Event boundary perception among the visually impaired in audio described films

Johansson, Roger; Rastegar, Tina; Lyberg Åhlander, Viveka; Holsanova, Jana

DOI:
10.31234/osf.io/7bmdx

2023

Document Version:
Early version, also known as pre-print

Event Boundary Perception Among the Visually Impaired in Audio Described Films

Roger Johansson[1*], Tina Rastegar[1], Viveka Lyberg-Åhlander[2, 3] & Jana Holsanova[4]

[1] Department of Psychology, Lund University, Sweden

[2] Department of Clinical Sciences in Lund, Logopedics, Phoniatrics and Audiology, Lund University, Sweden

[3]Speech Language Pathology, Faculty of Arts, Psychology and Theology, Åbo Akademi University, Turku, Finland

[4] Lund University Cognitive Science, Lund University, Sweden

Author note

* Correspondence concerning this article should be addressed to Roger Johansson, Department of Psychology, Lund University, Box 213, SE-221 00 Lund, Sweden, email: roger.johansson@psy.lu.se

Abstract

Audio description (AD) serves a critical role in making film narratives accessible to visually impaired audiences, aiming to enhance their viewing experience and comprehension. One method to assess the comprehension of film narratives, is through an event segmentation task, wherein participants delineate the narrative unfolding into distinct meaningful events. In the present study, both sighted and visually impaired participants engaged in such tasks. Sighted participants watched a Swedish film, while visually impaired participants experienced the same film with two AD versions—one explicitly expressing key event boundaries and another containing more implicitly conveyed ones. Our findings indicate that visually impaired participants perceived event boundaries similarly to sighted participants, suggesting that AD effectively conveys the event structure. However, in the AD version with implicit expressions, event boundaries were less likely to be recognized. These results shed light on event segmentation dynamics in films, emphasizing the importance of how event boundaries are presented in AD. This has significant implications for improving the cinematic experience for visually impaired viewers, emphasizing the need for clear, explicit information about event boundaries within AD.

Event Boundary Perception Among the Visually Impaired in Audio Described Films

In order to provide access to the content of moving images and audio-visual media for individuals with visual impairments, the practice of audio description (AD) is employed. AD consists of verbal descriptions of visual events, with the primary objective of enhancing the accessibility of visual information and facilitating a more comprehensive and immersive understanding, experience, and enjoyment of various forms of media, including films and television shows (Holsanova, 2016; 2022; Lopez, Kearney, & Hofstädter, 2022). To achieve this, audio describers meticulously select pertinent details from visual events, such as descriptions of environments, objects, individuals and their physical attributes, clothing, facial expressions, actions, gestures, and bodily movements. These selected details are then articulated verbally, utilizing vivid language to evoke inner mental imagery and enrich the process of meaning-making for the visually impaired audience (Holsanova, 2016; 2022), with the goal to reach "narrative equivalence" between the source film and the audio described film (Vandaele, 2012). However, the necessity for AD extends beyond mere comprehension, aligning with the principles outlined in the UN Convention on the Rights of Persons with Disabilities (CRPD), which emphasizes the importance of ensuring content accessibility for all (United Nations, 2006).

The production of AD entails a complex task that necessitates the integration of verbal descriptions with the sounds, voices, and dialogues present in the original audio-visual content. This synchronization requires precise timing during the narration process. For instance, it is imperative for the audio describer to refrain from speaking during dialogues and to discern when and how to elucidate specific sounds, such as identifying the source of gunshots (who is shooting at whom) or allowing sounds to convey their meaning independently, as in the case of a door knocking or a ringing telephone. In essence, the audio describer must continually assess *what* to describe, *when* to provide descriptions, and *how* to articulate them (Holsanova, 2016;

2022; Vercauteren, 2021). The effectiveness of this form of communication relies heavily on fundamental cognitive processes governing how sighted audio describers perceive and segment the unfolding narrative events within a film, and critically on how the visually impaired receivers perceive and conceptualize the structure, content, and segmentation of these events in relation to the provided AD (Holsanova, 2016; 2022; Vandaele, 2012).

Research conducted over the last three decades has extensively examined the structuring and differentiation of events, revealing that human cognition is equipped with inherent neurocognitive principles enabling the perception of events as distinct entities (Radvansky & Zacks, 2014; Kurby & Zacks, 2008; Zwaan & Radvansky, 1998). Consequently, a central inquiry in the context of successful AD pertains to how the sequence of events unfolding in a film aligns with the verbal narration conveyed through AD. Event segmentation also holds significance in the domain of computer-generated video description, an emerging field aimed at augmenting human-generated AD to make audio-visual media more widely accessible (Braun, Starr & Laaksonen, 2020; Starr, Braun & Delfani, 2020).

Notably, while previous research has recognized the importance of event segmentation in relation to AD production (Holsanova, Blomberg, Blomberg, Gärdenfors, & Johansson, 2023), prior investigations have not examined the perception of event segmentation among the visually impaired who are receiving the AD.

Event segmentation

Human existence transpires within the framework of time, where vast amounts of intricate information continuously interweave into an unbroken stream of experiences. However, our mental representations of the world, do not mirror this continuous stream. Instead, akin to the scenes in a movie, we comprehend and remember our experiences as distinct and meaningful chains of events (cf., Radvansky & Zacks, 2014). According to Zacks et al. (2001), an event is a perceptual time segment characterized by a discernible beginning and end, which can be defined as "a segment of time at a given location that is conceived by an observer to have a beginning and an end" (Zacks & Tversky, 2001; Kurby & Zacks, 2008). Thus, the process of event segmentation involves the demarcation of time segments into discrete events (Radvansky & Zacks, 2017), where the event boundaries, signifying transitions from one event to the next, typically manifest when significant narrative alterations occur, such as shifts in spatiotemporal configurations (Zacks, 2013). Despite individual variations, there exists a general consistency in how individuals segment events, suggesting the presence of shared cognitive processes (Radvansky & Zacks, 2017). Accumulating research demonstrates considerable agreement among individuals in discerning distinct and meaningful event units in narrative texts and films (e.g., Huff, Meitz, & Papenmeier, 2014; Newtson, 1973; Zacks et al., 2009; Zacks & Tversky, 2001), suggesting the existence of overarching cognitive principles governing the automatic processing and segmentation of event information. This concordance is corroborated by contemporary neurocognitive research, which has elucidated the neural mechanisms underlying event segmentation (e.g., Amoruso et al., 2013; Baldassano et al., 2018; Ezzyat & Davachi, 2011; Zacks et al., 2001). Notably, transient changes in specific brain activity have been observed not only during active event segmentation, but also during passive viewing of films (Zacks et al., 2001). Furthermore, event segmentation competence has been observed to influence memory outcomes. Specifically, when an individual's segmentation closely aligns with the group mean, their memory for associated events tends to exhibit superior performance (Sargent et al., 2013; Zacks, Speer, Vettel, & Jocaboy, 2006). Additionally, memories of information coinciding with event boundaries are often better retained (Huff et al., 2014; Pettijohn, Thompson, Tamplin, & Radvansky, 2016). Consequently, the manner in which we segment our experiences as they unfold within a narrative fundamentally shapes our perception, comprehension, and memory of it.

