Huff, M., Maurer, A. E., Brich, I., Pagenkopf, A., Wickelmaier, F., & Papenmeier, F. (2017). Construction and updating of event models in auditory event processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: 10.1037/xlm0000482

© 2017, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors permission. The final article will be available, upon publication, via its DOI: 10.1037/xlm0000482

Construction and Updating of Event Models in Auditory Event Processing

Markus Huff¹, Annika E. Maurer¹, Irina Brich², Anne Pagenkopf¹, Florian Wickelmaier¹ & Frank
Papenmeier¹

1: Department of Psychology, University of Tübingen, Germany

2: Leibniz-Institut für Wissensmedien, Tübingen

Correspondence concerning this article should be addressed to

Markus Huff, University of Tübingen, Department of Psychology, Schleichstr. 4, D-72076 Tübingen, Germany. E-mail: markus.huff@uni-tuebingen.de

Word count: 8003 words (Abstract: 213 words)

Abstract

Humans segment the continuous stream of sensory information into distinct events at points of change. Between two events, humans perceive an event boundary. Present theories propose changes in the sensory information to trigger updating processes of the present event model. Increased encoding effort finally leads to a memory benefit at event boundaries. Evidence from reading time studies (increased reading times with increasing amount of change) suggest that updating of event models is incremental. We present results from five experiments that studied event processing (including memory formation processes and reading times) using an audio drama as well as a transcript thereof as stimulus material. Experiments 1a and 1b replicated the event boundary advantage effect for memory. In contrast to recent evidence from studies using visual stimulus material, Experiments 2a and 2b found no support for incremental updating with normally sighted and blind participants for recognition memory. In Experiment 3, we replicated Experiment 2a using a written transcript of the audio drama as stimulus material allowing us to disentangle encoding and retrieval processes. Our results indicate incremental updating processes at encoding (as measured with reading times). At the same time, we again found recognition performance to be unaffected by the amount of change. We discuss these findings in the light of current event cognition theories.

Keywords: event cognition, auditory events, blind participants, reading times

Construction and Updating of Event Models in Auditory Event Processing

When reliving events from the past, humans can easily recreate some aspects of their experiences while other aspects remain harder to retrieve. In this context, it is important to note that unlike continuous environmental information, human memory is organized into distinct events (DuBrow & Davachi, 2013): When observing an ongoing activity, humans automatically segment the continuous stream of information into distinct *events*, which are divided by *event boundaries*. Memory for event boundaries is superior to memory for non-boundaries (Newtson & Engquist, 1976) and increases with increasing amount of change at event boundaries (Huff, Meitz, & Papenmeier, 2014). Elaborated processing of information is thought to be causal for these effects (Zacks, Speer, Swallow, Braver, & Reynolds, 2007). We studied the consequences of event boundary processing using an audio drama and a written transcript thereof. More specific, we tested recognition memory and measured reading times, thus allowing us to describe the perceptual and cognitive processes that are related with the processing of event boundaries in close detail (Radvansky & Copeland, 2010).

Consequences of Event Segmentation

Humans segment the continuous stream of information into distinct events (Newtson, 1973). Between two meaningful units (events) participants perceive an event boundary. Event segmentation and its consequences for perceptual, cognitive, and neuronal processing is studied in different research domains (natural event perception, Kurby & Zacks 2008, and narrative comprehension, Magliano & Zacks, 2011). With regard to natural event perception, Event Segmentation Theory (EST) (Zacks et al., 2007) proposes that event boundaries are perceived when predictions about the future development of the observed activity fail. Elaborated processing of information at event boundaries should result in higher long-term memory performance as compared to less elaborated information processing at non-event boundaries. On the other side, for narrative processing, Event Indexing (EI) describes how new semantic information is integrated within the current situation model (Zwaan, Langston, & Graesser, 1995).

Both research domains converge on the idea that an *event model* in working memory represents the current observation. In the narrative context, the event model describes the current situation – as described in a text or depicted in a movie – using several dimensions such as *character*, *time*, *location*, and *action* (Magliano & Zacks, 2011). A change in one or more dimension triggers updating processes – a new event model is constructed in the working memory that is in agreement with the new situation. The *updating* of such an event model is perceived as an event boundary (Magliano, Miller, & Zwaan, 2001; Speer & Zacks, 2005).

As measurements of neuronal activation (i.e. functional magnetic resonance imaging, fMRI, and electroencephalography, EEG) at event boundaries show, event segmentation is an automatic process. This was consistently found regardless of whether participants watched a video (Zacks et al., 2001; Zacks, Swallow, Vettel, & McAvoy, 2006), read a text (Swallow et al., 2010), listened to a narrative (Whitney et al., 2009) or to a music composition (Sridharan, Levitin, Chafe, Berger, & Menon, 2007). Behavioral measurements of event boundary processing, such as reading times, demonstrated that readers slow down at event boundaries (Rinck & Weber, 2003; Zacks, Speer, & Reynolds, 2009; Zwaan, Magliano, & Graesser, 1995; Zwaan, Radvansky, Hilliard, & Curiel, 1998).

The Event Boundary Advantage for Memory

Research studying memory for dynamic events shows that information present at event boundaries is encoded and retrieved more accurately than information present at non-boundaries (Newton & Engquist, 1976). Similarly, text comprehension research show that the working memory is “reset” after having perceived an event boundary. As a result, information that is present right before an event boundary is harder to remember than information from a current event (Speer & Zacks, 2005). Furthermore, event boundaries have a stronger connection to long-term memory than non-boundaries (Zwaan, Magliano, et al., 1995). Similar results were found with visual stimuli. In an experiment that tested memory for specific items shown in video clips manipulated whether the items were present at an event boundary, and whether there was event boundary between item disappearance and test (Swallow, Zacks, & Abrams, 2009). If asked to remember a specific item only seconds after it

disappeared from a video clip, memory performance for items that were presented at an event boundary and items that were tested within the same event (i.e. without event boundary between object disappearance and test) was highest. In contrast, if there was an event boundary between object disappearance and test, memory performance was significantly lower.

