Acta Psychologica 244 (2024) 104206

Contents lists available at ScienceDirect

Acta Psychologica

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

Time perception in film viewing: A modulation of scene's duration estimates as a function of film editing

Lydia Liapi [a], Elpida Manoudi [b], Maria Revelou [a], Katerina Christodoulou [a], Petros Koutras [c], Petros Maragos [c], Argiro Vatakis [a,*]

[a] Multisensory and Temporal Processing Laboratory (MultiTimeLab), Department of Psychology, Panteion University of Social and Political Sciences, Athens, Greece
[b] Department of Philosophy and History of Science, University of Athens, Athens, Greece
[c] Intelligent Robotic and Automation Laboratory, National Technical University of Athens, Athens, Greece

ARTICLE INFO

Keywords:
Filmmaking
Film editing
Time perception
Time estimation

ABSTRACT

Filmmakers and editors have empirically developed techniques to ensure the spatiotemporal continuity of a film's narration. In terms of time, editing techniques (e.g., elliptical, overlapping, or cut minimization) allow for the manipulation of the perceived duration of events as they unfold on screen. More specifically, a scene can be edited to be time compressed, expanded, or real-time in terms of its perceived duration. Despite the consistent application of these techniques in filmmaking, their perceptual outcomes have not been experimentally validated. Given that viewing a film is experienced as a precise simulation of the physical world, the use of cinematic material to examine aspects of time perception allows for experimentation with high ecological validity, while filmmakers gain more insight on how empirically developed techniques influence viewers' time percept. Here, we investigated how such time manipulation techniques of an action affect a scene's perceived duration. Specifically, we presented videos depicting different actions (e.g., a woman talking on the phone), edited according to the techniques applied for temporal manipulation and asked participants to make verbal estimations of the presented scenes' perceived durations. Analysis of data revealed that the duration of expanded scenes was significantly overestimated as compared to that of compressed and real-time scenes, as was the duration of real-time scenes as compared to that of compressed scenes. Therefore, our results validate the empirical techniques applied for the modulation of a scene's perceived duration. We also found interactions on time estimates of scene type and editing technique as a function of the characteristics and the action of the scene presented. Thus, these findings add to the discussion that the content and characteristics of a scene, along with the editing technique applied, can also modulate perceived duration. Our findings are discussed by considering current timing frameworks, as well as attentional saliency algorithms measuring the visual saliency of the presented stimuli.

1. Historical evolution of editing in film

Since the release of the first film around the end of the 19th century, cinematography has evolved and is currently considered one of the most significant forms of art and entertainment (Shimamura, 2013). This has partially been accomplished by successfully engaging the viewer in a virtual world that simulates our physical world (Brunick et al., 2013) in terms of the spatial and temporal continuity of events (e.g., Berliner & Cohen, 2011; Bordwell & Thompson, 2012), the unfolding of a dynamic optical flow (e.g., Gibson, 1979; Tan, 2018), or the segmentation of discrete events (i.e., viewers tendency to perceive boundaries when changes in the perceptual or conceptual features of the observed event occur; e.g., Kurby & Zacks, 2008; Magliano et al., 2001; Magliano & Zacks, 2011). Such simulations are possible due to the development of editing techniques, such as continuity editing, ensuring the smooth transition between shots by minimizing cut detection (i.e., the transition point between consecutive shots), thus inducing a sense of spatiotemporal continuity of a film's narrative to the viewer (Bordwell, 1985; Bordwell & Thompson, 2012; Cutting, 2005; Cutting and Candan, 2013; Shimamura, 2013; Smith, 2012). These techniques have mainly been developed through trial and error (Berliner & Cohen, 2011; Shimamura, 2013) and their continuous use led to their eventual establishment as means to guarantee that the viewer's sense of continuity is not disrupted.

* Corresponding author at: Department of Psychology, Panteion University of Social and Political Sciences, Leof. Syggrou Avenue 136, 176 71 Athens, Greece.
E-mail address: argiro.vatakis@gmail.com (A. Vatakis).

https://doi.org/10.1016/j.actpsy.2024.104206
Received 3 August 2023; Received in revised form 6 February 2024; Accepted 5 March 2024
Available online 10 March 2024

1.1. Editing techniques that affect a scene's perceived duration

The use of editing techniques has also been applied empirically in the filmmaker's attempts to manipulate timing, a particularly challenging task as a movie's physical duration (i.e., screen time or the time it takes to watch the movie) is not identical to its narrative duration. To deal with such time constraints, filmmakers utilize techniques that attempt to manipulate the perceived duration of a scene. Thus, elliptical editing (i.e., parts of an action are omitted but implied and shots are carefully linked together for action continuity) has been used for the compression of an action's duration (Media College, n.d.; Bordwell & Thompson, 2012; Mariani, 2018; Sieghartsleitner, 2018), while for the opposite effect (i.e., expansion of an action's duration), overlapping editing (i.e., parts of an action are repeatedly presented through consecutive shots, possibly from different angles) has been utilized (Media College, n.d.; Mariani, 2018; Reisz & Millar, 2010; Sieghartsleitner, 2018). Note that, depending on the medium, other editing techniques may be utilized. For instance, in music video clips or documentaries, a scene's duration may be compressed through the use of: a) time lapse (or fast motion), where time is compressed simply by speeding up the presented shot, b) transitions such as the fading out of one shot and into the next shot, or c) time remapping, where the speed of a segment varies between different shots (Media College, n.d.; Lindblom, 2015). Accordingly, expansion of a scene's duration can be accomplished by using slow motion, where timing is expanded by slowing down the speed of the presented action or through time remapping (Media College, n.d.; Lindblom, 2015). Cut minimization or long-takes, consisting of a single shot, serve to create a real-time experience to the viewer and allow the depiction of an action's duration as it would last in the physical world (Media College, n.d.; Ghosh, 2022). Proper use of such techniques can lead to a fast- or slow-paced movie segment that adds on to the narrative of the film and the goals of the filmmaker to induce a particular viewer state. However, even though these techniques have been systematically utilized in filmmaking, no study to date -according to our knowledge- has examined their perceptual validity, meaning it is still unknown whether this expansion or compression of a scene's timing is actually what the participants perceive or experience.

The use of scenes, movie segments or entire films to investigate aspects of time perception (or other cognitive functions) is valuable as such stimulation offers high levels of naturalism and complexity, thus facilitating the conduct of ecologically valid experiments with findings that can be easily generalized in real world settings. Moreover, the use of established editing techniques allows researchers to conduct easy to reproduce experiments (given that sufficient documentation of how the stimulation was recorded and edited is provided), thus, enhancing reliability of their findings. At the same time, the nature of cinematic stimuli allows the investigation of other low-level factors that affect duration perception, such as motion, shot length, or stimulus speed, without restricting experimenters to the use of simple stimulation, rarely met outside the settings of a laboratory. From a filmmaker's perspective, studies in this field can offer valuable insight on how various aspects of filmmaking affect viewers' sense of time and promote the creation of novel, experimental editing techniques.

1.2. Models of time perception

As we aim to investigate the potential effects of editing on the perceived duration of a scene, it is important to consider the potential predictions of the psychological models of time perception (it should be noted that we are focusing on timing at the level of a short scene and not a whole movie, large movie segments, or movie narrative, as this would be beyond the scope of the present study). Information processing models of timing support the existence of an internal clock that is based on a three-stage process: an accumulator that collects pulses emitted at a constant rate (i.e., pacemaker), a working memory component that encodes the specific time interval presented, and a decision-making component, where the ongoing pulses are compared to the reference memory for response selection (e.g., Gibbon et al., 1984; Treisman, 1963).

1.3. Attentional gate model of time perception

If the editing techniques for expansion, compression or real-time presentation of an action do in fact influence perceived time, we aim to examine whether this influence would be in line with the predictions of the Attentional Gate Model (AGM). The AGM posits that timing is affected by attentional allocation with resources devoted to temporal processing having an inverse relationship to the attentional demands of non-temporal processing (e.g., Block & Zakay, 1997; Zakay & Block, 1997). Thus, timing judgements for the editing techniques utilized here (i.e., elliptical editing, overlapping editing, and minimization of cuts to create perceptually compressed, expanded, and real-time scenes, respectively) can be defined by the main differentiating factor of these techniques. Given that cuts within a scene are not cues of temporal nature, based on AGM we assume that a higher number of cuts would result in a higher level of processing of non-temporal information. At the same time for scenes with fewer cuts, the attentional resources previously allocated to the processing of cuts (i.e., a non-temporal element) would now be available for the processing of a scene's temporal elements. Therefore, by placing the number of cuts as the key differentiating factor between the three editing techniques, we would expect that in the case of time expansion, viewers will have to process more non-temporal information as compared to the compressed and real-time scenes, thus resulting in fewer attentional resources devoted to the temporal processing of the attended interval. This, in turn, will result to a lower pulse accumulation, expressed by an underestimation for the expanded as compared to the compressed and real-time scenes, a prediction opposite to the filmmakers' intentions. The model's predictions also suggest an overestimation of real-time as compared to compressed scenes, given that fewer cuts are included in the former scenes and, thus, more attentional resources would be available for temporal information processing.

