditions (thus condition is a within-group variable), but each individual clip was shown in only one group-determined condition (**Fig. 1A**). For example, if group A saw *clip 1* in AO, they did not see that same clip in AV or VO. Participants were assigned to a group based on the time of their session, and each participant saw the clips in a different (randomized) order (**Fig. 1A**).

Procedure

The experimental task was explained to participants using a written script and written informed consent was acquired. They were seated in a comfortable chair facing a 24” computer screen (1152 x 864 pixel resolution, 72Hz, Viewsonic G90fb), with their heads stabilized using a chin rest located 60 cm from the computer screen. Stimuli and judgment tasks were presented using E-Prime 3 Pro software (Psychology Software Tools, Pittsburgh, USA). Eye position and pupil size of the left eye were recorded at 1000 Hz with a desk-mounted Eyelink 1000 (SR Research, Ottawa, Ontario, Canada) controlled by custom scripts using the E-Prime software. For one participant, we recorded the right eye due to simple anisometropia in their left eye.

In each experimental trial (**Fig. 1B**), participants viewed first a grey screen with a fixation cross for 2 sec (2000 ms); we refer to this period in the trial as the *baseline period*. Next, a film clip was presented (durations of 5200 ms to 12130 ms), followed by another grey screen and fixation cross for 4 sec (4000 ms); we refer to this post-clip period as the *test period*. Participants were then asked to make three cognitive judgments of each clip. In the first judgement task, participants were asked to estimate the Duration (in sec, 0-25 sec in one decimal increments) of the clip, via a slider controlled by the computer mouse. In the second task, they were asked to rate on a continuous scale their subjective experiences of Arousal [*How did this clip make you feel? - calm (0), neutral (3), aroused (5)*], and in the third task, they were asked to rate their subjective experiences of Valence [*How did you find this clip? - unpleasant (0), neutral (3), pleasant (5)*]. Arousal and Valence judgements were acquired using two separate sliders (1.0 to 5.0, with a one decimal increment). Five sec of silence and a grey viewing screen (with no fixation cross) followed, to allow pupil size to return to baseline and to stabilize before the next trial.

Eye tracker data acquisition and preprocessing

At the onset of the experiment, the eye tracker system was calibrated and validated to average visual angle Cartesian prediction error < 1° for each participant with a 9-point grid calibration. We offline preprocessed the eye position and pupil size data using custom MATLAB scripts. In order to ensure that our measurement of pupil size was not contaminated by artifacts such as blinks and saccades, we implemented a custom velocity-based thresholding procedure to identify the presence of such artifacts in the eye position time series (Smeets & Hooge, 2003). Artifacts lead to a significantly larger velocity relative to the average velocity when an observer maintains fixation. On each trial, we computed the velocity at each time point by applying a differential procedure to the position location data. The resulting time series was subsequently submitted to a 2nd order Butterworth filter (cutoff of 10 Hz) in order to remove high-frequency artifacts in the velocity profile. To detect artifacts in the eye position data, we computed the mean and standard deviation of the velocities across the time series for each trial and used these measures to calculate a trial-specific velocity threshold against which the velocity at each time point was compared. Inspection of the data revealed that a threshold of 2.5 standard deviations above the mean provided accurate classification of artifacts. Any time point corresponding to a velocity above the threshold was defined as an artifact, and onsets and offsets were defined by determining the first and last sample in a series of subsequent samples that exceeded the threshold. We removed all samples corresponding to time points where an artifact occurred for all subsequent analyses that involved pupil size measurements.

Pupil size varies not just as a function of acute changes in arousal in response to a stimulus, but also according to a number of confounding variables, such as longer-term fluctuations in arousal and lighting conditions (Aston-Jones & Cohen, 2005; Peysakhovich, Vachon, & Dehais, 2017). Moreover, pupil size as measured by the Eyelink 1000 system is defined in arbitrary units. In order to classify pupil size measurements in meaningful units and to reduce the influence of confounding variables, we calculated a relative measure of pupil size at each time point by subtracting the average pupil size in the 2000 ms baseline period from the pupil size during the test period. We measured pupil size during the clip as well; however, we include pupil size only from the test period in our subsequent analyses (see below, Temporal analysis of pupil size). The resulting values provide a measure of the change in pupil size relative to baseline, where negative values indicate a constricting of the pupil and positive values indicate pupil dilation. We used this relative pupil size in all subsequent analyses involving pupil size; for simplicity, we refer to this simply as pupil size.

Data Analysis

Statistical power

In the current study, three a priori effects were of interest. First, we predicted increased time estimation in the multimodal condition compared to unimodal conditions. Second, we predicted a significant association between arousal and time estimation. Third, we hypothesized that the strength of the association between arousal and time estimation could differ between sensory-modality conditions. With 28 subjects and 9 clips in each condition, the statistical power of detecting these three effects can be estimated in closed form under the repeated-measures ANOVA framework as within-subject factors. Using the G*power package and setting alpha at 0.05, a small effect (f = 0.10) has power of 0.15; a medium effect (f = 0.25) has power of 0.80, and a large effect (f = 0.40) has power of 1.00 (Faul, Erdfelder, Lang, & Buchner, 2007).

A previous study on the influence of modality by Wöllner et al. (2018) reported a large effect size of ($\eta^2_p = 0.20$, equivalent to f = 0.50). In addition, the primary hypothesis testing models employed in the current study are based on linear mixed-effect models that could offer better statistical power than repeated-measures variance tests (Muth, Bales, Hinde, Maninger, Mendoza, & Ferrer, 2015). Due to the numerical nature of fitting linear mixed-effect models, a direct a priori power calculation is not possible. For non-significant findings of interest in the current study, we conducted post-hoc numerical simulations to determine the statistical power for the examined effect in the linear mixed-effect models under various effect sizes (Brysbaert & Stevens, 2018; Green & MacLeod, 2015).

Statistical software

All data analyses were conducted using custom scripts written in the R programming language (Venables et al., 2013). Figures were generated using the ggplot2 package (Wickham, 2016). Mixed-effects models were fit using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) and statistical significance was assessed using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017).

Time Estimation Index

To quantify the accuracy of our participants' estimates of clip duration, we defined a Time Estimation Index that controlled for the fact that each clip varied in actual duration. Time Estimation Index was defined for each trial as the participants' estimate of clip duration on each trial divided by the clips' actual duration, minus 1. A negative value means that the observer underestimated the clip duration, a positive value indicates the observer overestimated the clip duration, while a value of 0 indicates the participant perfectly estimated clip duration.

Statistical modelling

For statistical analyses that included predictors with continuous variables (Clip Duration, Arousal, and Valence), we used linear mixed-effects models. After initially choosing 27 clips which were displayed during the experiment, we excluded two clips from subsequent analyses, because these were the only ones containing dialogue and written text. This resulted in a stimulus set of 25 unique film clips for each participant. Since the excluded clips resulted in uneven trial numbers in different conditions for different subjects, the excluded clips were treated as missing data in the linear mixed-effects models. Mixed-effects regression is an appropriate analysis technique for use in repeated measures designs because it employs a model structure that allows one to control for variance within each participant's responses (by controlling for the random effect of participant) as well as across participants (by specifying fixed effect predictors). The resulting regression coefficients (termed fixed effects) represent the magnitude of change associated with a unit change in each predictor variable after controlling for each of the other predictors. We used mixed-effects regression in two separate contexts: (1) to measure the temporal evolution (within the test period) of the predictiveness of pupil size coefficient on Time Estimation Index; and (2) to determine how predictor variables influence Time Estimation Index.

When testing whether a parameter associated with a fixed effect is significantly different from 0, we must know the null distribution of the parameter estimates and their associated test statistics. It is generally not possible to know the exact form of these distributions when employing mixed-effects modeling. To assess statistical significance, we approximated the relevant null distributions using Satterthwaite's method (Satterthwaite, 1946) using the lmerTest package.

Where appropriate, we used repeated-measures analysis of variance (RM-ANOVA) to compare the differences between levels of discrete predictor variables (AV, AO, VO conditions). All post-hoc pairwise comparisons (e.g., t-tests) were Bonferroni-controlled for multiple comparisons.

Temporal analysis of pupil size

We first determined the predictive effect of pupil size on Time Estimation Index while controlling for the other variables at each timepoint. This novel regression modelling analysis was inspired by previous investigations that predicted task performance from pupil size measurements (He, Heindel, Nassar, Siefert, & Festa, 2020). We include pupil size only from the test period (3rd panel in **Fig. 1B**) to ensure that measurements were not contaminated by differences between clips that were unrelated to our hypotheses (e.g., bright vs. dark luminance background). At each timepoint (ms) in the 4000 ms test period (3rd panel in Fig. 1B), we fit a linear mixed-effects model of the following form:

Time Estimation Index = β1 Intercept + β2 AV:AO + β3 AV:VO + β4 Pupil Size + β5 Valence + β6 Duration + β7 Participants (random variable)

To determine how the predictiveness of pupil size evolved over time, we extracted the p-value of the parameter associated with pupil size (*β4 Pupil Size*) at each timepoint and determined significance by comparing the p-values against a criterion of 0.05. We defined clusters by finding runs of significant timepoints. To do so, we calculated all onsets, defined as the first time point in which significance was achieved, and offsets, defined as the last time point of significance, across the full 4000 ms test period.