Thus, there is strong evidence that the transitions between old and new event models have implications for memory formation and that the content of working memory is modified at event boundaries. Moreover, these studies suggest that event boundary perception and maybe also further memory formation processes are modality independent (Zacks et al., 2009).

The Nature of the Updating Process at Event Boundaries

Whether event model updating is incremental – the consequences of change processing is additive as suggested by the additivity hypothesis of EI (Zwaan, Langston, et al., 1995), global – all information is updated (Zacks et al., 2009), or a combination thereof (Kurby & Zacks, 2012) is still a matter of debate. Evidence for incremental updating processes comes from studies that have demonstrated that the amount of change occurring at an event boundary influences the perception of the event boundary. The probability of perceiving an event boundary increases linearly with an increasing number of changing dimensions (Huff et al., 2014; Zacks et al., 2009). In addition, the neuronal activity within the relevant areas depends on the number of dimension changes in a positive linear manner (Speer, Zacks, & Reynolds, 2007). In the fields of discourse and text comprehension, researchers found further a mediating effect of the number of dimension changes on reading times. Event boundaries were read more slowly when the number of dimension changes increased (Zacks et al., 2009; Zwaan, Langston, et al., 1995; Zwaan & Radvansky, 1998). Interpreting these results, the authors concluded that the number of dimension changes is related to a higher processing effort. Although EI does not make specific predictions with regard to long-term memory formation, recent research using audiovisual narratives has shown that visual recognition performance increases linearly with an increasing number of dimension changes at an event boundary (*incremental updating effect*; Huff et al., 2014). These results can be interpreted in the light of Radvansky's event horizon model

(Radvansky, 2012). Like EST, the event horizon model (EHM) is based on the idea that humans represent current observations in the form of events in working memory. Importantly, EHM distinguishes between memory for a specific item and memory for a specific event. According to EHM, successful retrieval of a particular item is more likely if it is represented in more than one memory trace (i.e. the item was part of multiple event models). In contrast, if a specific item is part of multiple event models stored in memory and the task is to remember a specific event, retrieval interference impairs memory performance. Yet, EHM is silent with regard to the influence of distinctiveness (Hunt, 1995) of events on memory performance. Distinctiveness of an event increases with increasing number of changes at the adjacent boundaries. Thus, an increasing number of dimension changes might decrease competition during retrieval.

Yet, when it comes to describe the nature of the updating process it is essential to consider that a single measure (e.g., memory) might be insufficient (Radvansky & Copeland, 2010). In their study, Radvansky and Copeland (2010) examined the sensitivity of memory and reading time measures while perceiving and processing changes such as narrative shifts. Importantly, changes in the spatial dimension did not affect reading times but had a significant impact on memory probes. Whereas changes in the temporal dimension did affect reading times.

Visual, Audiovisual, and Auditory Event Processing

The reviewed empirical evidence so far is mainly based on studies that have used written texts or (audio-)visual dynamic scenes as stimulus material. Only a few studies have investigated event perception processes using auditory stimuli. The event segmentation process as a basic concept is comparable across modalities. As an example, the perception of an auditory narrative is also parsed into distinct events (Huff, Maurer, & Papenmeier, revision under review) and event boundaries in an auditory narrative evoke higher neuronal activity in the corresponding areas compared to the activation at non-boundaries (Whitney et al., 2009). Yet, to our knowledge, there are no studies testing memory effects of auditory event processing.

Some important differences between visual and auditory processing prevent a direct transfer of findings from research with visual stimuli to the processing of auditory stimuli. Memory capacity, for instance, is much higher in visual processing than in auditory processing. Visual memory is able to recognize 83% out of 10,000 pictures correctly (Standing, 1973), which is far superior to verbal memory (Standing, Conezio, & Haber, 1970). Moreover, the visual system has a massive memory capacity for objects, object details (Brady, Konkle, Alvarez, & Oliva, 2008), and dynamic audiovisual scenes (Meyerhoff & Huff, 2015). Even if participants are well trained in identifying auditory stimuli (e.g., musicians), they are still not able to recognize sounds to the extent it was found with visual stimuli (M. A. Cohen, Evans, Horowitz, & Wolfe, 2011). In conclusion, visual memory is clearly superior to auditory memory. Thus, it is an important research question whether the basic principles of event processing also hold for auditory narratives.

The Role of Expertise in Dynamic Event Processing

Does event perception change with prior knowledge in a specific domain or expertise in processing specific information? While some studies have shown that the influence of prior knowledge is quite subtle (Hard, Tversky, & Lang, 2006; Huff et al., 2017, revision under review), Graziano, Moore, and Collins (1988), demonstrated that children provided with relevant information right before the segmentation task segmented at a larger grain (Graziano et al., 1988). Importantly, persons with intellectual disability showed a different, less hierarchical event segmentation pattern as compared to a healthy student sample (Sebastian, Ghose, Zacks, & Huff, 2017). Given that most persons have lots of experience in audiovisual information processing – Germans spent almost 4 hours a day watching TV (“AGF Videoforschung,” 2017) – we cannot preclude that the capability to process auditory narratives is only little pronounced in the typical student sample.