Identifying event boundaries during film perception

When studying event segmentation, it is imperative to identify qualitative aspects marking the transition from one event to another. Various aspects for identifying event boundaries have been proposed, primarily grounded in studies of narratives across different media. One influential model is the event-indexing model (Zwaan, Langston, & Graesser, 1995; Zwaan & Radvansky, 1998). Rooted in the analysis of literary narratives (fiction), this model posits several indices involved in processing a narrative, including relationships among characters, time, and space, serving as foundational elements for event segmentation. In comprehending a narrative, one must monitor these indices and distinguish among the events occurring in different locations, at different times, or involving distinct characters that act within unique spatiotemporal contexts.

Extending the event-indexing model, subsequent studies have applied it to the analysis of cinematic narratives, yielding two prominent conceptual frameworks proposed by Zacks et al. (2009) and Cutting and Iricinschi (2015) for the purpose of event classification and segmentation. These frameworks exhibit certain disparities in the manner in which they demarcate event boundaries. Zacks et al. (2009) distinguish six distinct event change categories: Cause, Character, Goal, Object, Space, and Time, whereas Cutting and Iricinschi (2015) employ a more streamlined classification system comprising only three categories: Location, Time, and Character. Nevertheless, both frameworks share common ground by identifying temporal and spatial alterations as pivotal factors in event segmentation. Furthermore, they acknowledge that the perception of spatiotemporal event boundaries frequently coincides with the visual disruptions introduced by film cuts, even though cuts themselves have been found not to directly precipitate event segmentation (Zacks, 2021).

In the present study, we focus on changes in time and location. This choice is primarily motivated by the salience of temporal and locational changes, which have been central in prior research (e.g., Cutting and Iricinschi, 2015; Speer et al., 2003; Speer & Zacks, 2005; Zacks et al., 2010; Zacks, 2021) and are expected to be conveyed by the audio describer (see Holsanova et al., 2023; Vercauteren, 2021; Rai, Greening, & Petré, 2010). Spatiotemporal settings constitute fundamental narrative building blocks and are intrinsically connected to characters and their actions. Such settings can be associated with specific characters (e.g., characters linked to a particular pub, car, school, or city) or represent different time periods (e.g., flashbacks to the 1960s versus the present). Spatiotemporal settings have been demonstrated to be integral to AD (Remael & Vercauteren, 2015; Vandaele, 2012; Vercauteren, 2021; Vercauteren & Remael, 2015), with most AD guidelines recommending the inclusion of spatiotemporal features in descriptions (cf., Rai et al., 2010; Vercauteren & Remael, 2015).

The present study

A pivotal dimension in elevating the quality of AD pertains to assessing the degree to which the AD experience can be aligned with the experience of sighted individuals viewing a film without AD (Vandaele, 2012). To reach such "narrative equivalence" it is necessary to understand how AD influences the perception and segmentation of the unfolding chain-of-events. Despite the limited extent of research in this field, earlier investigations have contributed valuable insights. Our prior research (Holsanova et al., 2023) studied event segmentation in AD *production*, emphasizing a significant overlap between event boundaries and the resultant AD. However, there was large variability in how event boundaries were expressed, both within and across individual audio describers (Holsanova et al., 2023). In a study of audio dramas, Huff et al. (2018) conducted a comparative analysis involving both sighted and non-sighted participants, revealing that event segmentation in auditory contexts transcended modality, indicating a commonality in cognitive processes that is not contingent on auditory experience. However, it is worth noting that their research centered on audio dramas, which differ substantially from multimodal films that rely on a fusion of visual, verbal and aural cues (Holsanova 2022, Remael & Reviers 2018). Consequently, the current body of knowledge concerning event segmentation and the reception of AD among the visually impaired remains rather scant.

The aim of the present study was to investigate how non-sighted participants, experiencing a film with AD, perceive event boundaries in comparison to their sighted counterparts who view the same film without AD. We placed a particular emphasis on spatiotemporal event boundaries and examined the influence of AD on the perception of these event boundaries, specifically how explicitly articulated descriptions of event boundaries compared to more implicitly conveyed ones.

Our expectation was that event segmentation would occur in a comparable fashion between the two groups, and we hypothesized that the presence of implicit AD would diminish the likelihood of perceiving the corresponding event boundaries when compared to explicit AD and to visual film perception.

To accomplish these objectives, we conducted a study where sighted participants watched a film with the task to report when they encountered an event boundary. We then compared their performance to non-sighted participants who experienced the same film with AD, undertaking the same event segmentation task. Apart from that, we manipulated how event boundaries were expressed in the AD: either explicitly, with the inclusion of an explicit label denoting the event boundary, or implicitly, devoid of such explicit labels.

Methods

Participants

In total, a cohort of fifty-three adults (38 female) with a mean age of 38.5 years (SD = 12.5) participated in the study. Among these participants, 22 individuals met the criteria for non-sighted status, defined in accordance with the World Health Organization's standards for blindness or severe visual impairment. The remaining 31 participants were sighted individuals with no reported visual impairments.

The non-sighted participants were recruited through various organizations and networks dedicated to serving the visually impaired community in Sweden, including the Swedish Association of the Visually Impaired (SRF), the Swedish Braille Authority (Punktskriftsnämnden), the Swedish Agency for Accessible Media (MTM), and Young People with Visual Impairment (US). Five non-sighted participants were excluded. The exclusions consisted of two participants with only mild visual impairment, one participant due to technical complications and two participants due to failure to conform to instructions. Among the remaining 17 non-sighted participants, 10 were congenitally blind, and the remaining seven had lost their sight later in life, with a mean loss of sight at 19 years old (SD = 16.9) and with a mean duration of 30 years (SD = 21) of living with the condition. Within the sighted group, five participants were excluded; three were excluded due to technical issues, and two due to failure to conform to instructions.

As a result, the analyses and findings of this study are based on a total of 44 participants (28 female, 2 non-binary), with a mean age of 37.5 years (SD = 12.8). This participant pool consisted of 17 non-sighted individuals (10 female, 1 non-binary) with a mean age of 43.1 years (SD = 15.3) and 27 sighted individuals (18 female, 1 non-binary) with a mean age of 34.2 years (SD = 9.5).

All participants provided informed consent, either in written or oral form, and received compensation in the form of vouchers. Ethical approval was obtained in advance from the Swedish Ethical Review Authority under registration number 2019-03445. All research methods adhered to the ethical standards outlined in the Swedish Act concerning the Ethical Review of Research Involving Humans (2003:460) and the Code of Ethics of the World Medical Association, as articulated in the Declaration of Helsinki.