1.4. Change model of time perception

Stimulus speed has also been reported as a factor modulating an interval's perceived duration (e.g., Kaneko & Murakami, 2009; Makin et al., 2012; Sgouramani & Vatakis, 2014). Following the internal clock framework, stimulus speed is directly related to the accumulation of pulses by the pacemaker. More specifically, the presentation of a fast-moving stimulus leads to an acceleration of the pacemaker's firing rate, resulting in turn to a higher pulse accumulation and, therefore, an overestimation of the attended interval (e.g., Kaneko & Murakami, 2009; Makin et al., 2012; Matthews, 2011; Sgouramani & Vatakis, 2014). An alternative to this account, known as the change model, places change as the index of time passage, and thus, the number of changes as a determining factor of an interval's perceived duration, with a higher number of changes resulting in a perceptual expansion of the attended interval's duration (Poynter & Homa, 1983). By taking into account the change model in regard to the effect of stimulus speed in time perception, fast moving stimuli are assumed to contain a higher number of changes, which, in turn, leads to a perceived lengthening of the elapsed time as compared to slow moving stimuli (e.g., Brown, 1995; but see Sgouramani & Vatakis, 2014, for an examination of the effect of stimulus speed to time perception, while controlling for stimulus changes).

In terms of the editing techniques discussed in the present paper, according to the change model, judging an interval's duration is based on the ability to remember the sequence of events of a presented action and the ability to infer the duration of these sub-events (Block, 1982; Block & Reed, 1978; Fraisse, 1963; Poynter, 1983; Poynter, 1989; Poynter & Homa, 1983). The approach of the change model for prospective temporal estimation provides an alternative account of the potential effects of editing on the perceived duration of a scene. As stated by Poynter (1989), this model also considers the degree of contextual change between the sub-events that comprise an action (with a higher degree of change leading to a more evident effect of the number of changes on duration estimation), as well as the organization of these sub-events (with easily remembered sequences having a more evident effect on duration estimation). This model, therefore, predicts that editing techniques that require a higher number of cuts (i.e., changes) will result in the overestimation of the given scene's duration. Such prediction will point to an overestimation of the expanded scenes (with the highest number of cuts) as compared to the real-time (with the lowest number of cuts) and compressed scenes, but at the same time, the compressed scenes will be overestimated as compared to the real-time scenes, which is opposite to the expected result. Thus, it seems that based on the current models of timing, it is not easy to fully account for the expected outcome when editing a film.

One point, however, that we have not yet considered is the failure to actually detect a scene's edit. Studies have supported that participants cannot detect cuts in scenes that follow the continuity editing rules (even when they were explicitly asked to do just that) and further research has verified that adherence to these rules makes the viewers unaware of any cuts present in a scene (i.e., a phenomenon known as 'edit blindness'; e.g., d'Ydewalle & Vanderbeeken, 1990; Smith & Henderson, 2008). Such findings argue that scene editing is not expected to disrupt attention given that continuity editing rules are followed and that changes from one shot to the next will go undetected. Thus, there will be no need to allocate our attentional resources to non-temporal tasks and the number of changes present will not affect our timing judgements. Finally, it is worth noting that even when continuity editing rules are violated, viewers' attention is not disrupted as long as narrative continuity is preserved (e.g., Germeys & d'Ydewalle, 2007).

1.5. The present study

Here, we examined whether the compressed, expanded, and real-time editing techniques used in filmmaking led to differences in viewers' duration judgements and if so, whether these differences would reflect a compression, expansion, or sustention of a scene's perceived duration, respectively. Five short action events were recorded and edited following the continuity editing rules and the corresponding techniques for temporal manipulation (i.e., compressed-, expanded-, and real-time presentation). More specifically, the compressed versions of each action were created utilizing elliptical editing, where parts of the action were not depicted but implied, the expanded versions of each action were created through overlapping editing, where consecutive shots from different angles depicted repetitive parts of the action, and the real-time versions of each action, where cuts were minimized, with a single cut in the depicted action. Participants viewed these shots and were asked to verbally estimate each scene's duration (Penton-Voak et al., 1996). We expected that adherence to the continuity rules and edit blindness would eliminate the potential effects of attention disruption. On the basis that elliptical editing is used to compress an action (Bordwell & Thompson, 2012; Sieghartsleitner, 2018), real-time editing is used to present an action as it lasts in the physical world (Ghosh, 2022), and overlapping editing is used to expand an action's duration (Bordwell & Thompson, 2012). We aimed to examine whether expanded-time scenes would be overestimated as compared to compressed-time scenes, and whether real-time scenes would be overestimated as compared to compressed-time scenes. At the same time, we explored whether event type (i.e., different actions depicted in the five scenes used; see Methods) would have a differential effect on participants' duration estimates. We also utilized computational attentional models for visual saliency estimation of the stimuli we utilized (see below). More specifically, these signal analysis algorithms allowed the computation of a temporal curve indicating the visual saliency of the presented, dynamic stimuli as they unfolded. This, in turn, allowed for the detection of any potential attentional shifts due to the cuts, thus, addressing potential attentional resource expenditure to nontemporal information.

1.6. Methods

1.6.1. Participants

67 participants (48 female), aged between 18 and 39 years (Mean age = 27.03 years) took part in the experiment, as duration estimation at an interval level remains stable throughout this age range within adulthood (e.g., Beck, 1988; Espinosa-Fernández et al., 2003; Hancock & Rausch, 2010). All of them were naïve as to the purpose of the study and had self-reported normal hearing and normal or corrected-to-normal vision.

G* power (Faul et al., 2007; Faul et al., 2009) was used to perform an a priori power analysis for a repeated measures within factors analysis of variance (ANOVA) comparing time estimations between the three editing types used in the present study (compressed-, expanded-, and real-time). The effect size for this analysis was estimated based on Cohen's (1988) guidelines. This indicated that the best estimate of the editing type standardized mean difference was $f = 0.20$ meaning that compressed-scenes will be (on average) underestimated as compared to expanded- and real-time scenes and real-time scenes will be (on average) underestimated as compared to expanded-time scenes. This effect size was, therefore, used for the power analysis with the following input parameters: α (two-sided) = 0.05, power = 0.95, correlation among repeated measures = 1:2, non-sphericity correction $\varepsilon = 1$. The power analysis results suggested that $N = 66$ are required in this study to detect a difference between the editing types with 95 % probability.

Participants were recruited through advertisements posted online or printed and placed around the university campus. Students received extra course credit for their participation, while other volunteers took part after receiving a verbal description of the task in prior. The experiment was performed in accordance with the ethical standards laid down in the 2013 Declaration of Helsinki and according to the provisions of Greek law (4521/2018) given that the Ethics Committee at Panteion University of Social and Political Sciences came into force on 28 July 2021. Verbal informed consent was obtained from all participants.

1.6.2. Stimuli and apparatus

Five different events depicting everyday activities were recorded by a professional film director using a Nikon Digital SLR Camera D5100. A single camera was utilized for video recording with the attempt to create various takes of as similar as possible events/performances. Each of the five scenes depicted one of the following events: (a) a seated man lighting and smoking a cigarette (Cigarette scene), (b) a woman entering the living room, calling someone on the phone, and leaving the room (Phone scene), (c) a woman descending some stairs, opening the front door, and leaving the house (Stairs scene), (d) a woman making tea and drinking it (Tea scene), and (e) a woman placing paper in the typewriter and typing (Typewriter scene). These events were selected as they represent simple everyday life activities, familiar to most participants. Each event was recorded from multiple angles to collect the editing material necessary for the creation of the final scenes.

Each event was edited using the match-on-action editing technique, whereby a central action is presented from different views with consecutive shots as the action unfolds (Shimamura, 2013; Smith & Henderson, 2008). This is done by segmenting each action into shots presented consecutively with cuts interfering between them (for visualization purposes, an example of this technique within a scene could consist of a first shot depicting a woman walking towards a car, the next shot showing the woman in the driver's seat shutting the door, and a third shot with the woman driving on a highway, listening to music). All cuts were performed in accordance with the principles and key techniques used in continuity editing. At the end of the editing process, a professional editor reviewed and approved the editing of the five scenes. Each of the five match-on-action scenes (depicting the activities described above) was edited using elliptical, overlapping, and minimal editing to create compressed-, expanded-, and real-time scenes, respectively (see Fig. 1). For each event, the compressed-time version contained three cuts (i.e., the scenes were composed of four shots), the expanded-time version contained five cuts (i.e., the scenes were composed of six shots), and the real-time version contained a single cut (i.e., the scenes were composed of two shots). For each event, the first and last shots between the three timing versions (i.e., compressed, expanded, and real-time) were identical, so that editing would be applied only in between shots and not in the beginning or the end of each scene. The scenes are available here https://osf.io/8jxfy/.