If this is true, information processing in people, who have no experience with audiovisual information processing (such as blind people), should be different. In fact, there is evidence that blind people show an enhanced processing and memory of auditory stimuli compared to people with normal vision (Amedi, Raz, Pianka, Malach, & Zohary, 2003; Perleth & Effinger, 2001). Compared to normally

sighted people, blind people have an enhanced voice recognition and memory for voices (Bull, Rathborn, & Clifford, 1983; Röder & Neville, 2003). In addition, blind people are also faster in processing auditory cues, from simple tones (Röder, Rösler, Hennighausen, & Näcker, 1996) to complex sentences (Röder, Rösler, & Neville, 2000). Reorganization of neuronal occipital structures might be accountable for these differences (Bavelier & Neville, 2002; L. G. Cohen et al., 1997). Taken together, there are profound differences in the processing of auditory stimuli in normally sighted and blind people. Thus, dynamic event processing might also be different for each of these two groups.

Experimental Overview and Hypotheses

In the following, we will report five experiments that tested cognitive and perceptual consequences of event boundary perception and event model updating using auditory and textual stimulus material. Participants encoded a narrative (either an audio drama or a written transcript thereof) and completed a recognition test. Experiment 1a tested the event boundary advantage effect for auditory narratives. We expected event boundaries to be better recognized than non-boundaries (Newtson & Engquist, 1976). In Experiment 1b, we excluded the alternative that the event boundary advantage effect is due to systematic differences in item difficulty. Experiment 2a tested the incremental updating hypothesis (Huff et al., 2014) with normally sighted participants using an auditory narrative. If memory formation processes are comparable for audiovisual and auditory stimuli, we expect memory formation to increase linearly with an increasing amount of change. However, because of the specific differences between visual and auditory information processing (M. A. Cohen, Horowitz, & Wolfe, 2009) it might also be possible that recognition performance is not related to the amount of changes. Experiment 2b then tested the incremental updating hypothesis with blind participants. If the differences between normally sighted and blind people with regard to auditory information processing influence dynamic auditory information processes, we expect a different result pattern as compared to Experiment 2a. If, however, dynamic auditory information processing is a general perceptual-cognitive process that is independent of participants' experience with (audio-)visual and auditory information, we expect results to be comparable to Experiment 2a. In Experiment 3, finally, we disentangled encoding and retrieval

processes (Radvansky & Copeland, 2010). Using the written transcript of the audio drama as stimulus material we were able to measure reading times (as a measure for encoding processes) as well as recognition performance (as a measure for retrieval processes) within a single experiment. Participants read the narrative in a self-paced manner and completed a textual recognition test. According to the additivity hypothesis of EI (Zwaan, Langston, et al., 1995), we expect increasing reading times with increasing number of dimension changes.

Experiment 1a

Experiment 1a tested the event boundary advantage effect for auditory stimuli.

Method

Participants and design. All experiments were approved by the local ethics committee of the Leibniz-Institut für Wissensmedien, Tübingen (LEK 2015/032). Thirty-one normally sighted students of the University of Tübingen participated in this experiment for course credit or monetary compensation. We excluded three participants because they knew the presented audio drama episode. Thus, we analyzed the data of 28 participants (6 males), aged 19 to 44 years ($M = 25.00$, $SD = 7.51$) with normal hearing ability. Their first language was German. An a priori power analysis revealed that a sample size of at least 22 participants was needed to reach a power of .99 to detect an effect with the size of $d = 1.00$ (Newtson & Engquist, 1976).

We manipulated *event boundary* (event boundary vs. non-event boundary, within-subjects) as independent variable and measured participants' auditory recognition performance. In addition, we measured participants' experience with auditory narratives (i.e. audio dramas and audio books). We refer to this measure as *expertise*.

Stimulus material. As stimulus material we used the same edited episode of the German audio drama "The Three Investigators and the Secret of Terror Castle" (Körting, 2001) as in Huff et al. (revision under review), which is based on the same-named first title of the US book series "The Three

Investigators” (Arthur & Hitchcock, 1964). The episode lasted approximately 44 minutes. In the event segmentation study of Huff et al. (revision under review), native (persons who spoke the language of the audio drama) and non-native participants (persons who did not understand the language of the audio drama) segmented the audio-drama in two sessions into coarse- and fine-grained meaningful events. Based on the segmentation data of the native participants, we determined significant event boundaries using the segmag package for R (Papenmeier, 2014). For the auditory recognition test, we extracted 46 audio-clips with a duration of 4.5 seconds, 23 containing an event boundary as well as 23 clips without boundaries. Because there is no auditory equivalent to video stills (as used by Huff et al., 2014), we used short dynamic clips covering the whole event boundary. For each of the target items, a distractor matching the overall content of the episode and the features of the respective target item was extracted from different episodes of the same series. As an example, in case a target item contained a ringing phone, the corresponding distractor item also contained a ringing phone (see Table 1 for examples).

Expertise questionnaire. We measured expertise with audio dramas with five questions. The first two questions addressed general audio drama experience for the past and present (“Do you listen to audio dramas in general?” 0: never, 1: rarely, 2: often, 3: regularly). Questions three and four addressed audio drama experience with the “The Three Investigators” series for the past and present (“Have you listened to audio dramas of ‘The Three Investigators’ series?” 0: never, 1: rarely, 2: often, 3: regularly). Points for the “present” questions were doubled. Finally, participants indicated how many different episodes of the “The Three Investigator” series they have listened to. For this latter score, we calculated quartiles and awarded the points from 0 to 3.