Material

Two AD versions of the Swedish film *Skumtimmen* (Alfredson, 2013) were recorded by a professional audio describer. Skumtimmen is a Swedish-produced film with a theatrical release in 2013 (English title: Echoes from the Dead). The film is a drama-thriller centred around Julia whose son mysteriously disappeared 21 years ago. Now Julia has returned to her childhood home and the truth of past events is starting to resurface. The film takes place over three different time periods and has a runtime of 99 min. In the present study, participants were exposed to the first 43 min of the film.

Utilizing the event-index model (Zwaan et al., 1995; Zwaan & Radvansky, 1998) and its extension to motion pictures (Zacks et al., 2010; Cutting & Iricinschi, 2015), the film had previously undergone systematic coding for event boundaries aligning with spatial and temporal alterations (see Holsanova et al., 2023). Spatial changes either involved a transition from one location to another within the same time frame (e.g., from the outside to the inside of a house), or a distinct transition within one and the same location setting (e.g., the protagonist moving from one room to another in the same house). Temporal changes either concerned different points in time (e.g., shifts between different time periods through flashbacks and flashforwards) or a distinct change in the time of the day (e.g., a shift from morning to evening). In the 43 minutes of the film that was used in the present study, 92 spatiotemporal boundaries had been pre-coded (81 spatial changes, 11 temporal changes).

In the two AD versions, twenty salient event boundaries (10 spatial, 10 temporal) were expressed either explicitly or implicitly. For example, when the spatial setting changed from one to another, this event boundary was in the explicit version, explicitly expressed with a label for the new setting (e.g., “at the retirement home”, “at the police station”, “in the house”). In the implicit version, no explicit labels were expressed. Instead, the changes in spatial settings were implicitly expressed through associated characters and the visual content in the scene. For example, instead of explicating the location setting of the retirement home, this location change is only implied by the description “Julia and Lennart approaches Gerlof” (Julia is the protagonist and Gerlof is her father who is situated at the retirement home). Another example involves when the location setting changes for the protagonist Julia from her being outside to inside her house. Instead of explicating the spatial change to “in the house”, this is only implied by describing Julias activity: “Julia sits on her knees and empties a cupboard of aluminium pans, tin trays and bowls”.

For temporal changes, explicit labels for different time periods and time of the day were similarly expressed (“in the past”, “in the present”, “the 60’s”, “the next day”, “later in the evening”). In the implicit version, no such explicit labels were provided, instead the temporal changes were implied through characters associated with a certain time-period, or through descriptions of visual content associated with a time of the day. For example, rather than providing an explicit delineation of a flashback to the past, one may opt to convey this temporal transition by exclusively delineating the actions of a character concomitant with that historical epoch. By way of illustration, a transition from a contemporary occurrence to one transpiring in the past could be overtly articulated as follows: "Martin trembles and a tear runs down his cheek {event boundary} In the 1960s, Nils is running in the rain." However, as it has been previously established that Nils' activities transpire within the temporal confines of the 1960s, the temporal transition may be more implicitly manifested as follows: "Martin trembles and a tear runs down his cheek {event boundary} Nils is running in the rain." In this particular instance, the introduction of the event boundary is effected through the naming of Nils, without using an explicit reference to the temporal shift. Another example involves when the temporal setting changes for the protagonist Julia to the next day. Instead of explicating the temporal change to “the next day”, this is only implied by describing Julias activity in the morning of the next day: “Julia places a cardboard box in the open trunk of the red car” (in the previous scene she had been drinking wine in the evening).

In the present study, 8 of the non-sighted participants experienced the movie with the explicit AD, and the other 9 with the implicit AD.

Procedure and materials

The experiment was conducted remotely via video-call using the Zoom platform. Participants viewed the film by screen sharing, with the experimenter's camera deactivated and microphone muted to prevent any interference from background noise. At the commencement of each session, participants were provided with an explanation of the procedure and were required to provide verbal informed consent, which was audio-recorded and securely stored.

The study involved three versions of the movie: the original film for the sighted group (control condition) and two audio-described versions for non-sighted participants, randomly assigned. One audio-described version employed explicit descriptions for a subset of spatiotemporal event boundaries, while the other used implicit descriptions for the same subset of event boundaries. Both audio-described versions were developed in collaboration with a professional audio describer.

Participants were instructed to vocally indicate the onset of a new event (referred to as an "event boundary") by saying "Now." The experiment leader recorded each perceived event boundary using the multimedia annotation software ELAN 6.4 in segmentation mode (Sloetjes & Seibert, 2016). Before the formal experiment, participants practiced this procedure by watching a brief clip (approximately 2 minutes) from the TV series "Pippi Longstocking". The non-sighted participants experienced this clip with AD narrated by the same audio describer as in the main experiment.

Following the film viewing, participants provided background information, including age, gender, duration of visual impairment (for non-sighted participants), and any hearing impairments. All participants received a 50 SEK online voucher as compensation for their participation.

Data Analysis

The data were analyzed in three steps. First, we explored the agreement in perceiving event boundaries within the sighted and non-sighted groups, employing an unconditional data-driven approach based on the temporal points derived from participants' responses. Second, we scrutinized the performance of both the sighted and non-sighted groups in identifying spatiotemporal event boundaries by contrasting their responses with the pre-established event boundaries. In the third step, we examined the impact of using explicit versus implicit event boundary descriptions in AD for the non-sighted group, considering two versions of AD in relation to the pre-established event boundaries, as well as in relation to the responses of the sighted group.

Segmentation agreement

For the data-driven exploration of identified event boundaries, we first analyzed the number of identified event boundaries across the two groups over the whole film clip using a welch's t-test.

In the next step, we treated time as a continuous variable and estimated the perception of an event boundary by modeling it as a kernel density distribution function centered around the response of each participant. Time points extracted from participants' responses were subjected to a detailed analysis aimed at identifying convergent patterns using the *segmag package* for R (Papenmeier, 2016). The segmag package operates by calculating a collective norm that reflects the proportion of participant-defined segmentations for each specific time point (divided into 1-second bins). Subsequently, individual participant responses are correlated against this collective norm, taking into account variations in response frequencies. This correlation process yields a segmentation agreement value, which is used to obtain an approximation of the temporal instances at which event boundaries converged across participants. To ascertain which of the "agreement peaks" exhibited a statistically significant difference from others, a critical segmentation magnitude is determined by simulating responses under the null hypothesis that they are randomly distributed across participants (Papenmeier, 2016). Event boundaries with segmentation magnitudes surpassing the established cut-off were considered statistically significant at a confidence level of 95%.