The editing required for time manipulation and the different number of cuts for every event version (i.e., cuts of 3, 5, and 1 for compressed, expanded, and real-time versions, respectively) resulted in physical duration differences within a specific event, thus, static parts of the dynamic scene or -when not possible- static frames were added at the beginning and the end of each event across conditions so as to have stimuli of equal duration for the three editing techniques applied to each event[1] (e.g., Sgouramani et al., 2020; Sgouramani & Vatakis, 2014). When static frames were needed, these were extracted from the first and last frames of each video stimulus using Adobe Premiere Pro CS6. The final durations for the Cigarette, Phone, Tea, Stairs, and Typewriting scenes were 22, 25, 26, 16, and 16 s, respectively. The auditory stream was removed from all video clips. The stimuli (720 × 480) were presented on a black background on a Toshiba laptop. The experiment was performed using Presentation (Version 18.1, Neurobehavioral systems, Inc.). Participants responded using a standard laptop keyboard.

1.6.3. Procedure

Participants were seated approximately 60 cm from the monitor in a quiet, dimly lit room without the presence of external distractions, as they performed a verbal estimation task[2]. A short practice block was performed prior to the beginning of the main experiment to familiarize participants with the experimental procedure. Participants were instructed to carefully observe the presented actions throughout the entire experimental procedure. They were also informed that after each video's presentation, they would be asked to estimate its duration. They were asked to make estimations that would reflect each video's duration, not that of the depicted action in each trial, as we were interested in examining duration perception at the level of a scene instead of focusing on the perceived duration of a specific action. They were also instructed to avoid the use of any counting or other time-keeping strategies. At the beginning of each block the instruction "How long did the video last? Choose a time interval between x and y seconds" appeared, where "x and y" corresponded to a range of time and participants had to provide an estimate between the two values. After the presentation of each video, participants were asked to type the perceived duration. The task was self-paced, and participants had to provide a response to proceed to the next trial.

The experimental conditions were 15 in total (i.e., 3 temporal manipulation techniques for compression-, expansion-, and real-time presentation for each of the following scenes: Cigarette, Phone, Tea, Stairs, and Typewriter) with 5 randomized repetitions for each condition. The scenes were presented in two different blocks, with randomized order of stimulus presentation within each block and randomized block presentation. The events were paired based on their durations with the Cigarette, Phone, and Tea scenes composing one block with a response interval range of 15 to 35 s and the Stairs and Type scenes composing another block with a response interval range of 5 to 25 s. The duration of the experiment was approximately 40 min, and participants were allowed to take a break halfway through the experiment.

1.6.4. Visual signal saliency calculation

Bottom-up attention or saliency is based on the sensory cues of a stimulus captured by its signal-level properties (such as spatial, temporal, and spectral contrast, complexity, scale etc.). Like competitive selection, saliency can be attributed to the feature, the stream, or the modality level. Based on perceptual and computational attention modelling studies, efficient bottom-up models and signal analysis algorithms have been developed by Evangelopoulos et al. (2008, 2013) and Koutras et al. (2015) to measure the saliencies of both the auditory and visual streams in audiovisual videos of complex stimuli such as movie video clips. These saliencies can be integrated into a multimodal attention curve, in which the presence of salient events is signified by geometrical features such as local extrema and sharp transition points.

In the present study, we used the algorithms developed by Koutras et al. (2015) to compute a temporal curve indicating the saliency of the visual stream for the stimuli presented (see Fig. 2). For visual saliency estimation, an energy-based spatio-temporal model is used, based on the Itti et al. (1998) model (Koch & Ullman, 1985), which is more relevant to the cognition-inspired saliency methods. It uses biologically plausible spatio-temporal filters, like oriented 3D (space and time) Gabor filters (3D Gabor filters are often used in video processing to extract and analyze certain features of a dynamic frame sequence over space and time, such as motion and/or edge detection, object tracking or action recognition; Ray & Chakraborty, 2019) in order to extract visual features computed through a feature competition scheme, which is motivated by the experimental evidence of a biological counterpart in the human visual system (interaction/competition among the different visual pathways related to motion/depth and gestalt/depth/color, respectively; Kandel et al., 2000). In the first phase the initial color video is split into two streams: luminance and color contrast. Then follows the core stage of our perception-inspired computational model for visual saliency (Koutras & Maragos, 2015), which is applied both on luminance and color contrast channels. This process can be divided into three individual steps. The first step consists of the Spatio-Temporal Gabor filtering (Heeger, 1987; Koutras & Maragos, 2015), while the others include post-processing procedures like energy computation and dominant energy selection applied on the resulting energy volumes (with time being the third dimension). In this way, our approach achieves the detection of both the fastest changes in the video stimuli (e.g., flicker), as well as the slowest and more complex motion changes related with action events. In the last stage, the produced energy maps can be mapped to a 1D map giving time-varying attention curves. We employed a simple 3D to 1D mapping by taking the mean value for each 2D frame slice of each 3D energy volume.

Visual inspection of the saliency temporal curves of the visual stimuli presented shows that attention should be captured at various points of each video clip without, however, any specific direction (see Fig. 2). That is, the saliency points do not seem to be affected by the timing of the cuts in each video clip nor do they show some pattern that might explain the obtained behavioral results. The Cigarette and Tea scenes seem to have the most saliency points, while the Phone, Stairs, and Typewriter scenes the least. There are no major visible saliency

[1] Time perception is critical for various aspects of behaviour depending on an event's duration (Rhodes, 2018). Scales of time are divided in the millisecond range, the interval range (from seconds to minutes), and the circadian rhythm (a 24-h light/dark cycle). Even though the stimuli utilized in the present study are within the same range (i.e., interval range), potential differences in attentional allocation or any other confounding variables are avoided by using stimuli of equal duration between the three editing techniques applied here.

[2] Verbal estimation was preferred among other time estimation tasks (e.g., production, reproduction, bisection) given the long durations of the stimuli presented. The disadvantages of this task are the potential "wild estimates" (i.e., participants can sometimes give extremely strange estimates of duration) and quantization (i.e., participants' tendency to round up or down their responses; Wearden, 2015). However, introducing ranges of time intervals that a participant can select from can eliminate the potential disadvantages of this method (Wearden, 2015).