The expertise score was calculated by first summing up the scores. The final expertise score ranged from 0 (*no expertise*) to 21 (*high expertise*). Item statistics (Appendix A) and reliability of the scale, $\alpha = .81$, were calculated from the whole sample of Experiment 1a ($N = 31$). In Experiment 1a, the expertise scale ranged from 0 to 15 ($M = 7.29$, $SD = 3.88$), excluding participants knowing the presented episode.

Technical implementation. The experiment was programmed in PsychoPy2 (Peirce, 2007; 2008). Audio files were presented via headphones and participants adjusted the volume to a comfortable level at the beginning of the experiment.

Procedure. The experiment was conducted in individual sessions. Upon arriving, participants signed informed consent forms and were seated in front of a notebook computer. They were instructed to carefully listen to the audio drama and that this learning phase would be followed by a recognition test. After the learning phase, participants completed a distraction task (paper folding test, PFT) (Ekstrom, French, Harman, & Dermen, 1976). The distraction task lasted about ten minutes. Before the recognition test, demographic data was collected and participants received the instructions for the recognition test.

Target and distractor items were presented in random order. After each clip, participants indicated whether they recognized it by pressing either the “Yes” or the “No” key on a keyboard (assignment of keys was counterbalanced across participants) and their respective confidence on a 5-point scale: *no idea*, *not at all confident*, *moderately confident*, *fairly confident*, *very confident* (Perfect, Watson, & Wagstaff, 1993). After the test phase (duration 12 minutes), they were handed the expertise questionnaire. The experiment lasted a total of approximately 75 minutes.

Results and Discussion

We calculated sensitivity d_a (Verde & Rotello, 2003), that considers unequal familiarity distribution variances between hit and false-alarm distribution. In doing so, we combined participants’ yes/no responses and confidence (1 to 5) ratings. This resulted in a continuum with 1 to 5 as “old” and 6 to 10 as “new” answers. These ratings were submitted to maximum-likelihood estimation of the ROC curve for each participant, allowing the computation of d_a (Macmillan & Creelman, 2004). The average ROC curves are presented in Figure 1A. Sensitivity (d_a) was submitted to a two-way ANOVA with event boundary (event boundary vs. non-event boundary, within-subjects) and expertise (as measured with the expertise questionnaire, between-subjects) as independent variables. The analysis revealed a

main effect for event boundary, $F(1, 26) = 5.53, p = .027, \eta_p^2 = .18$, indicating that memory for event boundaries was better as compared to memory for non-event boundaries. We did not observe a main effect for expertise, $F(1, 26) < 1$, and no interaction between both factors, $F(1, 26) = 1.23, p = .277, \eta_p^2 = .05$. Taken together, this experiment extended the event boundary advantage effect to auditory narratives (Newton & Engquist, 1976). Further, this process is independent of participants' expertise (i.e. experience with audio dramas).

Experiment 1b

As an alternative, the results of Experiment 1a might also be explained with item difficulty. As compared to non-event boundary items, event boundary items might have been easier to encode and to recognize, consequently. Experiment 1b tested this assumption by asking participants to memorize the target items of the recognition test in random order (i.e. without the narrative context). Subsequently, we tested participants' memory with the same recognition test as in Experiment 1a. If item difficulty is causal for the event boundary effect of Experiment 1a, we expect a similar result pattern. However, if it is not item difficulty, we expect a different result pattern.

Method

Participants and design. Thirty-six students took part for course credit. All participants reported normal hearing abilities. We excluded five participants because they already knew audio drama. Thus, we analyzed data of 31 participants (7 males) ranging from 19 to 36 years ($M = 23.32, SD = 3.89$). The design was similar to Experiment 1a.

Stimulus material. In the learning phase, participants were presented with the target items of the recognition test of Experiment 1a. The recognition test was the same as used in Experiment 1a.

Expertise questionnaire. Expertise was measured as described in Experiment 1a. Item statistics are reported in Appendix A, the reliability of the expertise scale was at $\alpha = .77$. The expertise scale ranged from 0 to 17 ($M = 5.34, SD = 3.88$), excluding participants knowing the relevant episode.

Procedure. Procedure was similar as in Experiment 1a with the following exceptions. Instead of listening to the audio drama, participants memorized only the target items of the recognition test during the learning phase. Items were presented in random order. Experiment 1b lasted approximately 30 minutes.

Results and Discussion

Data analysis was similar to Experiment 1a. See Figure 1B for the descriptive data. In contrast to Experiment 1a, recognition performance was higher for non-boundary as compared to boundary items, $F(1, 29) = 17.62, p < .001, \eta_p^2 = .38$. Neither the main effect of expertise, $F(1, 29) < 1$, nor the interaction of both factors reached significance, $F(1, 29) < 1$. Taken together, in Experiment 1b we did not find an event boundary advantage effect as in Experiment 1a. Thus, and most important, the event boundary advantage effect cannot be explained by the fact that event boundary items are easier to learn than non-event boundary items. We thus conclude that the semantic structure of the narrative is a necessary prerequisite for the event boundary advantage effect.

Comparing Performance of Experiment 1a and Experiment 1b

To explore how the narrative context influenced participants' auditory recognition performance, we submitted the data of Experiment 1a and Experiment 1b to a mixed-factor ANOVA including the within-subjects factor event boundary (event boundary items, non-event boundary boundaries) and the between-subjects factor experiment (Experiment 1a, Experiment 1b). Results showed a significant interaction of event boundary and experiment, $F(1, 57) = 21.98, p < .001, \eta_p^2 = .28$. While event boundary items were better recognized than non-event boundary items when they were encoded within their narrative context, non-event boundary items were better recognized than event boundary items when they were encoded without narrative context. Further, performance in Experiment 1a was higher as compared to Experiment 1b, $F(1, 57) = 11.34, p = .001, \eta_p^2 = .17$, suggesting that there is a general memory benefit if the items were encoded in their narrative context. The main effect of event boundary was not significant, $F(1, 57) = 3.19, p = .080, \eta_p^2 = .05$.