Finally, we created 10-second bins and registered whether (or not) each participant indicated an event boundary within that bin. Then, based on previous approaches, we used point-biserial correlations to calculate the agreement between each individual's segmentation and the proportion of participants that identified a boundary at each 10-second bin (cf., Kurby & Zacks, 2011; Zacks, Swallow, Vettel, & McAvoy, 2016). The point-biserial correlations were conducted separately for the sighted and non-sighted groups. The larger the correlation, the better agreement with the group.

Identifying spatiotemporal event boundaries

In order to compare the event boundaries identified by participants with the pre-established event boundaries associated with spatiotemporal changes, we applied a "hit" criterion. A participant's response was considered a "hit" if it fell within a time window of -5 to +10 seconds relative to the onset of the pre-established event boundary. This relatively "generous" time window was chosen to accommodate two primary factors: the variability in individual responses and the extended nature of the experience of AD compared to visually perceiving the visual chain of events. Furthermore, in certain instances, the AD preceded the visual event boundary cue. Consequently, the timing of event boundaries identified in the context of AD was anticipated to exhibit lower precision and greater variability across different individuals.

Statistical analyses were carried out utilizing Generalized Linear Mixed-Effect Models (Gallucci, 2019) with the aid of Jamovi version 1.6.23 (The jamovi project, 2019). These models, incorporating data from all data points, offer enhanced statistical power compared to conventional analyses of variance. Participants and Event boundary items (the pre-established instances of spatiotemporal event boundaries) were included as random effects (intercepts). Model-fit assessment involved contrasting the deviance of the proposed model with an unconditional null model, encompassing solely the intercept and random factors. In constructing models with multiple independent variables, a backward selection approach was employed, commencing with a maximal model comprising all variables and interactions. Likelihood-ratio tests were then used to compare models, progressively eliminating non-significant effects until no further model changes yielded a significant likelihood-ratio test ($p < .05$). Models were fitted employing restricted maximum likelihood (REML), and Satterthwaite approximations were employed to evaluate the significance of individual predictors.

Results

Segmentation agreement

On average the non-sighted participants with AD registered 74 (SD = 58) event boundaries, and the sighted group without AD registered 67 (SD = 36) event boundaries. A Welch's t-test showed no significant difference between the two groups ($t(24.0) = 0.46$, $p = .65$,

$d = 0.15$). Thus, on a global level, the two groups reported a similar amount of event boundaries in their film experience.

Using the segmag package (Papenmeier, 2016), we explored the segmentation patterns across the sighted participants watching the film without AD and the non-sighted participants with AD. Figure 1 displays segmentation behavior across time for both groups. The segmentation plots show significant correspondence within groups, indicated by peaks highlighted in green, revealing similar response patterns across the two groups. However, the plots also suggest that there were more noise and less homogeneity in the segmentation behavior for the non-sighted group, which is expected given the lower number of participants in this group and the fact the AD is more variable in time when compared to the more immediate event changes provided by the visual input (e.g., when the spatiotemporal context is changed through scene cuts).

To analyze whether the non-sighted reported event boundaries at similar time points as the sighted, we first calculated segmentation agreement within groups, resulting in mean point-biserial correlations of $r = .53$ for the non-sighted and of $r = .58$ for the sighted. A Welch's t-test showed no significant difference in within-group-agreement between the two groups, $t(24.9) = 0.89$, $p = .38$, $d = 0.29$). Correlating both groups' agreement histograms showed a high correlation of $r = .70$, indicating that there were indeed extensive overlaps in when the two groups indicated their event boundaries. Nonetheless, calculating segmentation agreement for the non-sighted participants in relation to the sighted group, with the mean point-biserial correlation of $r = .37$, revealed that agreement was significantly lower than when comparing it to agreement within their own group ($t(16) = 10.5$, $p < .001$, $d = 2.56$).

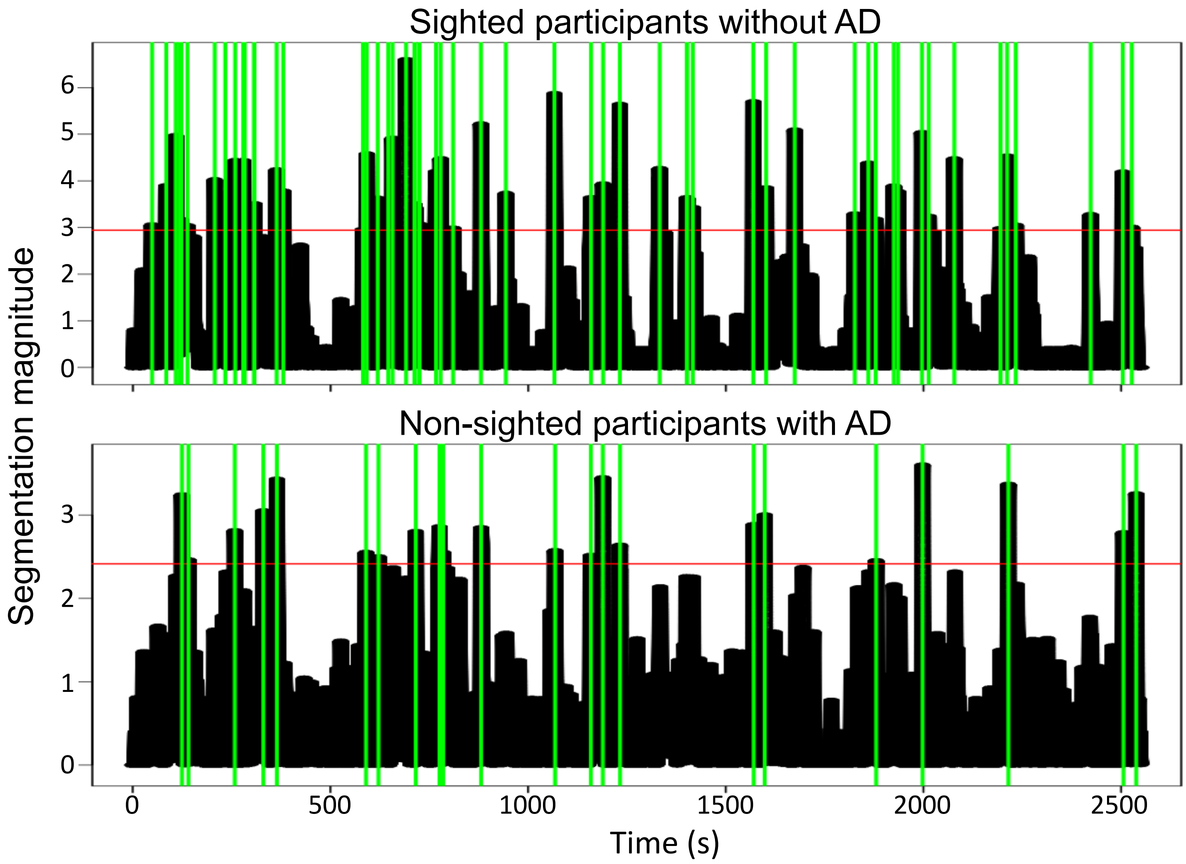

Fig. 1. *Event segmentation while watching the film "Skumtimmen" for both sighted participants without AD (top) and non-sighted participants with AD (bottom). Using the segmag package (Papenmeier, 2016) significant event boundaries (α < .05) are presented in vertical green lines based on their surpassing of critical cutoffs (represented by red horizontal lines) derived from bootstrapping procedures aimed at testing the null hypothesis that all key presses were randomly distributed throughout the duration of the experiment.*