To control for multiple comparisons in the above analysis, we implemented a cluster-size correction permutation analysis (Nichols & Holmes, 2001). We set out to define a null distribution of cluster sizes that specified what the expected distribution of cluster sizes would be under the null hypothesis that pupil size does not predict Time Estimation Index. For each of 1000 permutations, we randomly flipped the sign of the pupil size measurement at each time point for each trial for a randomly selected half of the participants included in this analysis (14 per permutation). The regression model described above was fit at each time point in the sign-flipped time series, and we determined the number and length of all clusters of significant pupil size coefficients across the time series in each permutation. Finally, we compared the cluster sizes for each cluster we observed in the empirical time series against the 95th percentile of this distribution (corresponding to a cluster length of 204 consecutive significant timepoints).

Principal Component Analysis of arousal measurements

On each trial, we collected two measurements of arousal: (1) the subjective report of the arousal participants experienced in response to each clip, and (2) the time series of pupil size measurement collected with the eye tracker. Given that it is likely the participant's arousal state produced correlated fluctuations in both of these measurements, we decided a priori to use a Principal Component Analysis (PCA) to extract the two orthogonal underlying components that uniquely explain variance in the dataset. For each trial, we calculated the average pupil size during the test period for the time points corresponding to the largest cluster where pupil size significantly predicted Time Estimation Index (from 376 ms to 2334 ms after clip offset). After z-scoring, the average pupil size and the self-reported arousal on each trial were submitted to PCA, resulting in two principal components (PCs; see **Fig. 2B** for a graphical depiction). The first PC accounted for 54% of the variance and was positively related to both subjective report and pupil size, yielding what was essentially an average of these two variables. We term this PC Composite Arousal. The second PC accounted for the remaining 46% of the variance and was positively related to subjective report while negatively related to pupil size, which essentially acts to control for the first PC; we refer to this as PC2 for simplicity. We calculated each trial's PC scores using the respective loadings for inclusion in our subsequent analyses.

Hypothesis testing

In order to determine how predictor variables influence Time Estimation Index, we fit two mixed-effects models that differed on whether we included an interaction between condition and Composite Arousal. Including interaction terms in a linear model slightly alters the interpretation of main effects in a three-level factor design. Hence, we report results from both models for completeness. We refer to these models as the Independent Effects Model and Interaction Model for simplicity. It should be noted that these two models do not constitute repeated hypothesis testing, as we only interpret the interaction terms from the Interaction Model.

The Independent Effects Model was fit using the following equation :

Time Estimation Index = β1 Intercept + β2 AV:AO + β3 AV:VO + β4 Composite Arousal + β5 PC2 + β6 Valence + β7 Duration + β8 Participants (random variable).

The Interaction Model was fit using the following equation:

Time Estimation Index = β1 Intercept + β2 AV:AO + β3 AV:VO + β4 Composite Arousal + β5 PC2 + β6 Valence + β7 Duration + β8 AV:AO x Composite Arousal + β9 AV:VO x Composite Arousal + β10 Participants (random variable)

Prior to inclusion in our models, we mean-centered the continuous predictor variables (Composite Arousal, Valence, and Duration) to ensure that the coefficients were interpretable across the two models. Since condition is a discrete variable with three levels, we included parameters that represent the planned contrasts between AV and AO (*β2 AV:AO*) and between AV and VO (*β3 AV:VO*), which represent the change in Time Estimation Index between the respective conditions at the mean value of each of the continuous predictor variables. Moreover, the Interaction Model included two additional parameters, *β8 AV:AO x Composite Arousal* and *β9 AV:VO x Composite Arousal*, which capture the changes in the association strength of the Composite Arousal variable between the respective conditions. We chose the maximal number of random-effects parameters that (1) controlled for between-participant differences in average performance and (2) afforded converging parameter estimates across multiple iterations of model-fitting. As a result, the Participants variable (Independent Effects: *β8*, Interaction: *β10*) corresponds to fitting unique intercepts for each participant for each predictor variable.

Results

Temporal analysis of relative pupil size

On each trial, pupil size was measured throughout the duration of the clip and the 4000 ms test period preceding the judgement task (3rd panel in **Fig. 1B**). **Fig. S1** depicts the temporal evolution of pupil size averaged across all participants and trials over this interval. This depiction demonstrates that pupil size fluctuated in a characteristic manner in the test period. Pupil size was large immediately following the film clip and then decreased dramatically, likely reflecting changes brought on by the visual transient that occurred in the transition from film clip to the grey background presented during the test period. Between 800 ms and 3000 ms of the test period, pupil size increases. Subsequently, as time from the film clip increased, pupil size decreased.

We first set out to determine whether fluctuations in pupil size in this test period predicted fluctuations in participants' duration estimates of the film clips and to characterize the temporal dynamics of this relationship. To do so, for each time point (ms) in this 4000 ms time window, we fit a linear model predicting Time Estimation Index from the fixed effects of pupil size and covariates of condition (AV, VO, or AO), clip Duration, and self-reported Valence as well as the random effect of participant. **Fig. 2A** depicts the pupil size parameter value as a function of time in the 4000 ms test period. The insets depict the model fits for two example timepoints to demonstrate how the relationship between pupil size and Time Estimation Index varies as a function of time. For example, at 135 ms after the onset of the test period, the pupil size parameter is slightly negative, indicating that as pupil size increases, Time Estimation Index decreases. However, at 858 ms into the test period, the pupil size parameter is positive, reflecting the fact that Time Estimation Index increases with increasing pupil size.

The temporal profile of the significance of the pupil size predictor provides a measure of the dynamics of the predictive relationship of pupil size on Time Estimation Index. Significant time points are depicted by black lines in **Fig. 2A**. Immediately after clip offset, pupil size does not predict Time Estimation Index, as reflected by p-values that exceed the criterion. However, beginning at 376 ms, there is a sustained significance of the pupil size predictor that continues until 2334 ms, after which p-values fluctuate above and below criterion. To further quantify the dynamics of this relationship, we defined temporal clusters by determining periods of repeated significance. The onset of a cluster was defined as the first time point in which significance was achieved, while the offset was defined as the last time point prior to the p-value being equal to or greater than criterion. This analysis yielded 10 clusters of significance ranging in size from 2 ms to 1959 ms. Overall, we observed that pupil size significantly predicted Time Estimation Index for 82.4% of the entire time window (3297 out of 4000 ms).

To ensure that this temporal analysis did not yield spurious results, we conducted a cluster size permutation test that allowed us to define a distribution of expected cluster sizes under the null hypothesis that pupil size does not predict Time Estimation Index, as described above (Methods: Temporal analysis of pupil size). Three clusters in the empirical data set were large enough to be deemed significant: an initial cluster that began early and extended for 1959 ms (ranging from 376-2334 ms relative to clip offset, indicated in **Fig. 2A** by the first black line), a subsequent cluster of length 873 ms that began almost immediately after the initial cluster (2337-3210 ms), and a cluster of length 354 ms that peaked later in the test period (3381-3735 ms). Of note, the two largest clusters occurred sequentially in time separated by a period of 2 ms where significance was not achieved (depicted by a gap in the black lines above the graph in **Fig. 2A**). Using this cutoff for cluster length, pupil size significantly predicted Time Estimation Index for 79.7% of the test period (3186 out of 4000 ms).

Composite Arousal

We collected two measures of arousal on each trial: a subjective report and the time series of pupil size measurements. Participant's arousal state likely produced correlated fluctuations in both of these measurements, so we a priori decided to use a Principal Component Analysis (PCA) to extract the two orthogonal underlying components that uniquely explain variance in the dataset (**Fig. 2B**). The PCA allowed us to isolate the underlying arousal component, which we term Composite Arousal. We next assessed whether Composite Arousal differed across conditions. **Fig. 2C** depicts the probability density distribution of Composite Arousal for each condition. Composite Arousal was highest in the AV condition (mean = 0.415, SEM = 0.09), intermediate in the AO condition (-0.079, 0.12), and lowest in the VO condition (-0.335, 0.07). RM-ANOVA confirmed that this difference across conditions was significant ($F_{(2, 25)} = 25.27$, $p < 0.001$). Post-hoc pairwise t-tests revealed that significance was driven by higher Composite Arousal in the AV condition compared to both the AO condition (AV-AO: $t_{27} = 4.10$, $p < 0.001$) and VO (AV-VO: $t_{27} = 7.64$, $p < 0.001$) conditions. The difference between the AO and VO conditions was marginally significant (AO-VO: $t_{27} = 2.48$, $p = 0.06$).

Time Estimation - The effect of arousal and condition

In order to determine which variables contributed to our participants' duration judgments, we fit a linear mixed-effects model predicting Time Estimation Index from predictor variables condition (AV, VO, or AO), Composite Arousal, Valence, and Duration; we refer to this model as the Independent Effects Model. The results are displayed in the second column of **Table 1A.** As shown in this table, all predictors in this model significantly predict the Time Estimation Index.