We thus conclude that the narrative context is an important factor for event memory. Yet, the results of this exploratory analysis should be regarded with caution because the participants were not randomly assigned to the two experiments.

Experiment 2a

In Experiment 2a, we tested the *incremental updating hypothesis* using normally sighted participants.

Methods

Participants and design. Thirty-three normally sighted participants, mostly students of the University of Tübingen, participated in this experiment for course credit or monetary compensation. We excluded 6 participants because they were familiar with the audio drama episode used. Thus, we analyzed the data of 27 participants (8 males), aged 18 to 30 years ($M = 22.96$, $SD = 3.31$), with normal hearing ability. Their first language was German. An a priori power analysis revealed that a sample size of at least 26 participants was needed to reach a power of .85 to detect an effect size of $\omega = .55$ (Huff et al., 2014).

We manipulated number of dimension changes (1 to 6 changes, within-subjects) as independent variable. As in Experiment 1a we measured recognition performance and expertise.

Stimulus material. The stimulus material for the second experiment was similar to Experiment 1a with the following exceptions: We used 46 items that were classified as event boundaries in Huff et al. (revision under review) and their respective distractors (see Appendix B). Two independent raters coded event boundaries for changes in the following six dimensions: time, location, person, action (Huff et al., 2014; Magliano & Zacks, 2011) and, additionally, the audio drama specific dimensions: narrator and music.

Time. A time change was coded for time leaps of either defined or undefined length (e.g. “The next day”, “shortly afterwards”, auditory cues such as owl voices), interrater reliability Cohen’s Kappa $\kappa = .88$ (J. Cohen, 1960, 1968).

Location. Location changes included changes in the whole setting or transitions between clearly defined areas/rooms, $\kappa = .96$.

Person. A change was coded if the constellation of protagonists after a boundary was different from before (including telephone calls), $\kappa = .87$.

Action. If a change in action, intention, or the conversation topic occurred an action change was coded, $\kappa = .78$.

Narrator. A narrator change was coded if either a beginning or an ending of a narrator sequence occurred during the event boundary, $\kappa = 1$.

Music. A music change was coded if either a beginning or an ending of a music sequence occurred during the event boundary, $\kappa = 1$.

Discrepancies were resolved through discussion. This led to one boundary with zero changes, 12 boundaries with one change, 13 boundaries with two changes, five boundaries with three changes, three boundaries with four changes, seven boundaries with five changes and four boundaries with six changes (Appendix B). The boundary at the end of the drama could not be coded and was therefore excluded from the analysis. The intercorrelations for the dimension changes are shown in Table 2.

Expertise questionnaire. The expertise questionnaire as well as the method for calculating the expertise score were described in Experiment 1a. Item statistics (Appendix A) and reliability of the expertise scale, $\alpha = .87$, were calculated from the whole sample of Experiment 2a ($N = 33$). The expertise scale ranged from 0 to 16 ($M = 6.32$, $SD = 4.05$), excluding the participants who knew the presented episode. Because of missing data, no expertise score could be calculated for two participants.

Technical implementation. Programming and presentation were equivalent to Experiment 1a.

Procedure. Procedure and duration of Experiment 2a were similar to Experiment 1a.

Results and Discussion

Two targets with their respective distractors were excluded from the analysis. These were the boundary with 0 dimension changes and the last boundary which could not be coded. Data analysis was similar as in Experiment 1a. To get a reliable number of data points per dimension change condition, we collapsed across the dimension change conditions with 1 and 2 changes as well as across conditions with 3, 4, 5, and 6 changes. The descriptive data are plotted in Figure 2A. Recognition performance was not influenced by the number of dimension changes as indicated by the non-significant main effect, $F(1, 25) < 1$. Further, the main effect for expertise was significant, $F(1, 25) = 5.54, p = .027, \eta_p^2 = .18$, indicating higher recognition performance with increasing expertise score (see Figure 2B). However, the interaction of both factors, $F(1, 25) = 2.67, p = .115, \eta_p^2 = .10$, was not significant.

The most important and at the same time surprising finding of this experiment is the non-significant relation between number of dimension changes and recognition performance. This is in contrast to recent findings that used audiovisual narratives (Huff et al., 2014). The significant main effect for expertise might indicate that information processing is different for low and high experienced listeners. However, because this factor did not interact with the amount of change present at the event boundary, we conclude that event model updating when perceiving auditory narratives is independent of experience, at least for the tested student sample. Although the present results suggest that event model updating is global (i.e. the whole event model is reset and built up from scratch) it could be argued that auditory information processing is less common for the tested sighted sample. Therefore, we tested a sample of blind participants in Experiment 2b using a similar experiment setup.

Experiment 2b

Experiment 2b replicated Experiment 2a with a sample of blind participants who had little or no experience with visual information processing and who were presumably more experienced with auditory information processing (Lessard, Paré, Lepore, & Lassonde, 1998; Röder et al., 2000).

Method

Participants and design. Twenty-seven blind participants participated in this experiment. They were recruited at schools and boarding schools for blind and visually impaired people. We excluded twelve participants – two because of data technical problems, two because of young age, one because of a high remaining vision, and five because they already knew the audio drama used. Further, data of two participants were incomplete – they did not fill-in the expertise questionnaire – we thus were not able to include them in the analysis. Thus, we analyzed 15 participants (9 male), aged 13 to 26 years ($M = 17.00$, $SD = 3.66$), with normal hearing ability. Their first language was German. For power analysis, see Experiment 2a.