Identifying spatiotemporal event boundaries

To examine how well the sighted and non-sighted groups identified the pre-established spatiotemporal event boundaries, we first created a null model (AIC = 4420, R^2 = 0.44), including only the dependent variable (event boundary hits) and the intercept (participants and event boundary items as random effects). To test the effect of experiencing the film without sight with AD with watching the film with sight without AD, the null model was then contrasted against models including fixed effects of group (non-sighted with AD, sighted without AD), and boundary type (spatial change, temporal change). The maximal model, including the fixed effects and their interactions (AIC = 4407, $R^2 = 0.44$), revealed a significant main effect of boundary type ($\chi^2(1) = 16.7$, $\beta = 1.63$, $SE = 0.40$, $z = 4.1$, $p < .001$). However, there were no effects of group ($\chi^2(1) = 2.52$, $\beta = 0.51$, $SE = 0.32$, $z = 1.59$, $p = .11$), or interaction between group and boundary type ($\chi^2(1) = 0.89$, $\beta = 0.24$, $SE = 0.26$, $z = 0.95$, $p = .34$). The best model fit, thus only included the fixed effect of boundary type. See Fig 2.

To analyze the timing of when the sighted and non-sighted groups responded in their correctly identified event boundaries (in relation to the onset of the pre-established event boundary), we first created a null model (AIC = 9639, $R^2 = 0.21$), including only the dependent variable (response latency) and the intercept (participants and event boundary items as random effects). To test the effect of experiencing the film, without sight with AD with watching the film with sight without AD, the null model was then contrasted against models including fixed effects of group (non-sighted with AD, sighted without AD), and boundary type (spatial change, temporal change). The maximal model, including the fixed effects and their interactions, (AIC = 9624, $R^2 = 0.22$), revealed a significant main effect of group ($F = 22.5$, $\beta = 1.12$, $SE = 0.24$, $t = 4.7$, $p < .001$). There were no effects of boundary type ($F = 0.68$, $\beta = 0.32$, $SE = 0.39$, $t = 0.83$, $p = .41$), or interaction between group and boundary type ($F = 0.59$, $\beta = 0.24$, $SE = 0.31$, $t = 0.77$, $p = .44$). The best model fit, thus only included the fixed effect of group. See Fig 2b.

To summarize, there were no significant differences in how well the non-sighted with AD and the sighted without AD identified the pre-established spatiotemporal event boundaries in their experience of the film narrative. However, the latency of the response (in respect to the pre-established event boundaries) was longer for the non-sighted with AD, independent of boundary type.

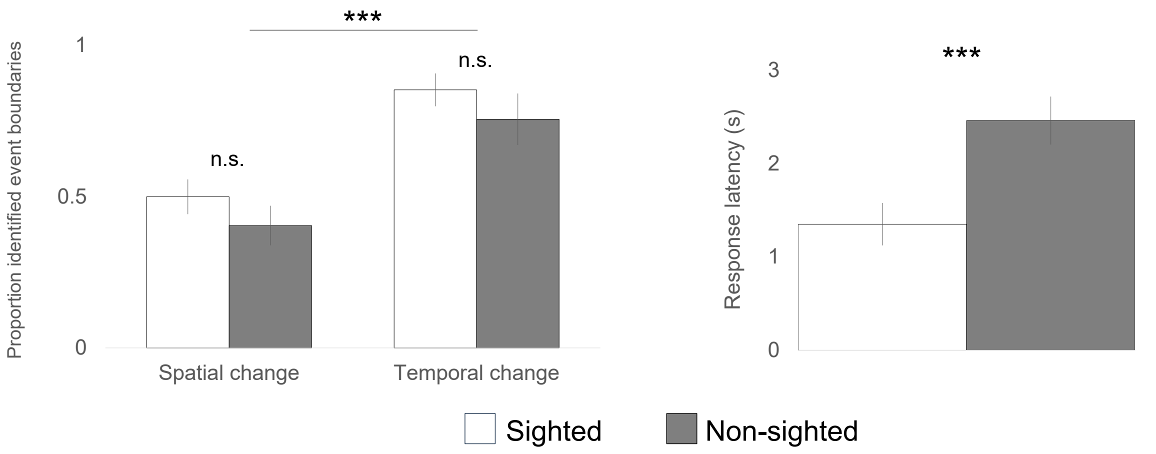

Fig. 2. *Event boundary identification across the sighted without AD and the non-sighted with AD. Error bars denote the standard errors of the mean. n.s., nonsignificant. ***p<.001.*

Explicit versus implicit event boundaries

To examine the impact of explicit versus implicit expressions of spatiotemporal event boundaries in AD, we examined how well the 20 event boundaries that were influenced by this manipulation were identified across the sighted and non-sighted groups. First, we created a null model (AIC = 911, R^2 = 0.48), including only the dependent variable (event boundary hits) and the intercept (participants and event boundary items as random effects). To test the effect of explicit vs implicit AD, the null model was then contrasted against models including fixed effects of group (non-sighted with explicit AD, non-sighted with implicit AD, sighted without AD) and boundary type (spatial change, temporal change). The best model fit, including the fixed effects (AIC = 904, R^2 = 0.48), revealed significant effects of group ($\chi^2(1) = 9.17$, p = .01), and boundary type ($\chi^2(1) = 4.24$, $\beta = 1.12$, $SE = 0.55$, $z = 2.1$, $p = .039$). Pair-wise group comparisons revealed significant differences between non-sighted with explicit and implicit AD ($\beta = 1.12$, $SE = 0.58$, $z = 2.0$, $p = .041$), and between sighted without AD and non-sighted with implicit AD ($\beta = 1.41$, $SE = 0.47$, $z = 3.0$, $p = .003$). No differences emerged between non-sighted with explicit AD and sighted without AD (β = 0.22, *SE* = 0.48, *z* = 0.46, *p* = .64). See Fig. 3.

When running the same analyses on the non-manipulated event boundaries, no significant differences across groups emerged (see supplementary materials). Thus, the reported effects are specific for the implicit versus explicit AD manipulation and not due to general differences in reporting event boundaries across the groups.