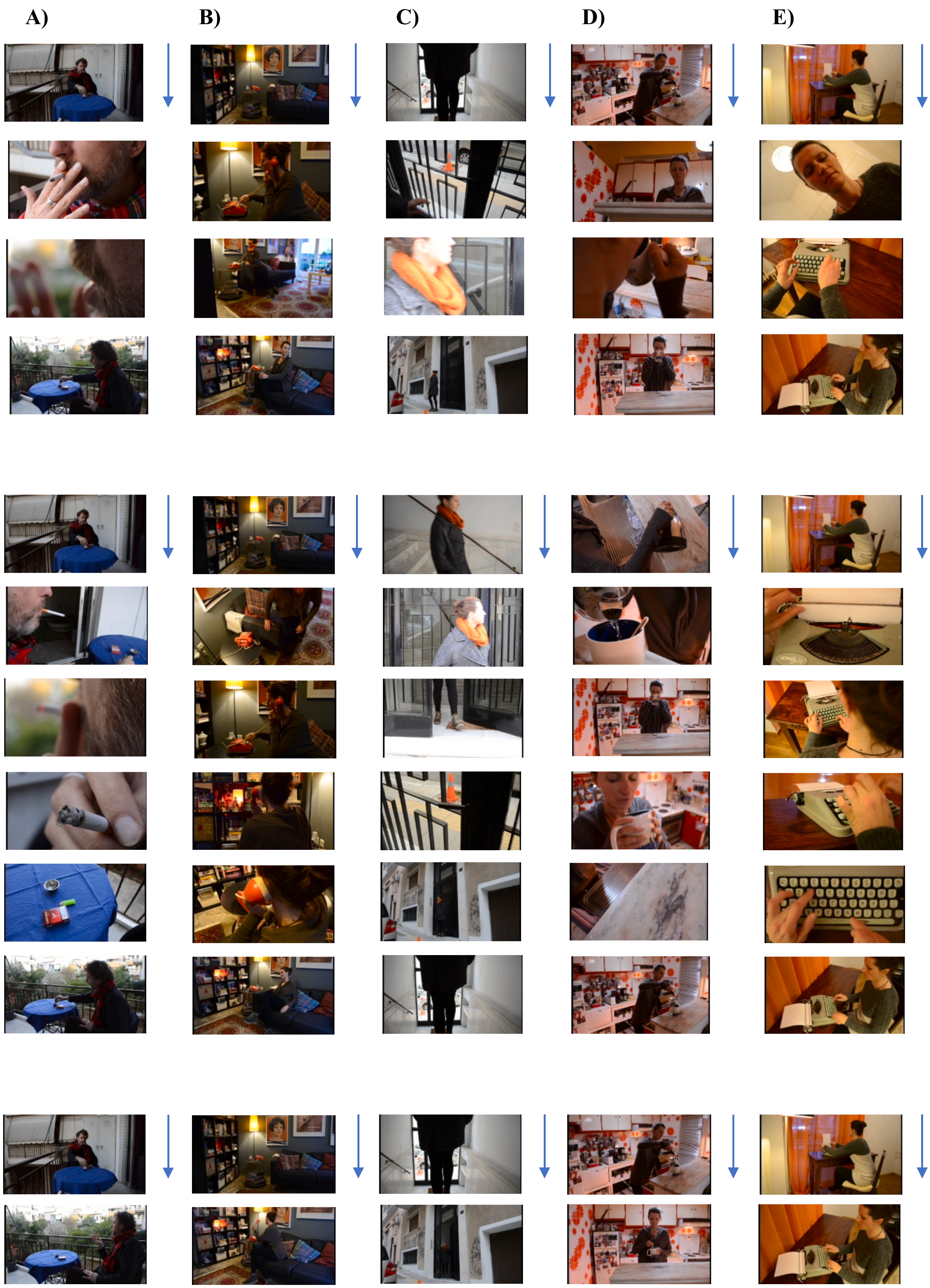

Fig. 1. Static frames extracted sequentially from the videos utilized as stimuli in our experiment. Every depicted frame was the first frame of the respective shot in each scene. Videos A (Cigarette), B (Phone), C (Stairs), D (Tea), and E (Typewriter) depict the events in their Compressed-, Expanded-, and Real-time versions, respectively.

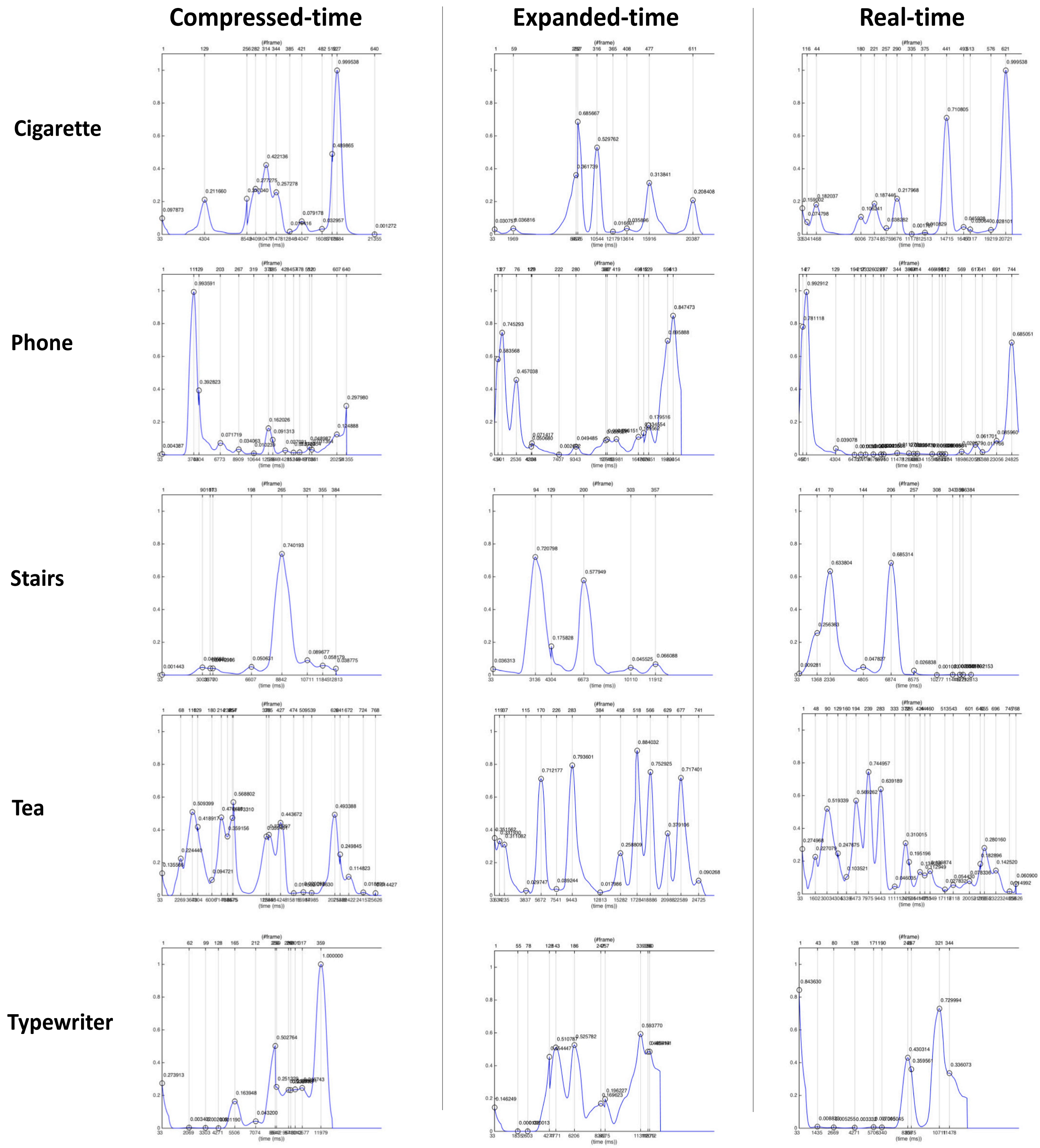

Fig. 2. Visual attentional curves indicating the saliency of the visual stream at each level of editing time (horizontally) and event type (vertically). The value at each time point reflects the overall saliency of the corresponding video frame, as derived by three feature cues: intensity, color, and motion.

differences between the compressed-, expanded-, and real-time edits of each scene. The model might provide the saliency points given the low-level features of each clip but cannot tell us how long or how strongly each point captures our attention away from timing. Overall, therefore, low-level saliency does not seem to account for any potential differences in time estimation.

1.7. Results

The experiment conducted in the present study, aimed to primarily investigate the effect of editing on participants' duration estimation. In this section, the main results of the collected data in our study are presented, focusing on participants' relative estimations of time and coefficient of variation (CV). We first obtained and averaged data from each participant separately across experimental conditions. Based on the average scores obtained in each experimental condition for each participant, we analyzed participants' relative estimated time scores and the CV to examine whether the perceived duration of the depicted events would be manipulated as a function of the Edit Type (compressed-, expanded-, and real-time editing). Relative estimated time was calculated by dividing the estimated time by the original duration of the presented video clip in each condition. This measure reflects how close participants' responses were to the physical duration of each video, demonstrating whether participants had under- or overestimated the duration of the presented video in each condition. The CV values were calculated by dividing the standard deviation by the mean duration judgement. This measure reflects the levels of variability in participants' responses in each condition with higher CV values indicating greater response variability. Repeated measures ANOVA were performed with Edit Type (compressed-, expanded-, and real-time) and Event Type (Cigarette, Phone, Stairs, Tea, and Typewriter scenes) as the within-participant factors. Event Type was included in both analyses to explore whether the effect of editing would change depending on the depiction of videos showing different everyday life activities. Bonferroni-corrected *t*-tests (where $p < 0.05$ prior to correction) were used for all post hoc comparisons.

1.7.1. Relative estimated time

The analysis of participants' relative estimated time scores revealed a significant main effect of Edit Type [$F(2,132) = 25.870$, $p < 0.001$, $\eta^2 = 0.282$], with the duration estimates for the expanded-time scenes being significantly higher ($M = 0.930$) than those for the compressed- and real-time scenes ($M = 0.878$ and 0.901, respectively), and the duration estimates for the real-time scenes were higher than those for the compressed-time scenes[3] (see Fig. 3). A significant main effect of Event Type was also revealed [$F(4,264) = 10.210$, $p < 0.001$, $\eta^2 = 0.134$], with higher duration estimates for the Cigarette scene ($M = 0.959$) as compared to all other scenes presented ($M = 0.922$, 0.892, 0.875, and 0.866 for the Phone, Stairs, Tea, and Typewriter scenes, respectively) and for the Phone scene as compared to the Tea scene.

A significant interaction between Edit Type and Event Type [$F(8,528) = 10.098$, $p < 0.001$, $\eta^2 = 0.133$] was also obtained. The duration of the compressed-time versions of the Cigarette, Stairs, Tea, and Typewriter scenes ($M = 0.921$, 0.850, 0.863, and 0.838, respectively) was underestimated as compared to their respective expanded-time versions ($M = 0.978$, 0.917, 0.898, and 0.926, respectively). Also, the compressed-time versions of the Cigarette and Stairs scenes were underestimated as compared to their real-time versions ($M = 0.978$ and 0.909, respectively), while the expanded-time versions of the Tea and Typewriter scenes were overestimated as compared to their real-time versions ($M = 0.865$ and 0.834, respectively; see Fig. 4).