We were primarily interested in the predictiveness of two variables: condition and Composite Arousal. After controlling for the influence of the other fixed and random effects, parameters pertaining to our planned contrasts between different conditions were both significant [AV:AO, $\beta = -0.031$, df = 666.754, S.E. = 0.014, $p < 0.05$; AV:VO, $\beta = 0.035$, df = 666.756, S.E. = 0.015, $p < 0.05$]. These results indicate that the average Time Estimation Index differed between the AV condition and the AO and VO conditions, respectively. To depict the average Time Estimation Index between each condition while controlling for confounding variables, the model-derived estimates of the average Time Estimation Index (intercept value) for each condition are displayed in **Fig. 3A**. As shown in this figure, in all conditions, clips tended to be judged shorter than their actual duration as indicated by coefficient values less than 0. The model-derived intercept values for each condition are reported on the right side of the graph. As indicated, the Time Estimation Index was shortest in the AO condition [$\beta_{AO} = -0.371$], intermediate in the AV condition [$\beta_{AV} = -0.340$], and longest in the VO condition [$\beta_{VO} = -0.305$]. Overall, these results indicate that time perception varied according to the perceived sensory modality of the clip.

The effect of Composite Arousal was also significant [$\beta = 0.030$, df = 676.139, S.E. = 0.006, $p < 0.001$], indicating that increased arousal is associated with increased subjective estimates of duration. To assess whether the influence of Composite Arousal varied between conditions, we fit a second linear mixed model that included two additional interaction parameters that compared whether the slope of the Time Estimation Index by Composite Arousal relationship differed between the AV condition and the AO and VO conditions, respectively. The results of this Interaction Model are shown in **Table 1B**. Both interaction coefficients were significant [AV:AO x Arousal, $\beta = -0.042$, df = 669.178, S.E. = 0.015, $p < 0.01$; AV:VO x Arousal, $\beta = 0.032$, df = 667.958, S.E. = 0.016, $p < 0.05$], indicating that the effect of arousal on Time Estimation Index differs between conditions (**Fig. 3B**). The direction of this interaction suggests that in the AO and VO condition, there is a weaker positive effect of arousal on the Time Estimation Index compared to the multimodal condition (AV; compare the slopes of the lines in **Fig. 3B**). In this figure, the steepest slope corresponds to the effect of Composite Arousal on the Time Estimation Index in the AV condition, suggesting that the positive effect of arousal is stronger in a multimodal context [Slope VO = 0.016; Slope AO = 0.026; Slope AV= 0.058].

Once the interacting effect of Composite Arousal is partialed out in the Interaction Model, there no longer is a significant difference between AV and AO (compare the values of the AV:AO parameters for each model in **Table 1**). We investigated whether the statistically non-significant result was due to underpowered tests to detect small effect sizes at the lower arousal regime being examined in the Interaction Model. We conducted post-hoc simulated power analyses of the AV:AO contrast and found an observed power of 0.266 (95% confidence interval: 0.239-0.295) for this fixed effect in the Interaction Model. We repeated the power calculations with small ($R^2 = 0.019$; $\beta = -0.07$) and medium ($R^2 = 0.135$; $\beta = -0.2$) effect sizes, obtaining simulated statistical powers at 0.98 and 1.0, respectively. Thus, this statistical power analysis shows that even if there is a true difference in Time Estimation Index between AV and AO conditions, the effect size is negligibly small. The employed linear mixed-effect models are sufficiently powered to detect an effect size of typical small magnitudes.

Additional findings

We also found significant effects of clip duration on Time Estimation Index. [$\beta = -0.019$, df = 664.021, S.E. = 0.004, $p < 0.001$]. This suggests that participants underestimate longer durations more than shorter durations. Valence is also a significant predictor, with increased valence associated with an increase in Time Estimation Index [$\beta = 0.013$, df = 671.037, S.E. = 0.006, $p < 0.05$]. The effects of these variables were identical in the two models (Table 1), since these variables were not involved in the interaction terms. We are reporting above the results from the Independent Effects Model.

As a post-hoc analysis, we tested whether the number of editing cuts in each clip (film transitions from one viewpoint to another as present in the original LIRIS database) affected Time Estimation Index. An additional linear mixed model (limited to trials containing clips viewed in AV and VO conditions, as no cuts were visible in the AO condition), revealed that the number of cuts predicted duration estimates, after controlling for variation in duration of clips and other fixed and random variables included in the previous models [$\beta_{434.9} = 0.019$, df = 434.9, S.E. = 0.005, $p < 0.001$]. This exploratory analysis suggests that an increase in the number of cuts can contribute to longer duration estimates.

Discussion

Our data show that participants' estimates of the duration of film clips differed from the actual durations, showing under-estimates in all stimulus conditions. Duration estimates varied according to the sensory modality of the film clips, with clips in the VO condition leading to the longest duration estimates (least under-estimated). Moreover, Composite Arousal positively influenced duration estimates across conditions and the effect of Composite Arousal on duration perception was strongest in the multimodal (AV) condition.

1- Crossmodal differences in duration perception

Our results indicate that the average duration estimates differed between the AV condition (multimodal) and the AO and VO conditions (unimodal), respectively. However, when we controlled for interactions between conditions and Composite Arousal, the difference in Time Estimation Index between AV and AO was no longer significant (see AV:AO result in **Table 1**). Two possible interpretations arise from this finding. First, the difference between AV and AO (as shown in **Fig. 3A**) is not independent of the effect of arousal - i.e., there is nothing inherently different between the two conditions that produces a difference in duration estimates that is independent of Composite Arousal. Alternatively, the null result in the Interaction Model is due to underpowered testing while the significant difference between the intercepts of AV and AO in the Independent Effects Model (**Fig. 3A**) represents more accurately the psychological differences in time perception between these two conditions. Our simulated power analysis confirms that we are sufficiently powered to detect a difference of small effect size between AV and AO. Our results therefore do not support a large enough difference between AV and AO conditions, and favor the interpretation that the difference between AV and AO is statistically interacting with the effect of arousal. On the other hand, the difference between the AV and VO condition remained significant in both models, suggesting that, after controlling for other variables, the VO condition led to the longest duration estimates (the VO clips were underestimated less).

Our findings concerning duration perception in the context of natural stimuli - film extracts - are thus at odds with the classic finding that “sounds are judged longer than lights” (Lustig & Meck, 2011). This was a surprising finding, as we had hypothesized, based on the results of Wöllner et al. (2018), that the addition of music to the film clips (multimodal condition) would yield an increase in duration estimates compared to duration estimates of unimodal clips. These differences between studies may reflect differences in experimental design, including the small number of excerpts (9 clips) used by Wöllner et al. (2018) as opposed to the 27 in our study, and the difference in clock-time duration ranges of the selected excerpts (267 - 40000 ms as opposed to 520 - 1213 ms in our study). Wöllner et al. (2018) found that the AV condition yielded higher physiological responses and increased duration estimates, and interpreted this to indicate that there was a correlation between physiological responses and duration estimates. In other words, they suggested that the increased duration estimates in the AV condition were mediated by the increased arousal observed in this same condition, however, they did not test statistically for a correlation between arousal and duration estimates. We indeed found a significant effect of Composite Arousal on duration estimates (as well as a significant interaction between Composite Arousal and condition). Given that arousal positively influences duration estimates, we suggest that the higher duration estimates in the AV condition (compared to the VO condition) in their experiment could be explained by the higher physiological measurements (arousal) revealed in the AV condition, rather than a condition-driven difference. Their study lacked a control for the effect of arousal on duration estimates which didn't allow them to discern the effect of arousal from the modality-driven effects on time perception.

Similar to our findings, Sanders and Cairns (2010) found that the addition of music to a video game caused players to underestimate the duration of play compared to the no-music condition. Interestingly, this effect was limited to their prospective timing experiment (where the participants were aware that they needed to make a duration estimate beforehand) whereas the music had no effect on duration estimates in retrospective timing tasks (where the participants learned after experiencing a duration that they needed to make a duration estimate) - i.e., music had an effect on experienced duration but not remembered duration. This finding, they argue, supports the theory that the two timing paradigms use different cognitive mechanisms to estimate duration. One variable has been found to influence prospective duration estimates but not retrospective ones, namely the complexity of the task (Block & Zakay, 1997). Increasing the processing complexity of a task decreases the prospective duration estimates. One possible interpretation of this phenomenon is that the processing complexity of a task limits the amount of attention participants can allocate to the monitoring of time. Similarly, shorter duration estimates in the AV condition (as opposed to VO) in our experiment could result from music attracting attention away from the passing of time such that the participant is less reliably able to estimate durations.

This interpretation would validate the notion in the film industry that music in film has, amongst other functions, an immersive role (also termed ‘absorption’) by making viewers less aware of time. Indeed, film-music’s emotional potential, as well as its effect on attention, are becoming more obvious (Cohen, 2010). It is thought that music elicits affective responses that in turn influence visual attention and physiological measures of the eye such as saccades and eye-blinks. For instance, eye-tracking studies show that emotions affect the way individuals attend to visual details (Ford, Tamir, Brunye, Shirer, Mahoney, & Taylor, 2010) and that music, as compared to silence, can lead to reduced eye-movement (longer fixations), fewer saccades, and increased number of eye-blinks (e.g., Mera & Stumpf, 2014; Schafer & Fachner, 2014). Furthermore, music can be used to guide attention by altering eye position, fixation duration, and the amplitude of saccades (Coutrot, Guyader, Ionescu, & Caplier, 2012).