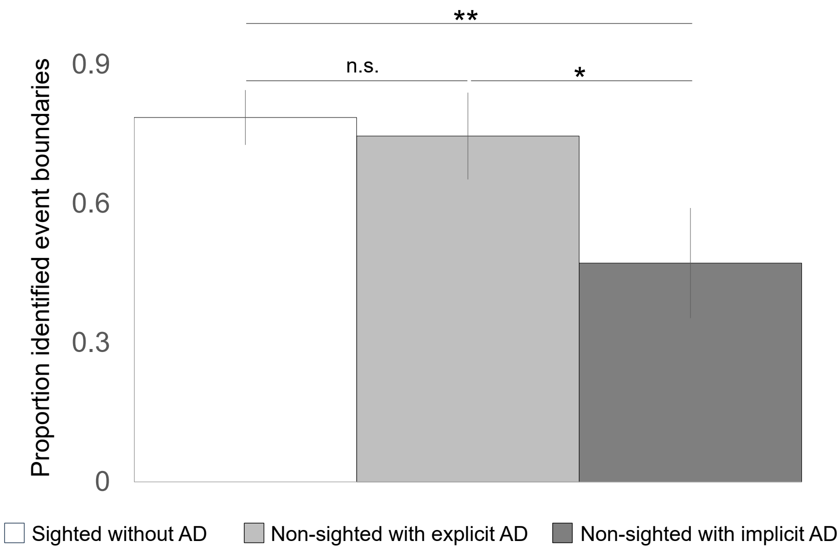

Fig. 3. *Identified event boundaries for explicit and implicit AD. Error bars denote the standard errors of the mean. n.s., nonsignificant. *p<.05, **p<.01.*

To analyze the timing of when the two groups correctly identified event boundaries (in relation to the onset of the pre-established event boundary), we first created a null model (AIC = 2653, R^2 = 0.28), including only the dependent variable (response latency) and the intercept (participants and event boundary items as random effects). The null model was then contrasted against models including fixed effects of group (non-sighted with explicit AD, non-sighted with implicit AD, sighted without AD) and boundary type (spatial change, temporal change). The best model fit, including only the fixed effect of group (AIC = 2646, $R^2 = 0.27$), revealed a significant main effect of group ($F = 6.4$, $p = .004$). Pair-wise comparisons revealed significant differences between sighted without AD and non-sighted with explicit AD ($\beta = 1.13$, $SE = 0.39$, $t = 2.9$, $p = .006$), and between sighted without AD and non-sighted with implicit AD ($\beta = 1.10$, $SE = 0.41$, $t = 2.7$, $p = .011$). No differences emerged between the non-sighted with explicit and implicit AD ($\beta = 0.03$, $SE = 0.50$, $t = 0.05$, $p = .96$). See Supplementary materials for corresponding analyses for the non-manipulated event boundaries.

Thus, the implicit AD did indeed decrease the probability of reporting spatiotemporal event boundaries, whereas the response performance for explicit AD was on par with the sighted control group. For identified event boundaries, there was, however, no difference in response latencies for explicit and implicit AD.

Discussion

The present study aimed to investigate how blind and sighted individuals perceive and understand event segmentation in audio-described films, with a specific focus on the impact of explicit and implicit AD on event boundary perception. Our findings contribute to the growing body of literature on event segmentation, multimodal processing, and accessibility for visually impaired individuals.

Our results suggest that both blind individuals with AD and sighted individuals without AD showed similar abilities to identify spatiotemporal event boundaries in the film. Despite differences in sensory input, participants in both groups demonstrated a comparable ability to perceive significant changes in the narrative, whether they relied on auditory and verbal cues (non-sighted with AD) or visual cues (sighted without AD). This discovery aligns with prior research demonstrating that event segmentation processes share common mechanisms, irrespective of sensory modality (Zacks et al., 2009; Radvansky & Zacks, 2017). For instance, Zacks et al. (2009) observed similar event segmentation performance across different presentation formats, such as listening to an oral narrative and reading the same story. Additionally, in their examination of audio dramas, Huff et al. (2018) reported comparable event segmentation abilities among sighted and visually impaired participants. In our present investigation, we extend these findings by illustrating that event segmentation in multimodal films, incorporating visual, verbal, and auditory cues (Holsanova, 2022; Remael & Reviers, 2018), is analogous between participants with intact visual perception and those with visual impairment who rely on AD. Consequently, the parallels in event segmentation appear to transcend variations in perceptual capabilities and presentation modes in these two contexts, underscoring the fundamental nature of the process involved in parsing meaningful information into distinct narrative events (cf. Radvansky & Zacks, 2017).

However, it is worth noting that blind participants with AD exhibited longer response latencies in identifying event boundaries compared to the sighted group. This discrepancy may be attributed to the nature of AD, where verbal descriptions of visual content are provided at a different pace than the real-time visual experience. The temporal misalignment between the AD and the film's actual events could account for the delayed responses among blind participants. This suggests that AD, while effective in conveying information, may still present challenges in terms of timing and synchronization for blind viewers.

Our study critically explored the impact of explicit versus implicit AD on event boundary perception. The results indicate that explicit AD, which explicitly described changes in spatial and temporal settings, improved event boundary perception for blind individuals compared to implicit AD. This finding suggests that providing clear and explicit descriptions of spatial and temporal changes in AD can enhance the understanding of event segmentation for blind viewers, making their experience more comparable to that of sighted individuals.

In contrast, implicit AD, which relied on indirect cues to convey changes in settings, led to a reduced ability to perceive event boundaries compared to both explicit AD and the sighted control group. This result highlights the importance of providing explicit and informative AD to ensure that blind viewers can fully grasp the spatiotemporal aspects of a film's narrative.

Our findings carry significant implications for the fields of audio description practices and the training of audio describers, and they hold particular relevance for the visually impaired end users (cf. Holsanova, 2016, 2022; Lopez et al., 2022). An essential facet of enhancing the accessibility and inclusivity of cinematic experiences is gaining a deeper understanding of how individuals with visual impairments process and segment the narrative sequence of events when assisted by audio description. Our current set of results underscores the critical need for audio describers to meticulously consider their articulation of event boundaries.

When producing AD, a pivotal consideration lies in striking a balance between the provision of an objective, impartial portrayal of visual content to accurately represent a specific moment and the inclusion of more subjective descriptions that encompass the interpretation of key narrative elements (cf., Rai et al., 2010; Remael, 2015). Our investigation offers substantiation that the elucidation of event boundaries, often necessitating a measure of interpretation, holds a recurring necessity for AD recipients. This elucidation serves as a foundational underpinning, facilitating the potential for recipients to engage in less biased encounters with the visual content contingent upon these delineations.

The process of event segmentation is equally essential for automated computer-generated video description (Braun, Starr, & Laaksonen, 2020; Starr, Braun, & Delfani, 2020). This necessitates training algorithms to identify actions within dynamic scenes, discern connections between frames and actions, recognize event boundaries, and effectively express them. A proper comprehension of how visually impaired end users perceive event boundaries in audio description can significantly contribute to the development of algorithms capable of accurately tracking referents across narrative depictions of time and space. This, in turn, may offer valuable assistance in establishing nominal and pronominal continuity and coherence, thereby addressing some of the key challenges inherent in the performance of such algorithms.