1.7.2. Coefficient of variation

The analysis of the CV values revealed a main effect of Event Type [$F(4,264) = 5.669$, $p < 0.001$, $\eta^2 = 0.079$], with higher response variability for the Stairs and Typewriter scenes ($M = 0.138$ and 0.136, respectively) as compared to the Phone scene ($M = 0.1180$). There was no effect of Edit Type [$F(2,132) = 0.714$, $p = 0.492$, $\eta^2 = 0.011$] and no interaction between Event Type and Edit Type [$F(8,528) = 1.541$, $p = 0.140$, $\eta^2 = 0.023$] was obtained (see Fig. 5).

Overall, we found that both Edit Type and Event Type had a significant main effect on participants' duration judgements, while a significant interaction between the two factors was also obtained. Moreover, a significant main effect of Event Type on participant' CV values was obtained, whereas the visual signal saliency analysis suggested that there were no differences in the low-level features between compressed-, expanded- and real-time scenes.

2. General discussion

In the present study, we investigated the perceptual outcome of the compressed, expanded, and real-time editing techniques in a scene's perceived duration. Despite their consistent use, this is one of the first studies (to our knowledge) to examine the temporal perceptual outcomes of this set of editing techniques. Overall, we found that participants overestimated the expanded-time versions of the presented scenes as compared to the corresponding compressed-time scenes (except for the Phone scene, where the pattern was there but it did not reach significance) and the compressed scenes were underestimated as compared to those of real-time scenes, but only for the Cigarette and Stairs scenes. Thus, each technique's effect was in line with its predicted outcome with some inconsistencies noted as a function of Event Type.

Examination of the data and the predictions made based on the models of timing do not show perfect alignment, suggesting that the number of cuts was not the key factor influencing participants' perceived duration of a scene. Considering our data in terms of the AGM (Block & Zakay, 1997; Zakay & Block, 1997) as a function of the number of cuts presented in a scene (i.e., more processing resources allocated to non-temporal than to temporal information), we would expect expanded-time scenes to be underestimated as compared to the compressed- and real-time scenes, while compressed-time scenes would be underestimated as compared to real-time scenes. Even though a single task was employed in the present study, participants were asked to pay attention to the presented actions, while also keeping track of time as duration estimations of each video had to be provided. Therefore, the AGM and its predictions have been considered here as participants had to attend both to temporal (i.e., duration) and to non-temporal aspects (i.e., content) of each scene. Our findings only partially verified these predictions with the compressed-time Cigarette and Stairs scenes being underestimated as compared to their respective real-time scenes (the rest of the scenes showed no statistical difference in compressed- vs. real-time scenes), while the prediction that expanded-time scenes would be underestimated as compared to compressed-time scenes was not found. Similarly, the change model of duration estimation (Poynter & Homa, 1983) only partially accounts for our findings in terms of the overestimation obtained for the expanded- as compared to the compressed-time scenes, with 5 versus 3 cuts, respectively, indicating that a higher number of changes (i.e., cuts) led to an expansion of the perceived duration. However, the overestimation of real- as compared to compressed-time scenes (with 1 versus 3 cuts, respectively) cannot be accounted for. The misalignment of the data with the model's predictions validates in a way that the number of cuts or their contextual saliency was not the key determining factor for estimating time in our experiment. This notion is also supported by the visual inspection of the attentional saliency computational analysis that did not indicate any specific patterns based on editing or any sharp transition points or local extrema around the timepoint of a cut. This analysis, therefore, indicates that our findings should not be attributed to the number of changes included in the presented scenes or any attentional factors related to them. This agrees with previous research that refers to the “blindness” of

[3] As the inclusion of Event Type as a variable in our analyses was exploratory, and the main scope of this study was to examine whether the application of different editing techniques affected perceived duration, a complementary mixed model analysis was conducted with relative estimated time as the dependent variable and Edit Type as a fixed effects variable. In this analysis, participants and Event Type were treated as random effects. This analysis was conducted to account for variation across different levels of the Event Type variable. A significant main effect of Edit Type was obtained [$F(2, 871) = 27.981$, $p < 0.001$], with relative time estimation being significantly higher for expanded-, as compared to compressed- and real-time scenes ($M = 0.940$, 0.889, and 0.912, respectively), and for real-time as compared to compressed-time, indicating that participants overestimated expanded as compared to compressed and real-time scenes, and real-time as compared to compressed scenes. No other effects were found. The effect of Edit Type is, thus, also evident with the use of this analysis, in line with the findings reported in the Results.

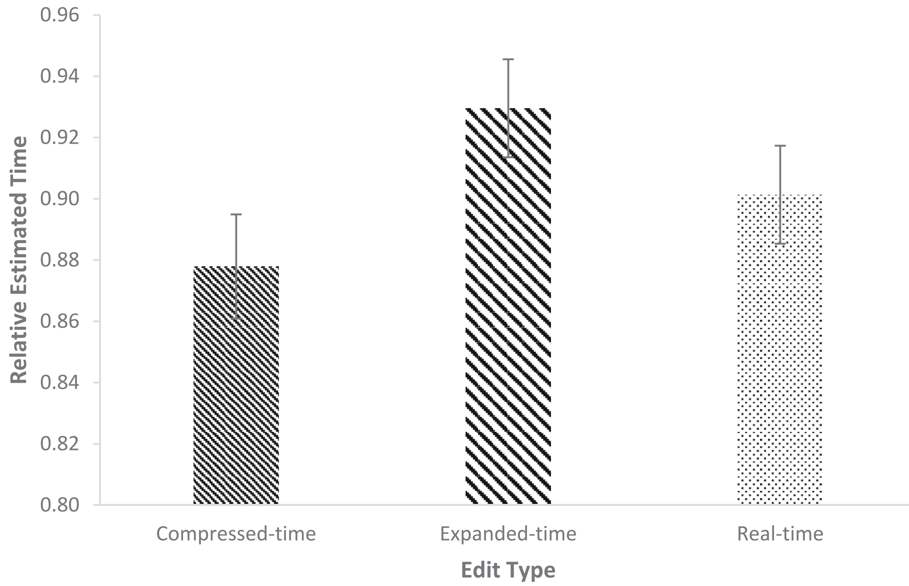

Fig. 3. Mean relative estimated time across Edit Type levels (Compressed, Expanded, and Real-time editing). Error bars represent the standard error of the mean.

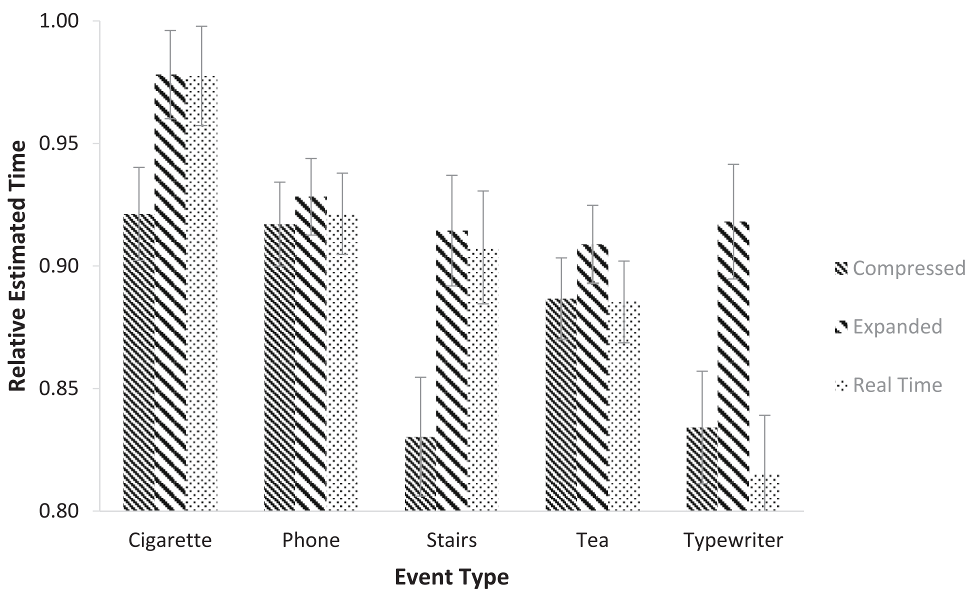

Fig. 4. Mean relative estimated time for every level of Edit Type (Compressed, Expanded, and Real-time editing) across each Event Type (Cigarette, Phone, Stairs, Tea, and Typewriter). Error bars represent the standard error of the mean.

viewers to cuts in a given scene (e.g., Smith & Henderson, 2008).