The idea that the difference between the AV and VO condition could arise as a result of music’s effect on attention is in line with a predominant psychological model that explains the influence of attention as well as arousal on duration perception - the pacemaker accumulator model (reviewed in Lake, 2016). According to this model, the rate of an internal time-keeping mechanism (pacemaker) determines the perception of duration. Attention is thought to influence the activity of a gate that modulates the number of pulses resulting from the pacemaker, that are fed to an accumulator (timekeeper). The accumulator collects a number of pulses which allows us to make duration estimates. Attention influences how many pulses are fed to the accumulator, such that directing the attention away from timing would result in a shortening effect - fewer pulses and shorter duration estimates (Brown, 1997). While our participants were explicitly asked not to count and to pay attention to the stimuli, they were aware that they had to give a duration estimate and must have allocated some attention to the passing of time. As we suggested, the presence of music could have modulated the amount of attentional resources attributed to the monitoring of time passing, but future studies should investigate the effect of attention-mediated changes of time perception in film.

Another possible interpretation as to why the AV condition yielded shorter duration estimates (**Fig. 3A**) compared to the VO condition (after controlling for other variables) is that participants favored one modality over the other in estimating duration. Studies have suggested that an internal auditory clock and visual clock run at different paces, potentially explaining modal differences in duration estimates (Penney et al., 2000; Wearden et al., 2006). Which clock do participants use to determine the perceived duration of a multimodal event, such as in film, where auditory and visual cues are presented simultaneously and can interact? Do viewers rely on a separate audiovisual clock (as hypothesized by Klink, Montijn, & Wezel, 2011)? Do viewers rely more on the auditory clock, meaning that a multimodal stimulus is perceived as having the same duration as the auditory stimulus? Is it perceived as having the same duration as the visual stimulus, or is it perceived as having a duration somewhere in between? While the Independent Effects Model suggested the latter to be true (**Fig. 3A**), the Interaction Model raises the possibility that differences in duration estimates between the AO and AV conditions are not independent of the effect of Composite Arousal (**Table 1**). When a stimulus is perceived through two perceptual tracks - auditory and visual, in our case - studies suggest that auditory processing dominates over visual processing in the perception of durations regardless of selective attention and temporal discriminability (Burr, Banks, & Morrone, 2009). This is in line with the modality-appropriateness hypothesis that suggests that auditory information is given priority over the visual modality in temporal tasks rather than spatial tasks, and vice-versa (Welch & Warren, 1980).

In a recent study, Amadeo, Campus, and Gori (2020) posited that supramodal recruitment of the auditory network is necessary for dealing with temporal representations in the visual system, suggesting that an early recruitment of the auditory cortex is implicated in temporal tasks regardless of the modality. The dominance of audition in temporal tasks is further supported by the fact that auditory training alters visual rhythm perception but not the other way around (Barakat, Seitz, & Shams, 2015). The lack of a significant difference between the AO and AV conditions in the Interaction Model (see AV:AO result in **Table 1**) might suggest that the brain prioritizes the auditory component in estimating time in film. This reliance on auditory time could explain why the duration estimates in a multimodal context were similar to estimates in the auditory-only condition - i.e. shorter than in the VO condition.

2- The effect of Composite Arousal on duration perception

Another central aspect of our research concerns the effect of arousal on duration perception in film. In our experiment, we collected two measures of arousal on each trial: a subjective report and the time series of pupil size measurements. There is extensive literature on understanding the role of arousal in psychological phenomena (including time estimation; see Lake, 2016 for review) and subjective reports and physiological measurements can provide valuable complementary measurements of psychological variables. Physiological measurements of arousal (such as pupil size) are not subject to the fallacies of self-report, like memory decay and noise due to response selection, and are therefore generally considered to be a more 'objective' measurement of arousal than self-report. For instance, in the context of auditory time perception, there are disagreements concerning the role of subjective arousal: Droit-Volet et al. (2013) correlated increased subjective arousal to higher subjective duration estimates, while Noulhiane, Samson, Mella, Rago, and Puthas (2007) found that subjective arousal had the contrary effect - high-arousal stimuli were judged shorter than low-arousing ones. In this study, while reported arousal was negatively correlated with estimated length, the auditory stimuli that were determined by the researchers as "emotional" were consistently judged to be longer than neutral ones. Noulhiane, Samson, Mella, Rago, and Puthas (2007) suggested that this effect reflected the predominant role of activation - physiological arousal - on time distortions, but provided no physiological measurement to back their claim. Given that the participants' arousal state likely produced correlated fluctuations in both the subjective and physiological measurements, we implemented a PCA to extract the underlying component corresponding to arousal, which we termed Composite Arousal. Composite Arousal was positively related to both subjective report and pupil size yielding what was essentially an average of the two variables. The resulting measure of arousal is a more rigorous measurement of the underlying arousal mechanisms influencing subjective and physiological measurements, as well as time perception, than simple subjective ratings.

Composite Arousal positively influenced duration perception across conditions. The pacemaker accumulator model offers a framework for understanding our behavioral results: Arousal is thought to determine the rate of activity of the pacemaker such that increased arousal leads to an increase in the number of pulses passing through the ‘gate’ and fed to the accumulator, yielding longer duration estimates (e.g., Zakay, 2005). The pacemaker accumulator model is often incorporated in behavioral and psychophysiological studies (Fontes et al., 2016); however, it has also been criticized for its reliance on interpretation from behavioral and psychophysical data and the lack of biological plausibility. In recent years, however, studies have begun to shed light on the neural circuitry of emotionally-driven time distortions. The anterior insula has emerged as a potential candidate for the integration of emotional experiences within an internal representation of time (Lake, 2016). The Striatal Beat Frequency Model (SBF) is another model that accounts for the influence of arousal on duration perception which relies on dopamine levels as driving the time distortions. Emotional events, such as our film excerpts, induce phasic changes in dopamine levels in the striatum, cortex and prefrontal cortex (Cheng, Tipples, Narayanan, & Meck, 2016). Such changes are thought to modulate the firing rate of cortical projections into the striatum, which in turn alter the rate of cortical oscillations and time distortions (Ogden, Henderson, McGlone, & Richter, 2019). As dopamine increases, so does the oscillation rate, leading to a perception of durations that are longer, whereas decreases in dopamine will slow the oscillation rate, leading to a shorter perception of durations.

The effect of arousal on time perception is well established - arousal leads to dilation of time. With the Interaction Model, we asked whether the rate of the influence of Composite Arousal on duration estimates differs between conditions. Results of this model suggest that in the AO and VO condition, there is a weaker positive effect of Composite Arousal on duration estimates compared to the multimodal condition. The steepest slope corresponds to the effect of Composite Arousal on the duration estimates in the AV condition (see **Fig. 3B**), suggesting that the positive effect of Composite Arousal is strongest in a multimodal context - i.e. arousal influences perceived duration more strongly when both auditory and visual information are simultaneously presented.

The differences between sensory modalities in the rate of the arousal effect suggests the existence of distinct neural underpinnings for our capacity to estimate duration in different sensory modalities. We found no research substantiating this novel inquiry, nor explaining differences in the rate of the effect of arousal on perceived duration, but previous research on multimodal perception have found distinct physiological and neurological activation in the context of audiovisual stimuli as a result of a bimodal interaction (Eldar, Ganor, Admon, Bleich, & Hendle, 2007; Baumgartner, Esslen, & Jäncke, 2006; Vines, Krumhansl, Wanderley & Levitin, 2006) including increased activity in brain areas associated with emotional processing (automatic ventral system of emotional perception), the medial temporal lobe memory system, and extrastriate visual processing areas (Baumgartner, Lutz, Schmidt, & Jäncke, 2006). Future studies should investigate the neural underpinnings of the increased rate of arousal-mediated distortions in a multimodal context.

3- Additional findings

Our measure of arousal - Composite Arousal - was highest in the AV condition, and lowest in the VO condition. This suggests that the addition of the musical component to the visual clip made them more arousing. Music can markedly enhance the emotional experience evoked by emotional pictures, as indexed by galvanic skin response, respiratory physiology and heart rate (Baumgartner, Esslen, & Jäncke, 2006). In the context of musical performances as well, bimodal conditions give rise to higher physiological (electrodermal activity measurement) and behavioral responses when participants are presented with audiovisual stimuli of musical performances, as opposed to AO and VO conditions (Levitin & Chapados, 2008). On the other hand, the audiovisual condition has also been shown to have stress alleviation effects (Shoda et al., 2016). A study by Thayer and Levinson (1983) investigated the effect of the sound-track on electrodermal measurements in three conditions: the film without accompanying music, the film with horror music (high-arousal score), and the film accompanied by the documentary music (low-arousal score). The authors found that the purpose-composed music was successful in increasing (horror music) and decreasing (documentary music) electrodermal activity measurements. Future studies could use a categorical distinction between highly-arousing music and calming music to clarify the effect of music when it has an increasing vs. decreasing effect on the physiological response to multimodal content. This classification could also help probe the difference between stimulative or sedative aspects of film-music on time perception and could shed light on the underlying musical properties (such as tempo - see Droit-Volet et al., 2013) that are responsible for the effects observed in our study.