Moreover, our findings align with the principles of the UN Convention on the Rights of Persons with Disabilities (CRPD), emphasizing the importance of making content accessible to all (United Nations, 2006). Accessible media, including AD, plays a vital role in ensuring that individuals with visual impairments can enjoy cultural and entertainment content on an equal footing with their sighted counterparts.

Limitations and future studies

The present study utilized a substantial portion of a film (43 minutes of "Skumtimmen") as stimuli, chosen for its distinct filmic spatiotemporal cuts. While this choice aligns with prior research (cf. Huff et al., 2018; Zacks et al., 2009) and offers the advantage of presenting a continuous stream of input with high ecological validity for a cinematic experience, it raises concerns about experimental control, especially in the context of explicit versus implicit manipulation. Since we selected 20 salient spatiotemporal boundaries within the film's narrative for our experimental manipulation, these instances remain dependent on the overall narrative, potentially diminishing the impact of the manipulation. A more controlled investigation employing multiple film clips with explicit and implicit audio description may be needed to further clarify this aspect. Nevertheless, it is worth noting that despite this potential limitation, the experimental manipulation successfully altered event segmentation identification as anticipated.

However, it is important to acknowledge that the film used in the study adheres to a typical narrative structure for a mystery thriller. Thus, the generalizability of the results to other narratives, film genres, and contexts remains uncertain. Furthermore, both the film and audio description were presented in Swedish, a language with particular linguistic resources for expressing spatiotemporal event boundaries (cf., Holsanova et al., 2023). To gain a more comprehensive understanding of event segmentation in audio description, it would be beneficial to conduct studies in various languages with diverse film materials.

In addition, the present study concentrated on salient spatiotemporal event boundaries. Future investigations should explore other aspects of event segmentation, such as character-driven event boundaries and different hierarchical levels of event segmentation, encompassing both coarse- and fine-grained events (cf., Huff et al., 2018; Zacks et al., 2009). Moreover, our study focused on the perception of event segmentation, which previous research has associated with narrative comprehension and memory (Sargent et al., 2013; Zacks, et al., 2006). However, as we did not directly assess comprehension, we lack direct evidence regarding participants' understanding of the narrative or whether the experimental manipulation influenced this understanding. Hence, future studies should investigate the specific impact of audio description within the context of event segmentation on comprehension.

Also, it is crucial to recognize that the visually impaired population is highly diverse, encompassing individuals with varying onset, degree, and reasons for impairment, contributing to substantial cognitive variability (e.g., Cattaneo & Vecchi, 2011). This heterogeneity may influence event segmentation capabilities. Therefore, future studies should delve deeper into specific visual impairments, including the examination of how audio description influences neural activity among congenitally blind individuals and those who have experienced late-onset visual impairment.

Finally, the present study employed video-call technology via the Zoom platform for the data collection. During this procedure, the film and audio material was synchronized and presented via the ELAN software through screen sharing. Consequently, it is important to acknowledge that some degree of experimental control was compromised due to inherent variations in participants' computer configurations. Furthermore, participant responses were recorded by the experiment leader within the ELAN software in tandem with their oral responses. This concurrent data capture method introduced an element of increased noise pertaining to response latencies and posed an augmented risk of potential response omissions. Nonetheless, it is noteworthy that these methodological adaptations were necessitated by the prevailing circumstances of the Covid-19 pandemic. The logistics and health considerations surrounding the pandemic rendered physical in-person participation infeasible, particularly for visually impaired participants located in various regions of Sweden, who would have incurred significant travel demands. Notwithstanding these challenges, we find it gratifying that the implemented experimental procedure yielded commendable results. It is important to emphasize that the anticipated experimental noise introduced by the procedure exhibits no discernible disparity between sighted and visually impaired participants. Consequently, it should not be deemed a pivotal factor influencing our main results.

Conclusion

Our study contributes valuable insights into event segmentation, multimodal processing, and accessibility for visually impaired individuals. Explicit AD was found to enhance the perception of event boundaries, making the viewing experience more comparable to that of sighted individuals. These findings have practical implications for the production of AD and highlight the ongoing importance of improving media accessibility to promote inclusivity and equal access to cultural and cinematic experiences for all individuals. Further research in this area can continue to advance our understanding of the complex interplay between sensory modalities and event cognition in the context of multimedia experiences.

Acknowledgments

This work was supported by a grant from FORTE 2018-00200 (Swedish Research Council for Health, Working Life and Welfare) and by TSI 2019 Lund University grant "Audio Description and Accessible Information" (2021-2025). We gratefully acknowledge Lund University Humanities Laboratory, Tina Weidelt for her thorough and professional work with providing and narrating the audio descriptions, and Linn Petersdotter for her persistent and careful work with data collection.

References

Alfredson, D. (Director). (2013). *Skumtimmen* [Echoes from the dead]. Yellow Bird Films AB, Fundament Film AB.

Amoruso, L., Gelormini, C., Aboitiz, F., Alvarez Gonzalez, M., Manes, F., Cardona, J. F., & Ibanez, A. (2013). N400 ERPs for actions: building meaning in context. *Frontiers in human neuroscience*, *7*, 57.

Baldassano, C., Hasson, U., & Norman, K. A. (2018). Representation of real-world event schemas during narrative perception. *Journal of Neuroscience*, *38*(45), 9689-9699.

Braun, S., Starr, K. & Laaksonen, J. (2020). Comparing human and automated approaches to visual story-telling. In Braun, S. & Starr, K. (Eds.), *Innovations in audio description research*. London: Routledge. 1–12.

Cutting, J., & Iricinschi, C. (2015). Re-presentations of space in Hollywood movies: An event-indexing analysis. *Cognitive science, 39* 434-456.

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological science*, *22*(2), 243-252.

Gallucci, M. (2019). GAMLj: General analyses for linear models. [jamovi module]. Retrieved from https://gamlj.github.io/.

Gerwien, J., & von Stutterheim, C. (2018). Event segmentation: Cross-linguistic differences in verbal and non-verbal tasks. *Cognition*, *180*, 225-237.

Holsanova, J. (2022). The Cognitive Perspective on Audio Description: Production and reception processes. In C. Taylor & E. Perego (Eds.), *The Routledge Handbook of Audio Description* (pp. 57-77). Taylor & Francis.

Holsanova, J. (2016). Cognitive approach to audio description. In: Matamala & Orero (eds.), *Researching audio description: new approaches*. Palgrave Macmillan.