A number of low-level features such as color, luminance, shot length, and motion affect the narrative of a film and guide the viewers to a better understanding of the film (Brunick et al., 2013). For instance, as it relates to temporal perception, Brunick and her colleagues stated that manipulation of shot duration affects a viewer's perception of tempo, as well as the type of information encoded while watching a movie. In the present study, the compressed-, expanded-, and real-time versions of each action were of the same duration. As a result, manipulating the number of cuts within scenes resulted in variable shot durations between the three edit types, with real-time scenes having the longest shot duration and the expanded scenes having the shortest. Longer shot durations may allow for higher non-temporal information processing in a scene as compared to shorter shot durations where attention must be guided to the central point of interest within each shot. This, however, cannot fully explain our findings given that real-time scenes were not underestimated in comparison to all other conditions tested (except for the Typewriter scene). Additionally, Adams et al. (2000) supported that timing during film viewing is influenced by the speed and rhythm of movement and cuts reflected by shot length and motion. This influence of shot length and motion to time perception during film viewing is different under specific circumstances, whereby the influence of one factor can outweigh the influence of the other (Adams et al., 2000). Moreover, the influence of stimulus speed in time perception has been systematically reported in studies using naturalistic stimuli (e.g., Grivel et al., 2011; Sgouramani & Vatakis, 2014). Visual inspection of the attentional saliency analysis of our stimuli showed saliency differences due to motion, color, and intensity in each scene and across scenes, however, no specific pattern could be derived. The differences noted could have accounted for some of our results, however, no specific manipulations of shot length, stimulus speed or motion were implemented in our stimulation as it was beyond the scope of our study. As shown in Table 1, each of the low-level factors mentioned above influence viewers' time perception, while watching films or movie segments.

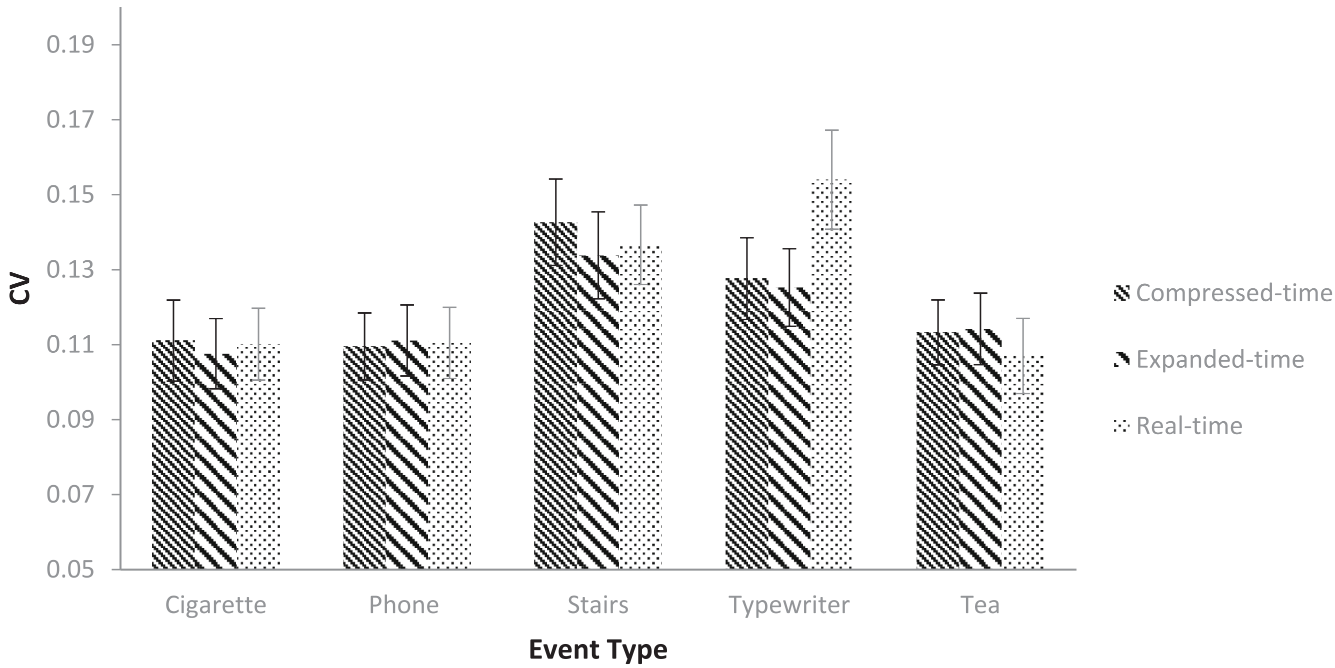

Fig. 5. Mean CV values for every level of Edit Type (Compressed, Expanded, and Real-time editing) across each Event Type (Cigarette, Phone, Stairs, Tea, and Typewriting). Error bars represent the standard error of the mean.

Table 1
A sample of low-level factors that influence time perception in films.

Factor	Influence on perceived time
Luminosity	Perceived duration is lengthened with increased luminance (e.g., Brigner, 1986; Lin, 2003).
Motion	Higher levels of stimulus motion lengthen perceived time (e.g., Brown, 1995; Eagleman, 2004).
Number of events	Sequences containing multiple distinct events result in an underestimation of the elapsed time as compared to sequences depicting a single event (e.g., Liverence & Scholl, 2012).
Pace	Fast-paced edited scenes are overestimated in terms of duration as compared to scenes with minimal to no editing (e.g., Balzarotti et al., 2021; Eugeni, 2018; Eugeni et al., 2019).
Shot length	Shots that are shorter in terms of length induce a high tempo feeling to the viewer (e.g., Broberg & Panagiotidis, 2022).
Stimulus speed	Higher levels of speed tend to lengthen perceived time as compared to lower levels of speed (e.g., Allingham et al., 2020; Grivel et al., 2011; Kaneko & Murakami, 2009; Makin et al., 2012; Matthews, 2011; Sgouramani & Vatakis, 2014; Tomassini et al., 2011).

Note. Pace in films is dependent upon multiple factors, such as rhythm of cuts, levels of motion or speed within shots, or shot length. Often, researchers use the term 'pace' referring to only one of these parameters, most commonly, shot length.

It should be pointed out that, as we manipulated the number of cuts within scenes while trying to maintain the same duration for the three edited versions of each action, this led to different shot durations between edit types, even though manipulation of shot length was beyond the scope of this study. Future studies, therefore, should be conducted to gain more insight on how these factors (i.e., number of cuts, shot duration), along with the aforementioned low-level factors (i.e., stimulus speed, motion) individually affect perceived duration within a scene.

Our analysis revealed a significant effect of Event type on participants' CV values, with time estimations for the Stairs and Typewriter scenes showing higher variability (i.e., lower timing sensitivity) than the respective estimations for the Phone scene. All the stimuli utilized in our experiment included movement as part of the presented actions, but some of the actions were more dynamic than others. The observed effect of Event type on participants' response variability could be driven by the dynamic differences between the presented actions and by the different proportion of static and dynamic shots between action types. Importantly, the levels of variability did not differ significantly between the edited versions of each scene and there was no interaction between Edit type and Event type on participants' CV values. Overall, the results of this analysis highlight that the effect of editing type in participants' duration estimation remained consistent between the different techniques examined in the present study.

The type of event presented also modulated estimation of duration with the Cigarette scene being overestimated as compared to all other scenes and the Phone scene being overestimated as compared to the Tea scene. Such differences indicate that the action depicted in a scene affects a viewer's perceived flow of time. In a recent study, Eugeni et al. (2019), investigated the effect of different actions (in terms of intentionality/goal-orientation and linearity/iteration) on duration estimation. The action presented affected participants' duration judgements with scenes showing repetitive actions with or without intent (i.e., an actor cutting a half-loaf of bread in two using a knife and an actor repeatedly moving a glass and a bread loaf on a table, respectively) to be overestimated as compared to their respective physical durations, while scenes with intended, linear action (i.e., an actor poured water into a glass and drank it) were veridical to their physical duration. The authors interpreted these findings using the Conjectural Application of Body Schemata (CABS) model for embodied time perception. This model is based on the hypothesis that the temporality of an action's body schemata operates as a matrix, where predictions about movement and action time (temporal characteristics of a scene, i.e., speed and duration) are made and tested as a scene dynamically unfolds. Thus, a linear and intentional action allows for more accurate predictions during the online viewing of a scene. This is not the case when the body schema of an action is less linear and goal-oriented, the matrix formulates weaker predictions and the comparison between these and the online experience of the viewer becomes an increasingly demanding task, as these predictions are constantly reformulated, leading to an overall overestimation of the action's duration. One could speculate that the Cigarette and Phone scenes are ambiguous in terms of linearity and intentionality as they depict repetitive activities with no specific intent (analogous to the act of cutting a half-loaf of bread in two and the act of repeatedly moving objects on a table that Eugeni and colleagues utilized) as compared to the other scenes presented (i.e., Stairs, Tea, and Typewriter, scenes analogous to the act of pouring water into a glass and drinking it), and, thus, based on the CABS model a temporal expansion would be expected in these scenes. Our findings seem to be in line with the model's predictions, however, it was beyond the scope of this study to examine the effect of action type on participants' duration judgements and an a priori classification of the presented actions would be required to draw any further conclusions. In the present study, the inclusion of the Event Type as a variable in the conducted analyses was exploratory and future studies should further examine the effect of the characteristics (i.e., intentionality, linearity) of an action to a viewer's perceived flow of time in cinematography.