Valence was also a significant predictor in our results. Increased valence was associated with an increase in duration estimates; however, its effect was relatively small. In musical terms, valence may be understood as the range of affective responses from sad (low valence) to happy (high valence), that is to say, negative versus positive emotions. Alternatively, valence may be understood as the spectrum between ‘unpleasant’ and ‘pleasant’ music. Sad (negative) music may be perceived as pleasant (Kawakami et al., 2013; Sachs, Damasio, & Habibi, 2015), thus valence in a musical context may be better understood as describing pleasantness, rather than positiveness. We found no published studies that reported an effect of musical valence on time perception when valence is taken as the latter. On the other hand, Droit-Volet et al. (2013) found that manipulating the pleasantness of music, by creating tonal vs atonal (backwards) versions of the same pieces, had an impact on duration estimates - i.e. pleasant music was judged shorter than unpleasant music. Our results thus differ from their conclusion. The unpleasant (atonal) condition in their experiment was created by reversing the audio track which resulted in a strong contrast between pleasant and unpleasant stimuli. It is possible that the difference in findings could be explained by the fact that our clips didn't explore the full range of emotional experiences evoked in the context of film. A film genre that was missing entirely from our dataset is the horror genre. Aversive stimuli (negative valence) are known to cause time dilation more strongly than positive stimuli (Droit-Volet & Meck 2007), which may be important from an evolutionary point of view (Mella et al., 2011). Such an inclusion could have contributed in creating a wider spectrum of valence.

Finally, a post-hoc analysis revealed that the number of editing cuts predicted duration estimates. This visual property was found to increase duration estimates. With the caveats that this was a post-hoc analysis, we hypothesize that the visual discontinuities arising from cuts lead to longer duration estimates. Future studies could investigate the qualitative aspect of the cut, such as jump cuts (omission of time) or match cuts (cuts where shots are matched by the action or subject), and how these affect the experience of continuity and time. Inserting controlled editing manipulations that make use of human cognitive predispositions in future research could provide valuable insights into our perceptual mechanisms at the basis of time and continuity, core elements of cinematic experiences. We, therefore, urge future research to collaborate with composers and film-makers to create the desired experimental stimulus material.

Conclusion

In summary, we have provided evidence suggesting that the perceptual modality affects duration perception in the context of film such that, after controlling for other variables, VO clips yielded longer duration estimates compared to AO and AV clips. While the Independent Effects model showed a significant difference in duration estimates between the AO and AV conditions, this difference was no longer significant after controlling for the interaction between conditions and arousal. We provide two interpretations for the effects of sensory modality on time perception which rely on attention allocation and modal dominance in duration perception. We also discuss the potential role of music in this context, as well as its effect on physiological measurements. Moreover, arousal positively influenced duration estimates across conditions. The rate of this effect was stronger in the multimodal condition which warrants future research. We also show that time perception is influenced by other internal factors (e.g. experience of valence/pleasantness) and external factors (e.g. actual clip duration and editing manipulations). Despite a long history, time perception research is still a long way from understanding how seemingly effortless duration estimates are performed in a multimodal context but film extracts are an insightful and ecologically valid way of studying the influence of sensory modality, attention, emotions and other internal/external factors on time perception.

Acknowledgments

The authors would like to thank Amitai Shenhav for his help in conceiving the paradigm, Elena Festa for providing equipment, and Marc Perlman for his insightful feedback. An earlier version of this study was submitted by Mattis Appelqvist-Dalton to fulfill the requirements for an ScB degree with Honors from Brown University. This study was supported by the Independent Concentration Fund (Brown University).

Open Practices Statement:

The data and materials for all experiments are available on the Brown Data Repository (https://doi.org/10.26300/ke1d-f930) and the experiment was not preregistered.

References

Amadeo, M. B., Campus, C., & Gori, M. (2020). Visual representations of time elicit early responses in human temporal cortex. *NeuroImage*, *217*, 116912. https://doi.org/10.1016/j.neuroimage.2020.116912

Arstila, V. (2012). Time slows down during accidents. *Frontiers in Psychology*, *3*, 196. https://doi.org/10.3389/fpsyg.2012.00196

Aston-Jones, G., & Cohen, J.D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403-450. doi: 10.1146/annurev.neuro.28.061604.135709

Barakat, B., Seitz, A. R., & Shams, L. (2015). Visual rhythm perception improves through auditory but not visual training. *Current Biology*, *25*(2), R60-R61. https://doi.org/10.1016/j.cub.2014.12.011

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Baveye, Y., Dellandrea, E., Chamaret, C., & Chen, L. (2015). Liris-accede: A video database for affective content analysis. *IEEE Transactions on Affective Computing*, *6*(1), 43-55. DOI: 10.1109/TAFFC.2015.2396531

Baumgartner, T., Esslen, M., & Jäncke, L. (2006). From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *International Journal of Psychophysiology*, *60*(1), 34-43. doi: 10.1016/j.ijpsycho.2005.04.007

Behar, I., & Bevan, W. (1961). The perceived duration of auditory and visual intervals:

Cross-modal comparison and interaction. *The American Journal of Psychology*, *74*(1), 17-26.

Beeman, W. O. (1981). The use of music in popular film: East and West. *India International Centre Quarterly*, *8*(1), 77-87.

Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review*, *4*(2), 184-197. doi: 10.3758/BF03209393

Block, R. A., Hancock, P. A., & Zakay, D. (2010) How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica*, *134*(3), 330-343. https://doi.org/10.1016/j.actpsy.2010.03.006

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition,* 1(1). doi: 10.5334/joc.10

Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, *59*(7), 1118-1140.

Burr, D., Banks, M.S., & Morrone, M.C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*, *198*(1), 49–57. doi: 10.1007/s00221-009-1933-z

Cai, Z. G., & Wang, R. (2014). Numerical magnitude affects temporal memories but not time encoding. *PLoS ONE*, *9*(1), e83159.https://doi.org/10.1371/journal.pone.0083159

Chapados, C., & Levitin, D. J. (2008). Cross-modal interactions in the experience of musical performances: Physiological correlates. Cognition, 108(3), 639-651. doi: 10.1016/j.cognition.2008.05.008

Cheng, R. K., Tipples, J., Narayanan, N. S., & Meck, W. H. (2016). Clock speed as a window into dopaminergic control of emotion and time perception. *Timing & Time Perception*, *4*(1), 99-122. https://doi.org/10.1163/22134468-00002064

Cohen, A. J. (2011). Music as a source of emotion in film. In Patrik N. Juslin & John Sloboda (eds.), *Handbook of Music and Emotion: Theory, Research, Applications*. (pp. 1099 - 1036) Oxford University Press. doi:10.1093/acprof:oso/9780199230143.003.0031

Coutrot, A., Guyader, N., Ionescu, G., & Caplier, A. (2012). Influence of soundtrack on eye movements during video exploration. *Journal of Eye Movement Research*, *5*(4), 2. doi: 10.16910/jemr.5.4.2

Craig, A. D. (2009). Emotional moments across time: a possible neural basis for time perception in the anterior insula. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1933-1942. doi: 10.1098/rstb.2009.0008

Droit-Volet, S., Meck, W. H., & Penney, T. B. (2007). Sensory modality and time perception in children and adults. *Behavioural Processes*, *74*(2), 244-250. doi: 10.1016/j.beproc.2006.09.012

Droit-Volet, S., Fayolle, S. L., & Gil, S. (2011). Emotional state and time perception: mood elicited by films. *Frontiers in Integrative Neuroscience*, *5*, 33. doi:10.3389/fnint.2011.00033.

Droit-Volet, S., Bueno, L. J., & Bigand, E. (2013). Music, emotion, and time perception: the influence of subjective emotional valence and arousal? *Frontiers in Psychology*, *4*, 417. doi: 10.3389/fpsyg.2013.00417.

Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in*

Neurobiology, *18*(2), 131-136. doi: 10.1016/j.conb.2008.06.002

Eldar, E., Ganor, O., Admon, R., Bleich, A., & Hendler, T. (2007). Feeling the real world: limbic response to music depends on related content. *Cerebral Cortex*, *17*(12), 2828-2840.

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

Fraisse, P. (1957). *Psychologie du temps*. Presses Universitaires de France.

Fontes, R., Ribeiro, J., Gupta, D. S., Machado, D., Lopes-Júnior, F., Magalhães, F., ... & Velasques, B. (2016). Time perception mechanisms at central nervous system. *Neurology International*, *8*, 5939. doi:10.4081/ni.2016.5939

Ford, B. Q., Tamir, M., Brunyé, T. T., Shirer, W. R., Mahoney, C. R., & Taylor, H. A. (2010). Keeping your eyes on the prize: Anger and visual attention to threats and rewards. *Psychological Science*, *21*(8), 1098-1105. doi: 10.1177/0956797610375450

Gil, S., & Droit-Volet, S. (2012). Emotional time distortions: the fundamental role of arousal. *Cognition and Emotion*, *26*(5), 847-862. DOI: 10.1080/02699931.2011.625401

Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493-498. https://doi.org/10.1111/2041-210X.12504

He, M., Heindel, W. C., Nassar, M. R., Siefert, E. M., & Festa, E. K. (2020). Age-related changes in the functional integrity of the phasic alerting system: a pupillometric investigation. *Neurobiology of Aging*, *91*, 136-147. https://doi.org/10.1016/j.neurobiolaging.2020.02.025

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459-491. doi: 10.1037/0033-295x.96.3.459.