We used the same design as described in Experiment 2a. In addition to the mentioned expertise and confidence measurements, we collected the remaining visual acuity and the time point of the beginning of blindness.

Material and Procedure

Stimulus material. Stimulus material in the learning phase was similar to Experiment 2a. To test the items for the auditory recognition test, we conducted a pre-study in which we asked seven blind persons, aged 22 to 32 years ($M = 25.67$ years, $SD = 3.98$), to segment the audio drama into meaningful events. Segmentation was comparable except for four event boundaries that were perceived by the sighted but not by the blind participants (two boundaries with an action change as well as two boundaries with an action change and a location change). Therefore, we excluded these four items (including the respective distractor items) from the recognition test (see also Appendix B).

Distraction task and expertise questionnaire. For distraction between learning and testing phase, we used a haptic puzzle distraction task in equivalence to the PFT used in Experiment 1a and 2a. The reliability of the expertise scale was $\alpha = .92$. We report the item characteristics in Appendix A. In Experiment 3, the expertise scale ranged from 0 to 20 ($M = 6.63$, $SD = 5.94$), excluding participants who were familiar with the used episode.

Technical implementation. Programming and presentation were equivalent to Experiment 2a with the following exception: All written instructions were read aloud via a computer voice. This voice

was recorded using the default screen reader program implemented in OS-X 10.9. The answer keys including the number keys 1 to 5 were marked with glue dots so that the blind participants could easily find them.

Procedure

Procedure was similar to Experiment 2a with the following exceptions. The experimenter read information about the experiment to the participants. Next, participants signed the informed consent forms before they were seated in front of the notebooks to hear the audio drama via headphones. After listening to the drama, demographic information was collected and participants completed the distraction task. Participants then proceeded with the recognition test. The experiment lasted a total of 90 minutes.

Results and Discussion

Data analysis was similar to Experiment 2a. Descriptive data are plotted in Figure 3. Again, recognition performance was not influenced by the number of dimension changes as indicated by the non-significant main effect, $F(1, 14) = 1.30, p = .273, \eta_p^2 = .09$. Further, neither the main effects for expertise, $F(1, 14) = 3.25, p = .093, \eta_p^2 = .19$, nor the interaction of both factors, $F(1, 14) = 3.50, p = .082, \eta_p^2 = .20$, reached significance.

Most important, the blind participants' recognition performance resembled that of the normally sighted participants in Experiment 2a. Recognition performance was independent of the number of dimension changes and, as in Experiment 1a with normally sighted participants, expertise (i.e. experience with audio dramas) did not influence memory performance in blind participants either. We thus propose that the underlying process of memory formation is similar in normally sighted and blind participants.

Comparing Performance of the Sighted vs. Blind Participants

In a next step, we compared recognition performance across Experiments 2a and 2b to check if there are differences with respect to sensitivity that can be traced back to the participants' visual abilities. We submitted the data to an ANOVA including the between-subjects factor experiment (Experiment 2a, Experiment 2b) and the within-subjects factor number of dimension changes. Neither the main effects of experiment, $F < 1$, and number of dimension changes, $F(1, 42) = 1.43$, $p = .239$, $\eta_p^2 = .03$, nor the interaction of both factors, $F < 1$, were significant. Taken together, the main result of Experiment 2a and 2b – auditory recognition performance was independent of the number of dimension changes – did not depend on participants' visual abilities. However, because of its exploratory nature, the results of this analysis need to be considered with caution.

Experiment 3

Memory for event boundaries is higher than for non-event boundaries (Experiment 1a and 1b) and is independent from the number of dimension changes at an event boundary (Experiment 2a and 2b). Although the results of Experiment 2a and 2b suggest that event model updating is global, there is still the possibility that other indicators of updating provide further information of the process. As an example, Radvansky and Copeland (2010) showed that the same narrative shift (such as a shift in the spatial dimension) differently affects reading times and memory probes. Applying this logic to the present set of experiments, event model updating might still be incremental but the used recognition memory measure might be insensitive for the underlying process. As basic processes of dynamic event processing are assumed to be independent of modality (Zacks et al., 2009), we used a written transcript of the audio drama as stimulus material to disentangle encoding and retrieval processes. During encoding, participants read the audio drama in a self-paced manner allowing us to measure reading times. In addition, we also measured participants' recognition memory.

If updating is incremental we expect reading times to increase with an increasing number of dimension changes. This result would indicate a higher cognitive effort in processing text passages with a higher number of dimension changes depicted in the respective text (Radvansky & Copeland, 2010; Zwaan & Radvansky, 1998). If, however, updating is global, we expect reading times to be independent

of the number of dimension change. Increased effort during encoding could be related to higher memory performance in the recognition test. However, if processes during encoding and retrieval are independent, we expect a similar result pattern as in Experiment 2a.

Method

Participants and Design. We recruited 31 participants. We excluded six participants, who already knew the audio drama. Thus, we analyzed data of 25 participants (six male), aged from 18 to 34 years ($M = 21.36$, $SD = 3.49$). All participants reported sufficient hearing abilities and German as being their native language.