In another recently conducted study, Eugeni (2018) examined the role of both the type of action and the editing style on participants' duration judgements. Three types of actions were presented, varying in terms of how familiar, goal-oriented, transitive, and ingestive they were (i.e., a clip showing an actor: a) pouring and drinking some water was classified as a familiar, transitive, ingestive, and clearly goal-oriented action, b) cutting some bread was classified as a familiar, transitive, clearly goal-oriented, but not ingestive action, and c) randomly moving some objects on a table was classified as an unfamiliar, non ingestive, and ambiguously goal-oriented action). Each action was edited according to 3 styles: master shot (no cuts and editing), slow-paced editing (5 shots and 4 cuts), and fast-paced editing (11–13 shots and 10–12 cuts). Eugeni found that both editing style and type of action affected participants' duration judgements. Specifically, in terms of editing style, he found that fast-paced edited scenes were further overestimated as compared to master shots while, in terms of action type, the unfamiliar and not clearly goal-oriented action was further overestimated as compared to the other two action types. Considering the number of cuts in the three editing styles utilized by Eugeni, that is, fast-paced editing, slow-paced editing and master shot, there is a correspondence to the expanded-time, compressed-time, and real-time scenes, respectively, utilized in our study. Our findings are partially in line with those of Eugeni's regarding the effect of editing type, given that in our study the expanded scenes (i.e., those containing the highest number of cuts among the three types of editing examined) led to duration overestimation like the fast-paced editing implemented by Eugeni. In our study, however, even though real-time scenes contained less cuts than compressed scenes, the former was further overestimated as compared to the latter. This was not the case in Eugeni's study, where no difference in participants' duration judgements was found between slow-paced edited scenes and master shots. These differences could be due to the fact that time versus pace editing may be different, thus it would be of interest for future studies to investigate the role of editing pace, along with motion speed and shot length in duration judgements, to clarify how each of these parameters may affect viewer's time estimates.

The stimuli utilized in the present study allowed for a naturalistic depiction of everyday life activities, following the editing principles applied to manipulate a scene's perceived duration. It should be noted, however, that we addressed the effect of editing at the level of a scene and not at a larger scale (i.e., within a movie or larger movie segments), or the effect of narrative. Previous studies have shown that narrative plays a crucial role in guiding viewers' attention (e.g., Cohen et al., 2015; Germeys & d'Ydewalle, 2007), and timing models such as the AGM predict that the allocation of attentional resources is directly linked to time perception (e.g., Block & Zakay, 1997; Zakay & Block, 1997). In the present experiment, we utilized short scenes without a specific theme, as this was one of the first attempts to examine and specify the effects of editing and action on time perception. Future studies should be conducted to further examine these effects over scenes or larger movie segments and investigate the role of narrative on viewers' time perception.

Overall, the present findings suggest that the elliptical, overlapping, and real-time editing techniques can successfully be utilized by filmmakers to manipulate a scene's perceived duration. This pattern, however, was not consistent across scenes, with some scenes showing no effect across conditions (i.e., phone scene), while in other scenes the real-time manipulation was underestimated as compared to the compressed-time editing conditions. Current models of timing do not fully account for our results, a finding that could be attributed to low-level feature differences (Adams et al., 2000; Brunick et al., 2013; Cutting, 2016), attentional saliency effects, or to the overall effect of the scene's narrative (even in such a short scale of stimulation). In order to further break down how the characteristics of a scene and the use of editing interact on viewers' perceived duration, future studies including the manipulation of low-level factors (such as stimulus speed, motion, shot length, or luminance), and the characteristics of an action (such as intentionality, and linearity), along with the application of different editing techniques would be valuable to investigate this issue in depth. Moreover, with the use of larger movie segments, more research should be conducted to further investigate the effect of narrative in duration perception with the use of cinematographic material.

CRediT authorship contribution statement

Lydia Liapi: Writing – original draft, Investigation, Formal analysis. **Elpida Manoudi:** Writing – review & editing, Data curation. **Maria Revelou:** Writing – review & editing, Data curation. **Katerina Christodoulou:** Writing – review & editing, Data curation. **Petros Koutras:** Writing – review & editing, Formal analysis. **Petros Maragos:** Writing – review & editing, Formal analysis. **Argiro Vatakis:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

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

Data availability

Data can be found here.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 964464.

References

Adams, B., Dorai, C., & Venkatesh, S. (2000). Study of shot length and motion as contributing factors to movie tempo. In *Proceedings of the 8th ACM international conference on multimedia* (pp. 353–355). New York, NY: ACM.

Allingham, E., Hammerschmidt, D., & Wöllner, C. (2020). Time perception in human movement: Effects of speed and agency on duration estimation. *Quarterly Journal of Experimental Psychology, 74*(3), 559–572. https://doi.org/10.1177/1747021820979518

Balzarotti, S., Cavaletti, F., D'Aloia, A., Colombo, B., Cardani, E., Ciceri, M. R., … Eugeni, R. (2021). The editing density of moving images influences viewers' time perception: The mediating role of eye movements. *Cognitive Science, 45*(4). https://doi.org/10.1111/cogs.12969

Beck, C. T. (1988). Norm setting for the verbal estimation of a 40-second interval by women of childbearing age. *Perceptual and Motor Skills, 67*(2), 577–578. https://doi.org/10.2466/pms.1988.67.2.577

Berliner, T., & Cohen, D. J. (2011). The illusion of continuity: Active perception and the classical editing system. *Journal of Film and Video, 63*(1), 44–63. https://doi.org/10.1353/jfv.2011.0008

Block, R. A. (1982). Temporal judgments and contextual change. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(6), 530–544. https://doi.org/10.1037//0278-7393.8.6.530

Block, R. A., & Reed, M. A. (1978). Remembered duration: Evidence for a contextual-change hypothesis. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 656–665. https://doi.org/10.1037/0278-7393.4.6.656

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

Bordwell, D. (1985). *Narration in the fiction film*. Madison: The University of Wisconsin Press.

Bordwell, D., & Thompson, K. (2012). *Film art: An introduction* (10th ed.). New York, NY: McGraw-Hill.

Brigner, W. L. (1986). Effect of perceived brightness on perceived time. *Perceptual and Motor Skills, 63*(2), 427–430. https://doi.org/10.2466/pms.1986.63.2.427

Broberg, F., & Panagiotidis, A. (2022). *Impact of shot length and motion on cinematic tempo (thesis)*. Sweden: KTH Royale Institute of Technology.

Brown, S. W. (1995). Time, change, and motion: The effects of stimulus movement on temporal perception. *Perception & Psychophysics, 57*(1), 105–116. https://doi.org/10.3758/bf03211853

Brunick, K. L., Cutting, J. E., & DeLong, J. E. (2013). Low-level features in film: What they are and why we would be lost without them. In A. Shimamura (Ed.), *Psychocinematics: Exploring cognition at movies* (pp. 133–148). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199862139.003.0007.

Cohen, A. L., Shavalian, E., & Rube, M. (2015). The power of the picture: How narrative film captures attention and disrupts goal pursuit. *PLoS One, 10*(12), Article e0144493. https://doi.org/10.1371/journal.pone.0144493

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Cutting, J. E. (2005). Perceiving scenes in film and in the world. In J. D. Anderson, & B. F. Anderson (Eds.), *Moving image theory: Ecological considerations* (pp. 9–27). Carbondale, IL: University of Southern Illinois Press.

Cutting, J. E. (2016). Narrative theory and the dynamics of popular movies. *Psychonomic Bulletin & Review, 23*(6), 1713–1743. https://doi.org/10.3758/s13423-016-1051-4

Cutting, J. E., & Candan, A. (2013). Movies, evolution, and mind: From fragmentation to continuity. *The Evolutionary Review, 4*(3), 25–35.

d'Ydewalle, G., & Vanderbeeken, M. (1990). Perceptual and cognitive processing of editing rules in film. In R. Groner, G. d'Ydewalle, & R. Parham (Eds.), *From eye to mind: Information acquisition in perception, search, and reading* (pp. 129–139). Amsterdam: Elsevier Science Publishers B.V. (North-Holland).