Holsanova, J., Blomberg, J., Blomberg, F., Gärdenfors, P., & Johansson, R. (2023). Event segmentation in the audio description of films: a case study. *Journal of Audiovisual Translation.*

Huff, M., Maurer, A. E., Brich, I., Pagenkopf, A., Wickelmaier, F., & Papenmeier, F. (2018). Construction and updating of event models in auditory event processing. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(2), 307–320. https://doi.org/10.1037/xlm0000482

Huff, M., Meitz, T. G., & Papenmeier, F. (2014). Changes in situation models modulate processes of event perception in audiovisual narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1377.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in cognitive sciences*, *12*(2), 72-79.

Kurby, C. A., & Zacks, J. M. (2011). Age differences in the perception of hierarchical structure in events. *Memory & cognition*, *39*(1), 75-91.

Lopez, M., Kearney, G., & Hofstädter, K. (2022). Seeing films through sound: Sound design, spatial audio, and accessibility for visually impaired audiences. *British Journal of Visual Impairment*, *40*(2), 117–144. https://doi.org/10.1177/0264619620935935

Lyberg-Åhlander, V., Holsanova, J., & Johansson, R. (2023). *The Influence of Narrative Specificity and Voice Quality when Listening to Audio Descriptions: a Comparison of the Sighted and the Blind*. PsyArXiv.

Magliano, J. P. & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive Science*, 35 (6), pp. 1489-1517.

Newtson D (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 28(1), 28–38.

Papenmeier, F. (2016). *Package "segmag": Determine Event Boundaries in Event Segmentation Experiments*. https://cran.r-project.org/web/packages/segmag/index.html

Pettijohn, K. A., Thompson, A. N., Tamplin, A. K., Krawietz, S. A., & Radvansky, G. A. (2016). Event boundaries and memory improvement. *Cognition*, *148*, 136-144.

Radvansky, G. A., & Zacks, J. M. (2014). *Event cognition*. Oxford University Press.

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. In *Current Opinion in Behavioral Sciences* (Vol. 17, pp. 133–140). Elsevier Ltd. https://doi.org/10.1016/j.cobeha.2017.08.006

Rai, S., Greening, J., & Petré, L. (2010). A comparative study of audio description guidelines prevalent in different countries. *Londra: Royal National Institute of Blind People*.

Remael, A., Reviers, N. & Vercauteren, G. (eds.) (2015). *Pictures painted in words: ADLAB Audio Description Guidelines*, EUT Edizioni Università DiTrieste, Trieste.

Remael, A., & Reviers, N. (2018). Multimodality and Audiovisual Translation: Cohesion in Accessible Films. In L. Pérez-González (Ed.), The Routledge handbook of audiovisual translation (pp. 260–280). London: Routledge.

Remael, A. & Vercauteren, G. (2015). Spatio-temporal settings. In: *Pictures painted in words: ADLAB Audio Description Guidelines*, EUT Edizioni Università DiTrieste, Trieste, Section 2.1.2, 24–27.

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., Eisenberg, M. L., & Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241-255.

Santin, M., Van Hout, A., & Flecken, M. (2021). Event endings in memory and language. *Language, Cognition and Neuroscience*, *36*(5), 625-648.

Sasmita, K., & Swallow, K. M. (2022). Measuring event segmentation: An investigation into the stability of event boundary agreement across groups. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01832-5

Sloetjes, H., & Seibert, O. (2016). Measuring by marking: The multimedia annotation tool ELAN. In A. Spink, G. Riedel, L. Zhou, L. Teekens, R. Albatal, & C. Gurrin (Eds.), *Proceedings of measuring behavior 2016* (pp. 492–495). Dublin City University.

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of memory and language*, *53*(1), 125-140.

Speer, N. K., Swallow, K. M., & Zacks, J. M. (2003). Activation of human motion processing areas during event perception. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(4), 335-345.

Stawarczyk, D., Wahlheim, C. N., Etzel, J. A., Snyder, A. Z., & Zacks, J. M. (2020). Aging and the encoding of changes in events: The role of neural activity pattern reinstatement. *Proceedings of the National Academy of Sciences*, *117*(47), 29346-29353.

United Nations. (2006). *Convention on the Rights of Persons with Disabilities and Optional Protocol UNITED NATIONS*. https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html

The jamovi project (2019). Jamovi. (Version 1.6.23) [Computer Software]. Retrieved from https://www.jamovi.org.

Vandaele, J. (2012). What meets the eye. Cognitive narratology for audio description. *Perspectives*, *20*(1), 87-102.

Vercauteren, G. (2012). A narratological approach to content selection in audio description. Towards a strategy for the description of narratological time. *MonTI* 4 (2012: 207-231).

Vercauteren, G. (2016). A translational and narratological approach to audio describing narrative characters. *TTR*, XXVII (2), 71–90.

Vercauteren, G. (2021). Insights from Mental Model Theory and Cognitive Narratology as a Tool for Content Selection in Audio Description. *Journal of Audiovisual Translation, 4*(3), 6–24.

Vercauteren, G., & Remael, A. (2015). Spatio-temporal settings. In A. Maszerowska, A. Matamala, & P. Orero (Eds.), *Audio description. New perspectives illustrated* (Vol. 112, pp. 61–80). John Benjamins Publishing Company.

Walczak, A., & Fryer, L. (2017). Creative description: The impact of audio description style on presence in visually impaired audiences. *British Journal of Visual Impairment*, *35*(1), 6–17. https://doi.org/10.1177/0264619616661603

Zacks, J. M. (2013). Constructing Event Representations During Film Comprehension. In *Psychocinematics: Exploring Cognition at the Movies*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199862139.001.0001

Zacks, J. M. (2013). Constructing event representations during film comprehension. In A. P. Shimamura (Ed.), *Psychocinematics: Exploring Cognition at the Movies* (pp. 227-243). Oxford University Press.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature neuroscience*, *4*(6), 651-655.

Zacks, J. M. & Magliano, J. P. (2011). Film, Narrative, and Cognitive Neuroscience. In: *Art & the Senses*, ed. D. P. Melcher and F. Bacci. New York: Oxford University Press.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, *138*(2), 307.

Zacks. J. M., Speer N. K., Swallow K.M., Maley C.J. (2010). The brain's cutting-room floor: Segmentation of narrative cinema. *Frontiers in Human Neuroscience*, 4(168): 1–15.

Zacks, J. M., Speer, N. K., Vettel, J. M., & Jacoby, L. L. (2006). Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychology and aging*, *21*(3), 466.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29–58. https://doi.org/10.1037/0096-3445.130.1.29

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological science*, *6*(5), 292-297.

Zwaan, R. A. & Radvansky, G. A. (1998). Situation Models in Language Comprehension and memory. *Psychological Bulletin*, 123(2), 162–185.

Zwaan, R. A., Magliano, J. P. and Graesser, A. C. (1995) ‘Dimensions of Situation-Model Construction in Narrative Comprehension’, *Journal of Experimental Psychology: Learning Memory and Cognition,* 21, 386–397.