Material and Procedure

Stimulus material. Basis of the stimulus material was a transcript of the audio drama (see Experiment 1a). We presented the narrative in *sentences*. Each speaker was indicated by its name (printed in bold letters) at the beginning of his part, for example: ‘**Peter:** Look, there is the castle!’. The noise information such as ‘*door is closing*’ or ‘*parrot squawks*’ was printed in italic letters (see Figure 4 for an example). For designing the recognition test, we transcribed all recognition items of Experiment 2a. The recognition items included speaker and noise information in a similar way as the textual information during encoding (see Table 1 for examples).

Distraction task and expertise questionnaire. For distraction between learning and testing phase, we used the PFT as in Experiment 1a. The reliability of the expertise scale was $\alpha = .81$. We report the item characteristics in Appendix A. In Experiment 3, the expertise scale ranged from 0 to 13 ($M = 6.36$, $SD = 4.01$), excluding participants who already knew the narrative.

Technical implementation. Technical implementation was done in PsychoPy 2 (Peirce, 2007). The participants were instructed to press the spacebar whenever they were ready to read the next part of the transcript.

Procedure. Procedure was similar to Experiment 2a with the exception that participants read the narrative (instead of listened to it as in Experiment 2a). We presented each part (including speaker

and related noise information) separately on-screen and instructed participants to read the text self-paced by pressing the spacebar when they felt ready to proceed to the next part (Rinck & Weber, 2003). The experiment lasted about 90 minutes.

Results and Discussion

We analyzed reading times during initial reading as well as memory performance in the recognition test.

Reading times. We included the same event boundaries as in Experiments 2a for this analysis. Individual reading time data were divided by the number of words per part and subsequently aggregated on the participant level, resulting in six measures – one for each dimension change condition – for each participant. We fitted a linear mixed effect model (lme) with *number of dimension changes* as continuous fixed effect as well as random intercept and random slope for the *participant* effect and random intercept and slope for the *item* effect (Barr, Levy, Scheepers, & Tily, 2013). We analyzed the model parameters with a type-II ANOVA using the Anova function of the car package (Fox & Weisberg, 2010). This analysis revealed a significant main effect of number of dimension changes, $\chi^2(1) = 8.71$, $p = .003$. Thus, indicating that reading times increased linearly with increasing number of dimension changes (see Figure 5A).

Recognition performance. Analysis was similar to Experiment 2a. See Figure 5B for the descriptive statistics. Neither the main effect of number of dimension changes, $F(1, 23) = 3.11$, $p = .091$, $\eta_p^2 = .12$, and expertise, $F(1, 23) = 3.88$, $p = .061$, $\eta_p^2 = .14$, nor the interaction of both factors were significant, $F(1, 23) < 1$. Thus, with regard to the number of dimension changes, results of Experiment 3 are similar to those of Experiments 2a and 2b.

Taken together, the results of Experiment 3 replicate previous research – reading times increased with increasing number of dimension changes (Zwaan & Radvansky, 1998). Further, and equally important, they also replicate the memory test results of Experiments 2a and 2b with text stimuli

– recognition performance was not related with the number of dimension changes, thus suggesting that the results are independent of modality (at least across the auditory and textual modality).

The present experiment allowed us to disentangle encoding and retrieval processes (Radvansky & Copeland, 2010). Most important, as encoding effort (as measured with reading times) increases with amount of change, we propose that event model updating is incremental as suggested by the additivity hypothesis of Event Indexing (Zwaan, Langston, et al., 1995). This result further supports the claim of Radvansky and Copeland (2010) according to which only multiple measures allow the exact description of event cognition processes.

General Discussion

In five experiments, we investigated perceptual and cognitive consequences of event boundary perception. Participants listened to an audio drama or read the written transcript thereof and completed a recognition test afterwards. Results of Experiment 1a replicated the event boundary advantage effect – higher memory for event boundaries as compared to non-event boundaries (Newtson & Engquist, 1976). Experiment 1b ruled out the alternative explanation that the event boundary advantage effect is caused by lower difficulties of the event boundary items. Instead, in case the participants encoded the target items without their respective context, we no longer observed an advantage in the recognition test. In case there is no event model representing the narrative, no updating is possible and, hence, there is no advantage for event boundary items. We thus conclude that the basis of the event boundary advantage for memory is updating of the present event model in working memory. In Experiment 2a and 2b, we studied how the amount of change presented at event boundary relates to memory performance. We observed no influence of the number of dimension changes on recognition performance. This was true for both sighted and blind participants and is in contrast to recent research using audiovisual stimuli, which described event model updating as being incremental in nature (Huff et al., 2014). To solve this conflict and because we only tested recognition memory so far, we aimed at finding a further indicator for event processing in Experiment 3 (Radvansky & Copeland, 2010). Participants read a transcribed version of the audio drama in a self-paced manner and completed a

textual recognition test afterwards. This allowed us to disentangle encoding and retrieval processes and to describe the event encoding process using reading times. Importantly, as reading times increased with an increasing amount of change at an event boundary, we suggest that updating is indeed incremental. Probably, recognition memory was too coarse to measure these effects.

To the best of our knowledge this is the first study that studied basic processes of event cognition (such as memory formation) using auditory narratives. An important finding was that we replicated the memory advantage for event boundaries (Newtson & Engquist, 1976), thus, suggesting that auditory stimuli are processed in a similar way as visual stimuli. Further, auditory narratives had the advantage that we were able to explore how experience with auditory information processing influences dynamic event processing. Both, the expertise score of the sighted participants and the results of the blind participants suggest that auditory information processing is not influenced by these. Finally, by transcribing the auditory narratives we were able to use a further indicator of event comprehension (Radvansky & Copeland, 2010), namely reading times and to link auditory event processing to the well-established domain of text comprehension research (Zwaan, Langston, et al., 1995). We thus conclude that perceiving and understanding dynamic events does neither depend on the modality of the stimulus material (Zacks et al., 2009) nor on the expertise and capability of the perceiver. Instead, we propose that event perception is general perceptual-cognitive mechanism.