Kawakami, A., Furukawa, K., Katahira, K., & Okanoya, K. (2013). Sad music induces pleasant emotion. *Frontiers in Psychology*, *4*, 311. https://doi.org/10.3389/fpsyg.2013.00311

Klink, P. C., Montijn, J. S., & van Wezel, R. J. (2011). Crossmodal duration perception involves perceptual grouping, temporal ventriloquism, and variable internal clock rates. *Attention, Perception, & Psychophysics*, *73*(1), 219-236. doi: 10.3758/s13414-010-0010-9.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmer test package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. Doi: 10.18637/jss.v082.i13

Lake, J. I. (2016). Recent advances in understanding emotion-driven temporal distortions. *Current Opinion in Behavioral Sciences*, *8*, 214-219. doi: 10.1016/j.cobeha.2016.02.009

Lipscomb, S. D., & Kendall, R. A. (1994). Perceptual judgement of the relationship between musical and visual components in film. *Psychomusicology: A Journal of Research in Music Cognition*, *13*(1-2), 60-98. https://psycnet.apa.org/doi/10.1037/h0094101

Lissa, Z. (1965). *Ästhetik der Filmmusik* (p. 140). Berlin: Henschelverlag.

Lustig, C., & Meck, W. H. (2011). Modality differences in timing and temporal memory throughout the lifespan. *Brain and Cognition*, *77*(2), 298-303. doi: 10.1016/j.bandc.2011.07.007.

Marks, M. M. (1997). Music and the silent film: Contexts and case studies, 1895-1924. Oxford University Press on Demand.

Matthews, W. J. (2015). Time perception: the surprising effects of surprising stimuli. *Journal of Experimental Psychology: General*, *144*(1), 172-197. doi: 10.1037/xge0000041.

Mella, N., Conty, L., & Pouthas, V. (2011). The role of physiological arousal in time perception: Psychophysiological evidence from an emotion regulation paradigm. *Brain and Cognition*, *75*(2), 182-187. http://dx.doi.org/10.1016/j.bandc.2010.11.012

Mera, M., & Stumpf, S. (2014). Eye-tracking film music. *Music and the Moving Image*, *7*(3), 3-23. https://doi.org/10.5406/musimoviimag.7.3.0003

Merchant, H., Harrington, D. L., & Meck, W. H. (2013). Neural basis of the perception and estimation of time. *Annual Review of Neuroscience*, *36*, 313-336. DOI: 10.1146/annurev-neuro-062012-170349

Meumann, E. (1893). Beitrage zur psychologie des zeitsinns. *Philosophische Studien*, *9*, 264-306.

Muth, C., Bales, K. L., Hinde, K., Maninger, N., Mendoza, S. P., & Ferrer, E. (2016). Alternative models for small samples in psychological research: applying linear mixed effects models and generalized estimating equations to repeated measures data. *Educational and Psychological Measurement*, *76*(1), 64-87. https://doi.org/10.1177/0013164415580432

Nichols, T.E., & Holmes, A.P., 2001. Nonparametric permutation tests for functional neuroimaging experiments: a primer with examples. *Human Brain Mapping, 15*, 1e25. https://doi.org/10.1002/hbm.1058

Noulhiane, M., Mella, N., Samson, S., Ragot, R., & Pouthas, V. (2007). How emotional auditory stimuli modulate time perception. *Emotion*, *7*(4), 697-704. DOI: 10.1037/1528-3542.7.4.697

Oakes, S. (2003). Musical tempo and waiting perceptions. *Psychology & Marketing, 20*(8), 685-705. https://doi.org/10.1002/mar.10092

Ogden, R. S., Henderson, J., McGlone, F., & Richter, M. (2019). Time distortion under threat: Sympathetic arousal predicts time distortion only in the context of negative, highly arousing stimuli. *PloS ONE*, *14*(5), e0216704. https://doi.org/10.1371/journal.pone.0216704

Ortega, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2014). Audition dominates vision in duration perception irrespective of salience, attention, and temporal discriminability. *Attention, Perception, & Psychophysics*, *76*(5), 1485-1502. doi: 10.3758/s13414-014-0663-x.

Penney, T. B., Gibbon, J., & Meck, W. H. (2000). Differential effects of auditory and visual signals on clock speed and temporal memory. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1770–1787. doi:10.1037/0096-1523.26.6.1770

Peysakhovich, V., Vachon, F., & Dehais, F. (2017). The impact of luminance on tonic and phasic pupillary responses to sustained cognitive load. *International Journal of Psychophysiology*, *112*, 40-45. doi: 10.1016/j.ijpsycho.2016.12.003.

Pfeuty, M., Dilharreguy, B., Gerlier, L., & Allard, M. (2015). fMRI identifies the right inferior frontal cortex as the brain region where time interval processing is altered by negative emotional arousal. *Human Brain Mapping*, *36*(3), 981-995. doi: 10.1002/hbm.22680

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

Sanders, T., & Cairns, P. (2010). Time perception, immersion and music in videogames. *Proceedings of the 24th BCS Interaction Specialist Group Conference,* 160-167. https://dl.acm.org/doi/pdf/10.5555/2146303.2146327

Sachs, M. E., Damasio, A., & Habibi, A. (2015). The pleasures of sad music: a systematic review. *Frontiers in Human Neuroscience*, *9*, 404. https://doi.org/10.3389/fnhum.2015.00404

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*(6), 110-114. doi: 10.2307/3002019.

Schäfer, T., & Fachner, J. (2015). Listening to music reduces eye movements. *Attention, Perception, & Psychophysics*, *77*(2), 551-559. DOI: 10.3758/s13414-014-0777-1

Smeets, J. B. J., & Hooge, I. T. C. (2003). Nature of variability in saccades. *Journal of Neurophysiology*, *90*(1), 12–20. https://doi.org/10.1152/jn.01075.2002

Tarkovsky, A. (1988). *Sculpting in time* (K. Hunter-Blair, Trans.). University of Texas Press.

Thayer, J. F., & Levenson, R. W. (1983). Effects of music on psychophysiological responses to a stressful film. *Psychomusicology: A Journal of Research in Music Cognition*, *3*(1), 44-52. https://psycnet.apa.org/doi/10.1037/h0094256

Tse, P. U., Intriligator, J., Rivest, J., & Cavanagh, P. (2004). Attention and the subjective expansion of time. *Perception and Psychophysics*, *66*, 1171-89. doi: 10.3758/bf03196844

van Wassenhove, V., Buonomano, D. V., Shimojo, S., & Shams, L. (2008). Distortions of subjective time perception within and across senses. *PLoS ONE*, *3*(1), e1437. doi: 10.1371/journal.pone.0001437

Venables, W., Smith, D., & the R Core Team (2013) An introduction to R, Vol. 34. Notes on R:

A programming environment for data analysis and graphics. Version 4.0.2 (2020–06–22).

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, *101*(1), 80-113. doi: 10.1016/j.cognition.2005.09.003

Wearden, J. H., Edwards, H., Fakhri, M., & Percival, A. (1998). Why "sounds are judged longer than lights": Application of a model of the internal clock in humans. *The Quarterly Journal of Experimental Psychology*, *51B*, 97–120. doi: 10.1080/713932672

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638. https://psycnet.apa.org/doi/10.1037/0033-2909.88.3.638

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Switzerland: Springer.

Wöllner, C., Hammerschmidt, D., & Albrecht, H. (2018). Slow motion in films and video clips: Music influences perceived duration and emotion, autonomic physiological activation and pupillary responses. *PloS One*, *13*(6), e0199161. doi: 10.1371/journal.pone.0199161

Zakay, D. (2005). Attention et jugement temporel. *Psychologie Française*, *50*(1), 65-79. https://doi.org/10.1016/j.psfr.2004.10.004

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends in Hearing*, *22*, 2331216518777174. doi: 10.1177/2331216518777174