Eagleman, D. M. (2004). Time perception is distorted during slow motion sequences in movies. *Journal of Vision, 4*(8), 491. https://doi.org/10.1167/4.8.491

Espinosa-Fernández, L., Miró, E., Cano, M., & Buela-Casal, G. (2003). Age-related changes and gender differences in time estimation. *Acta Psychologica, 112*(3), 221–232. https://doi.org/10.1016/s0001-6918(02)00093-8

Eugeni, R. (2018). What time is in? Subjective experience and evaluation of moving image time. *Italian Journal of Cognitive Sciences, 1*, 81–96. https://doi.org/10.12832/90973

Eugeni, R., Balzarotti, S., Cavaletti, F., & D'Aloia, A. (2019). It doesn't seem it, but it is. A neurofilmological approach to the subjective experience of moving-image time. In A. Pennini, & A. Flazone (Eds.), *The extended theory of cognitive creativity: Interdisciplinary approaches to performativity* (pp. 243–266). Springer International Publishing. https://doi.org/10.1007/978-3-030-22090-7.

Evangelopoulos, G., Rapantzikos, K., Potamianos, A., Maragos, P., Zlatintsi, A., & Avrithis, Y. (2008). Movie summarization based on audio-visual saliency detection. In *Proceedings of the international conference on image processing (ICIP 2008)*.

Evangelopoulos, G., Zlatintsi, A., Potamianos, A., Maragos, P., Rapantzikos, K., Skoumas, G., & Avrithis, Y. (2013). Multimodal saliency and fusion for movie summarization based on aural, visual, and textual attention. *IEEE Transactions on Multimedia, 15*(7), 1553–1568. https://doi.org/10.1109/TMM.2013.2267205

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

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*, 175–191. https://doi.org/10.3758/BF03193146

Fraisse, P. (1963). *The psychology of time*. Oxford, England: Harper & Row.

Germeys, F., & d'Ydewalle, G. (2007). The psychology of film: Perceiving beyond the cut. *Psychological Research, 71*(4), 458–466. https://doi.org/10.1007/s00426-005-0025-3

Ghosh, A. (2022). Analysis of single-shot and long-take filmmaking: Its evolution, technique, mise-en-scène, and impact on the viewer. *Indian Journal of Mass Communication and Journalism, 2*(2), 4–12. https://doi.org/10.54105/ijmcj.b1023.122222

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences, 423*, 52–77. https://doi.org/10.1111/j.1749-6632.1984.tb23417.x

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Grivel, J., Bernasconi, F., Manuel, A. L., Murray, M. M., & Spierer, L. (2011). Dynamic calibration of our sense of time. *Neuropsychologia, 49*, 147–150. https://doi.org/10.1016/j.neuropsychologia.2010.11.004

Hancock, P. A., & Rausch, R. (2010). The effects of sex, age, and interval duration on the perception of time. *Acta Psychologica, 133*(2), 170–179. https://doi.org/10.1016/j.actpsy.2009.11.005

Heeger, D. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America, 4*(8), 1455–1471.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*, 1254–1259.

Kandel, E., Schwartz, J., & Jessell, T. (2000). *Principles of neural science* (4th ed.). New York: McGraw-Hill.

Kaneko, S., & Murakami, I. (2009). Perceived duration of visual motion increases with speed. *Journal of Vision, 9*(14), 1–12. https://doi.org/10.1167/9.7.14

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4*, 219–227.

Koutras, P., & Maragos, P. (2015). A perceptually based spatio-temporal computational framework for visual saliency estimation. *Signal Processing and Image Communication, 38*, 15–31.

Koutras, P., Zlatintsi, A., Iosif, E., Katsamanis, A., Maragos, P., & Potamianos, A. (2015). Predicting audio-visual salient events based on visual, audio and text modalities for movie summarization. In *Proceedings of the international conference on image processing (ICIP 2015)*.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12*(2), 72–79. https://doi.org/10.1016/j.tics.2007.11.004

Lin, C. C. (2003). Effects of illumination, viewing distance, and lighting color on perception time. *Perceptual and Motor Skills, 96*(3), 817–826. https://doi.org/10.2466/pms.2003.96.3.817

Lindblom, O. (2015, May 26). *Images in time: Expressing and manipulating time in cinema*. Videomaker. https://www.videomaker.com/article/c18/17968-images-in-time-expressing-and-manipulating-time-in-cinema/.

Liverence, B. M., & Scholl, B. J. (2012). Discrete events as units of perceived time. *Journal of Experimental Psychology: Human Perception and Performance, 38*(3), 549–554. https://doi.org/10.1037/a0027228

Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology, 15*(5), 533–545. https://doi.org/10.1002/acp.724

Magliano, J. P., & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive Science, 35*(8), 1489–1517. https://doi.org/10.1111/j.1551-6709.2011.01202.x

Makin, A. D. J., Poliakoff, E., Dillon, J., Perrin, A., Mullet, T., & Jones, L. A. (2012). The interaction between duration, velocity and repetitive auditory stimulation. *Acta Psychologica, 139*, 524–531. https://doi.org/10.1016/j.actpsy.2012.01.013

Manipulating Time in Video Production. (n.d.). Retrieved from https://www.mediacollege.com/video/editing/time/.

Mariani, L. (2018, October). Film language. cinemafocus.eu https://www.cinemafocus.eu/Linguaggio%20cinematografico/suono.html.

Matthews, W. J. (2011). How do changes in speed affect the perception of duration? *Journal of Experimental Psychology. Human Perception and Performance, 37*, 1617–1627. https://doi.org/10.1037/a0022193

Penton-Voak, I. S., Edwards, H., Percival, A., & Wearden, J. H. (1996). Speeding up an internal clock in humans? Effects of click trains on subjective duration. *Journal of Experimental Psychology: Animal Behavior Processes, 22*(3), 307–320.

Poynter, W. D. (1983). Duration judgment and the segmentation of experience. *Memory & Cognition, 11*(1), 77–82. https://doi.org/10.3758/BF03197664

Poynter, W. D. (1989). Judging the duration of time intervals: A process of remembering segments of experience. In I. Levin, & D. Zakay (Eds.), *Time and human cognition: A life-span perspective* (pp. 305–321). Amsterdam: Elsevier.

Poynter, W. D., & Homa, D. (1983). Duration judgment and the experience of change. *Perception & Psychophysics, 33*(6), 548–560. https://doi.org/10.3758/bf03202936

Ray, K. S., & Chakraborty, S. (2019). Object detection by spatio-temporal analysis and tracking of the detected objects in a video with variable background. *Journal of Visual Communication and Image Representation, 58*, 662–674. https://doi.org/10.1016/j.jvcir.2018.12.002

Reisz, K., & Millar, G. (2010). *The technique of film editing* (2nd ed.). London: Focal Press.

Rhodes, D. (2018). On the distinction between perceived duration and event timing: Towards a unified model of time perception. *Timing & Time Perception, 6*(1), 90–123. https://doi.org/10.1163/22134468-20181132

Sgouramani, H., Chatziioannou, I., & Vatakis, A. (2020). Modulating subjective timing through looming and receding biological motion. *Timing & Timing Perception, 8*(3–4), 221–238. https://doi.org/10.1163/22134468-bja10006

Sgouramani, H., & Vatakis, A. (2014). "Flash" dance: How speed modulates perceived duration in dancers and non-dancers. *Acta Psychologica, 147*, 17–24.

Shimamura, A. P. (2013). Psychocinematics: Issues and directions. In A. Shimamura (Ed.), *Psychocinematics: Exploring cognition at movies* (pp. 1–26). New York: Oxford.

Sieghartsleitner, M. (2018). Aspects of elliptical editing. *International Journal of Film and Media Arts, 3*(2), 148–163.

Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections: The Journal for Movies and the Mind, 6*(1), 1–27. https://doi.org/10.3167/proj.2012.060102

Smith, T. J., & Henderson, J. M. (2008). Edit blindness: The relationship between attention and global change blindness in dynamic scenes. *Journal of Eye Movement Research, 2*(2), 1–17. https://doi.org/10.16910/jemr.2.2.6

Tan, E. S. (2018). A psychology of the film. *Humanities and social sciences communications, 4*(1), 1–20. https://doi.org/10.1057/s41599-018-0111-y

Tomassini, A., Gori, M., Burr, D., Sandini, G., & Morrone, M. C. (2011). Perceived duration of visual and tactile stimuli depends on perceived speed. *Frontiers in Integrative Neuroscience, 5*. https://doi.org/10.3389/fnint.2011.00051

Treisman, M. (1963). Temporal discrimination and the indifference interval. Implications for a model of the "internal clock". *Psychological Monographs, 77*, 1–31.

Wearden, J. H. (2015). Mission: Impossible? Modelling the verbal estimation of duration. *Timing & Time Perception, 3*(3–4), 223–245. https://doi.org/10.1163/22134468-03002051

Zakay, D., & Block, R. A. (1997). Temporal cognition. *Current Directions in Psychological Science, 6*(1), 12–16. https://doi.org/10.1111/1467-8721.ep11512604