Figure 1

Experimental Design

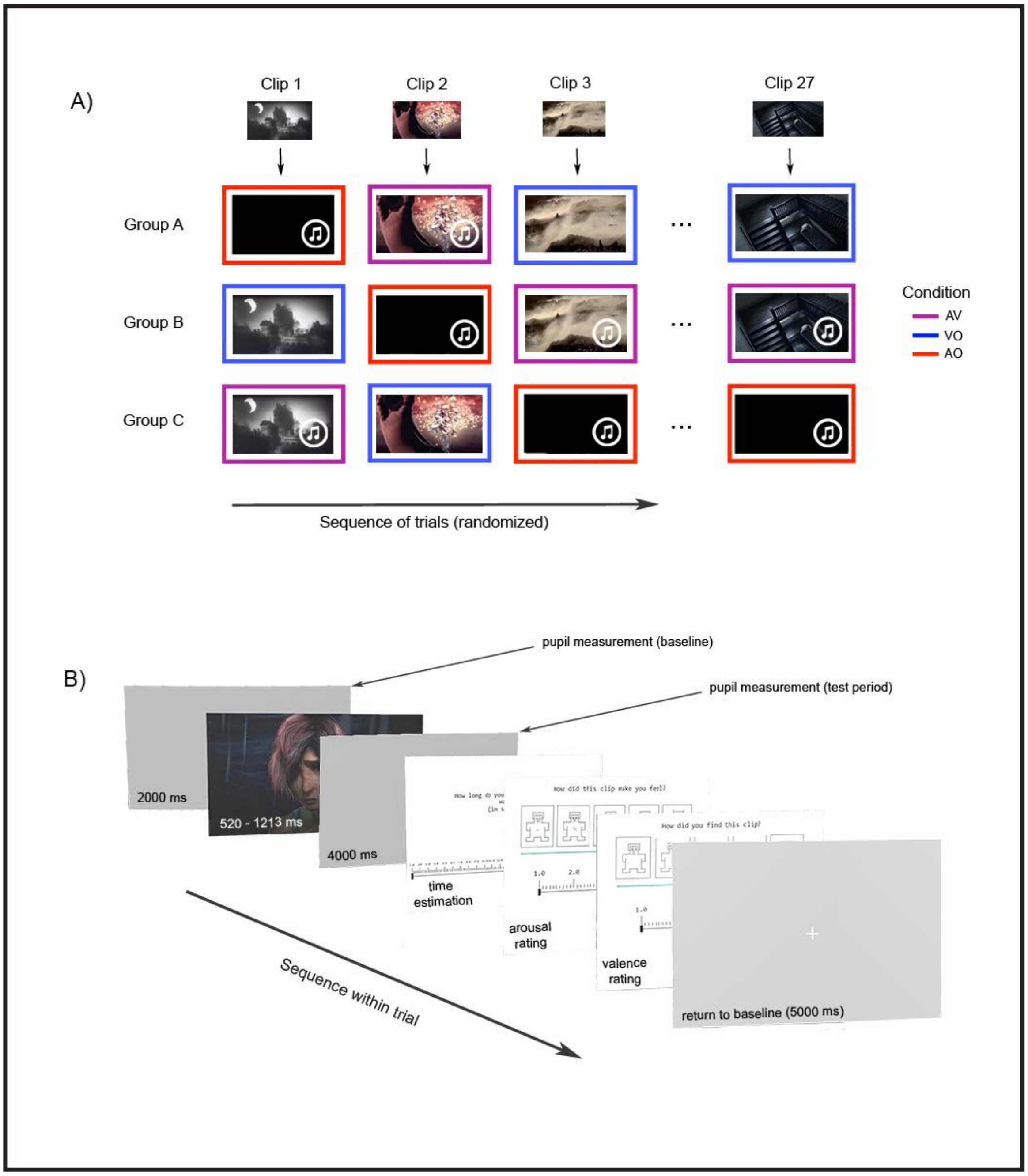


Fig. 1 A) Sequence of experimental trials, stimuli, and group division. Each group of participants (A, B, C, left column) viewed all 27 clips (top row, clips 1, 2, 3, and 27 are shown) and each clip was shown in only one group-specific condition (9 clips in each condition). Each participant saw the clips in a different (randomized) order. The color boxes identify condition (right side labels). **B)** Sequence within an individual trial. The 1st panel is a grey screen (with a fixation cross) displayed for 2000 ms (baseline period), the 2nd panel is the film clip (between 520 and 1213), followed by another grey fixation screen for 4000 ms (3rd panel, pupil measurement interval). The 4th, 5th and 6th panels correspond to the judgement tasks (time estimation, arousal rating and valence rating) and the 7th panel corresponds to 5000 ms of silence and a grey viewing screen (return to pupil size baseline).

Figure 2

Temporal Analysis of Pupil Size Effect and Modality-driven Differences in Composite Arousal

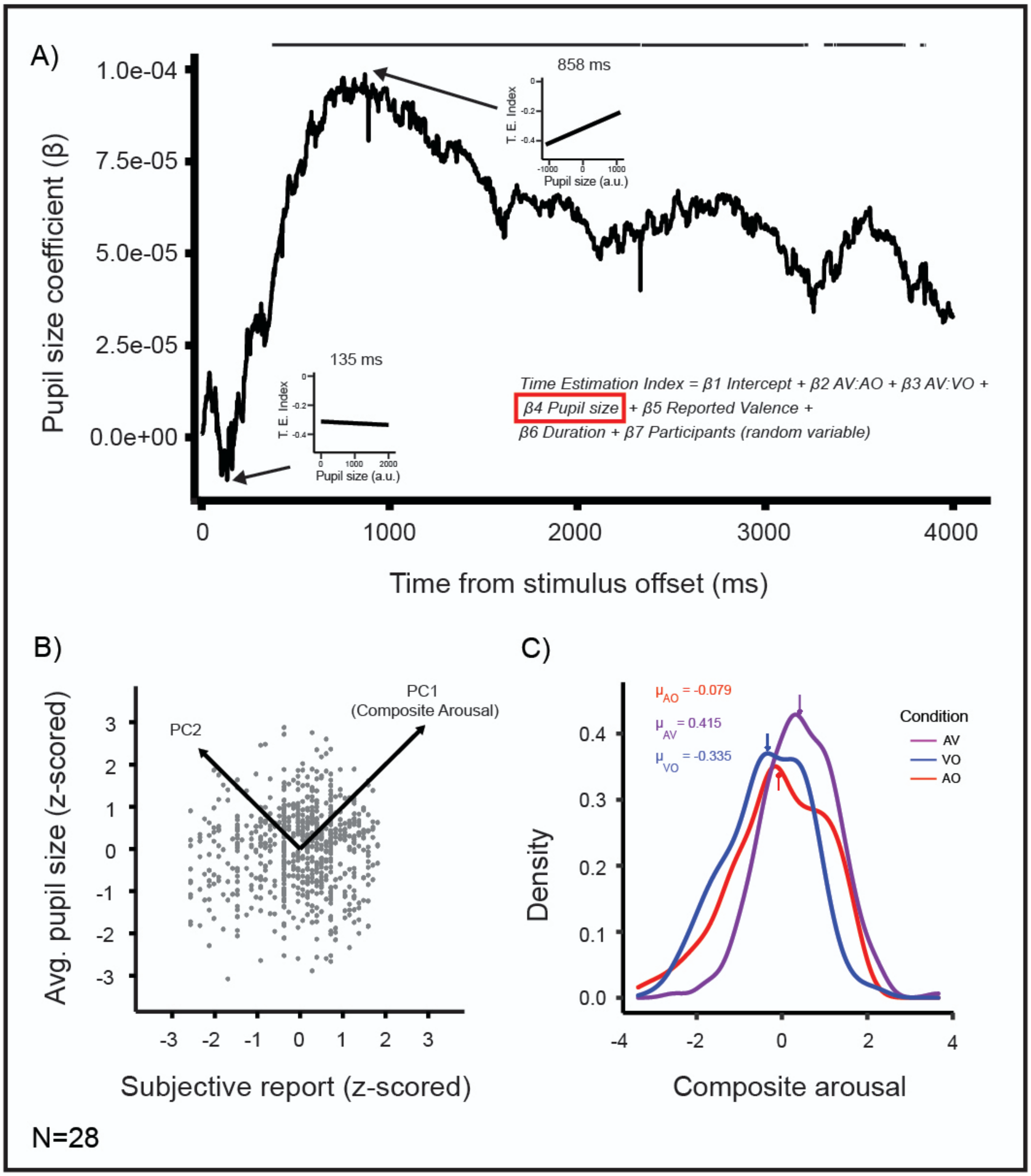


Fig. 2 A) Pupil size coefficient (**β4;** extracted from the time estimation linear mixed model displayed in the graph) as a function of time in the 4000 ms time interval following clip offset (x-axis). The statistical significance at each time point in the test period (p-value < 0.05) is depicted by the horizontal black lines above the graph. Ten clusters of significance ranging in size from 2 ms to 1959 ms are displayed (3297 out of 4000 ms, or 82.4% of the entire interval). The two inset graphs show the slope of the parameter value at two distinct time points (135 ms and 858 ms), showing the effect (slope depicted as a black line) of pupil size on Time Estimation Index when the pupil size value used in the model is restricted to those unique time points. The axes of these inset graphs correspond to the Time Estimation Index (y-axis) and pupil size (x-axis) in arbitrary units (a.u.). **B)** Average pupil size (z-scored) as a function of subjectively reported arousal (z-scored). Each dot corresponds to one trial. Both measures were subjected to a PCA that yielded two principal components; the corresponding dimensions are depicted with arrows. Full explanation of analysis is included in the Methods section. **C)** The density distribution of Composite Arousal (derived from Principal Component Analysis) for each Condition. The value for AV (purple line) is positive while those for AO (red line) and VO (blue line) are negative.

Figure 3

Arousal and Condition-driven Differences in Time Estimation Index

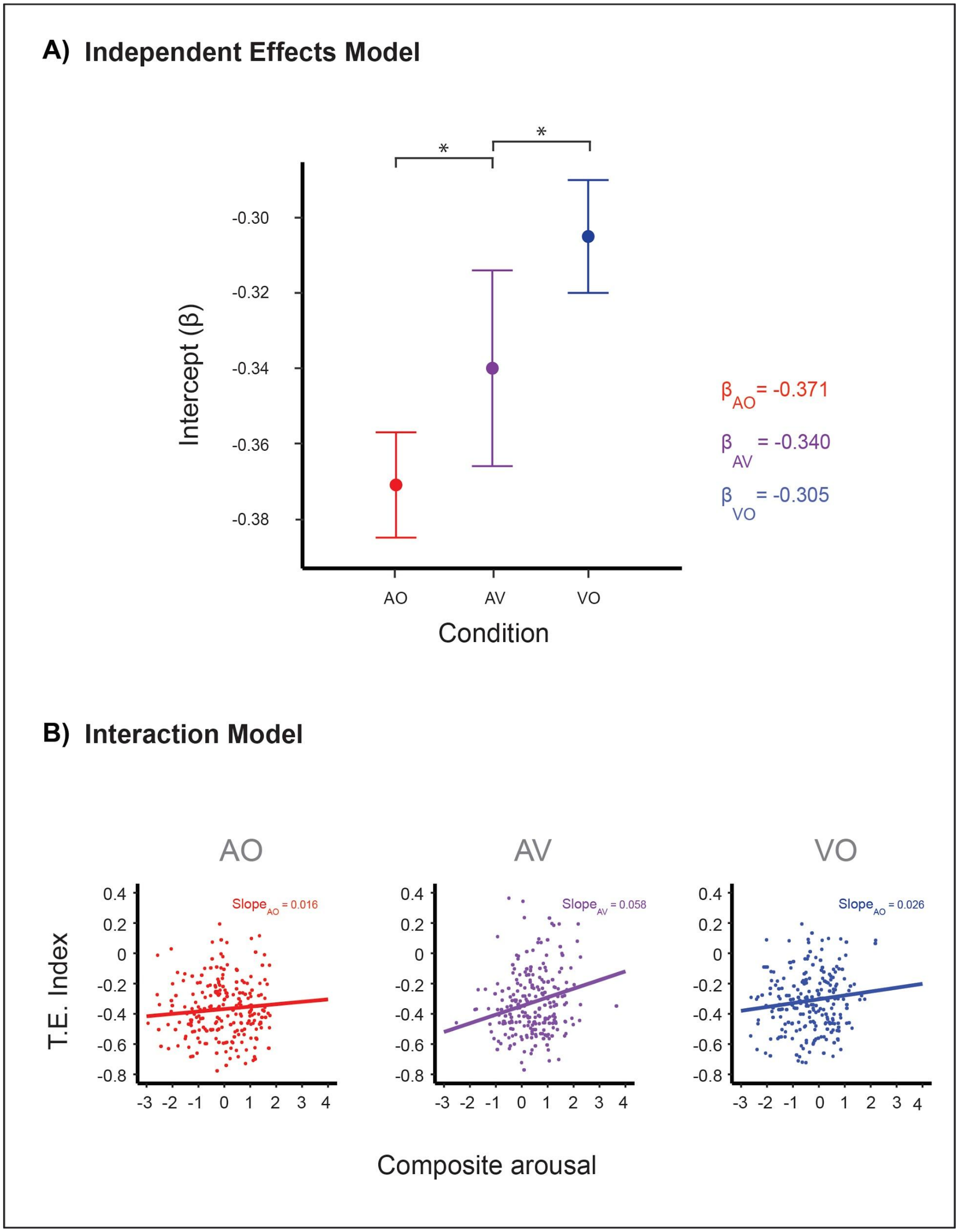


Fig. 3 A) The Independent Effects Model-derived estimates of the average Time Estimation Index (Intercept (β), y axis) for each condition. Data are means plus one standard deviation. The mean intercept estimates are provided to the right. Note that the significant difference between AO and AV displayed in this graph is not significant in the Interaction Model (see **Table 1**). **B)** The Interaction Model-derived estimates of the effect of Composite Arousal on Time Estimation Index differs between conditions. Slope values are provided in each individual plot. The steepest slope is seen in the AV condition.

Table 1

A)

Independent Effects Model

	β	df	t
AV:AO	-0.031^{*} (0.014)	666.754	-2.133
AV:VO	0.035^{*} (0.015)	666.756	2.304
Composite Arousal	0.030^{***} (0.006)	676.139	4.720
PC2	-0.00004 (0.007)	682.115	-0.005
Duration	-0.019^{***} (0.004)	666.021	-4.978
Valence	0.013^{*} (0.006)	671.037	2.043
Intercept	-0.340^{***} (0.026)	33.744	-13.170
Observations			700
Log Likelihood			265.654
Akaike Inf. Crit.			-513.351
Bayesian Inf. Crit.			-472.391

B)

Interaction Model

	β	df	t
AV:AO	-0.020 (0.015)	665.156	-1.366
AV:VO	0.045^{**} (0.015)	664.874	2.934
Composite Arousal (AV)	0.058^{***} (0.012)	669.913	4.845
PC2	0.003 (0.007)	680.251	0.388
Duration	-0.019^{***} (0.004)	664.023	-4.957
Valence	0.013^{*} (0.006)	669.014	2.033
AV:AO x Arousal	-0.042^{**} (0.015)	669.178	-2.808
AV:VO x Arousal	-0.032^{*} (0.016)	667.958	-2.076
Intercept	-0.352^{***} (0.026)	35.534	-13.448
Observations			700
Log Likelihood			262.979
Akaike Inf. Crit.			-503.957
Bayesian Inf. Crit.			-453.895

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Results from the two mixed-effects models that differed on whether we included two parameters that estimated the interaction between condition and Composite Arousal: **A)** Independent Effects Model; **B)** Interaction Model.

Figure S1

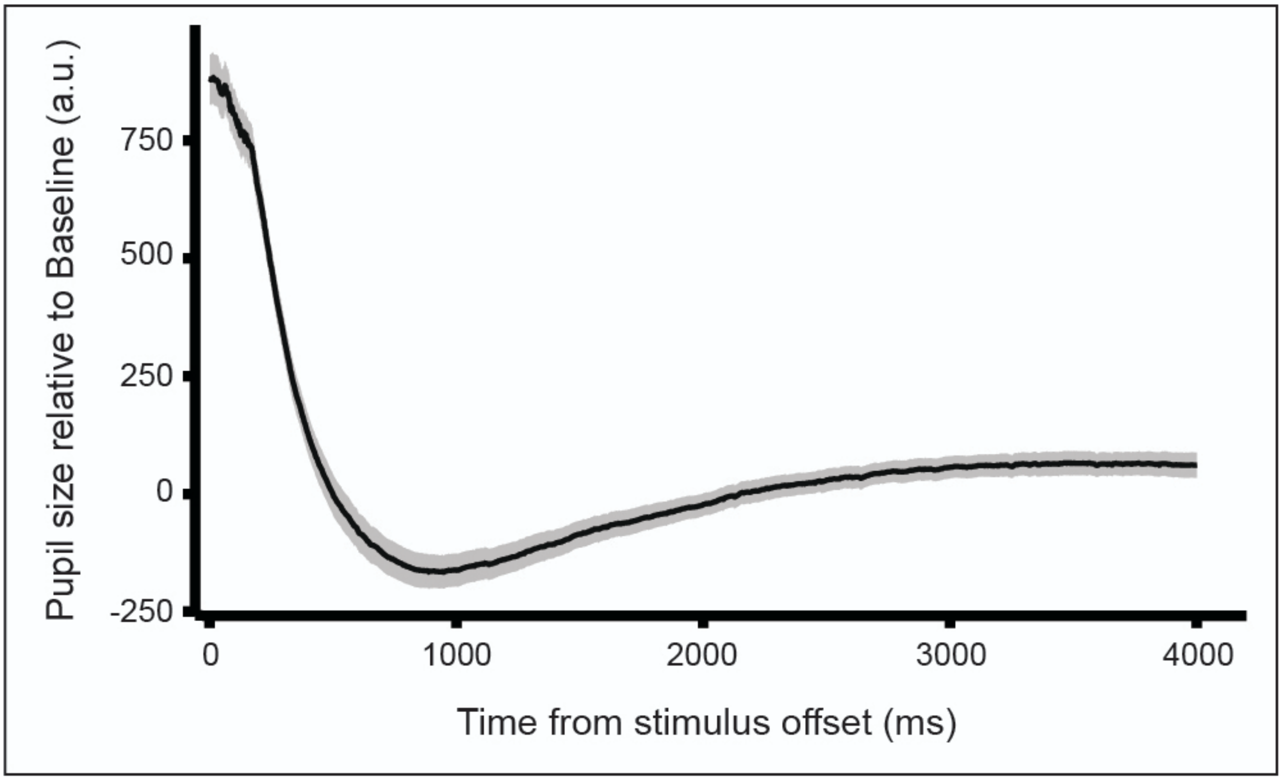


Fig. S1 Average pupil size in arbitrary units (a.u.) at each time point in the 4000 ms time interval after each clip. The gray envelope shows the 95% confidence interval.