Theoretical Implications

The results of this project relate to present theories of event cognition. Both event segmentation theory (EST; Zacks et al., 2007) and event indexing (EI; Zwaan, Langston, et al., 1995; Zwaan, Magliano, et al., 1995; Zwaan & Radvansky, 1998) propose that the current working memory representation (“event model”) is updated at event boundaries. However, EST and EI make different predictions with regard to the nature of this process – whereas EST proposes updating to be global, EI proposes it to be incremental. To date, most empirical results found support for the latter, suggesting

that this process is incremental as both reading times (Zacks et al., 2009; Zwaan, Magliano, et al., 1995; Zwaan et al., 1998) and recognition performance increases linearly with the number of dimension changes (Huff et al., 2014). The present results add important information to our understanding of this process. First and foremost, it is important to mention that, in general, event boundaries are remembered to a higher degree than non-boundaries (Experiments 1a and 1b). Thus, the basic mechanism of event model updating is comparable across modalities and supports the basic assumption of EST that event processing (including event model construction and updating as well as subsequent memory-related processes) is multimodal in nature.

This view is further supported by the results of Experiments 2a and 2b, which showed no difference between sighted and blind participants with regard to recognition memory. Although blind people are highly experienced in processing auditory information – in comparison to people with normal vision, they show specific differences (presumably due to reorganisation of occipital structures) such as a higher memory capacity for auditory information and faster processing of auditory stimuli (Amedi et al., 2003; Bavelier & Neville, 2002; Bull et al., 1983; M. A. Cohen et al., 2009; Perleth & Effinger, 2001; Röder et al., 2000) – they had no advantage in processing dynamic events.

Second, the results of the present study suggest that event model updating is incremental, thus supporting the additivity hypothesis of EI (e.g., Zwaan, Langston, et al., 1995). This conclusion is based on the combination of the results of Experiments 2a, 2b, and 3. If one only considers memory performance, one might come to the conclusion that updating is global (i.e. the whole event model is build-up after the perception of an event boundary). However, because reading times increased linearly with an increasing number of dimension changes, thus reflecting increased effort (Radvansky & Copeland, 2010), we conclude, that updating of event models is incremental.

Limitations

This is one of the first studies to use an audio drama as stimulus material. Although we employed exactly the same procedure (including instructions and statistical analyses) to determine the event boundaries (Huff et al., revision under review), there might be some differences between

audiovisual and auditory stimuli. In particular, the auditory recognition test items differed substantially from the audiovisual items typically used. Whereas most studies examining audiovisual material used static video stills to test recognition memory, we used audio clips of 4.5 seconds around the corresponding event boundaries of the audio drama. In contrast to the video stills that depict the plot immediately after the event boundary, the dynamic auditory test items of the present study cover parts of the event before and the event after the event boundary. At this point, further research is necessary to study retrieval-based influences on recognition memory in event perception research: for example, by comparing free recall test performance with recognition test performance or think-aloud measures, as it is used in text comprehension research (Kurby & Zacks, 2012).

It is important to note that we used just a single audio drama to explore event model construction and updating processes. We decided to use this specific auditory narrative because it has a lot of dimension changes and it has just one unfolding narrative thread. Thus, this stimulus material meets most of the criteria which made *The Red Balloon* (Lamorisse, 1956) very popular in cognitive research (Baggett, 1979; Zacks et al., 2009). However, more research using a broader range of stimuli (such as audio books or recorded conversations) is needed to explore the boundary conditions of the effects described in this project. However, because of the robustness of the present results (e.g., event boundary advantage) and their fit with current theoretical concepts (EST, EHM, EI), we are confident that the present findings will prove to be a solid basis for future research.

Conclusion

Perceiving and remembering changes in dynamic events such as narratives triggers the perception of an event boundary and the updating of the event model in working memory, consequently. Previous research found that memory is higher for event boundaries than for non-event boundaries (*event boundary advantage effect*) and that memory performance increases with an increasing amount of change at an event boundary (*incremental updating effect*). We studied processes of event model construction and updating using an auditory narrative and a written transcript thereof. We replicated the event boundary advantage effect but also found that memory performance was unrelated to the amount

of changes at event boundaries. However, because additional measures reflect increasing effort during encoding, we propose that updating of event models is incremental. Our experiments have demonstrated that more than one dependent measure is needed to describe processes of event model construction and updating.

Acknowledgement

This research was supported by the “Juniorprofessorenprogramm” funded by the Ministry of Science, Research and the Arts Baden-Württemberg, this research grant was awarded to MH.

Open practices

Data and analyses scripts (for the statistical programming language R) have been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/fmy4u/>

Author Contributions

AEM and MH developed the study concept and the study design. AEM, IB, and AP programmed the experiments. Data collection was performed by IB, AP, and student assistants under the supervision of AEM. Data analysis was performed by MH and AEM in consultation with FP and FW. MH and AEM drafted the manuscript and FP provided critical revisions. All authors approved the final version of the manuscript for submission.

References

- AGF Videoforschung. (2017). Retrieved March 30, 2017, from <https://www.agf.de/daten/tvdaten/sehdauer/>
- Amedi, A., Raz, N., Pianka, P., Malach, R., & Zohary, E. (2003). Early “visual” cortex activation correlates with superior verbal memory performance in the blind. *Nature Neuroscience*, 6(7), 758–766. <https://doi.org/10.1038/nn1072>
- Arthur, R., & Hitchcock, A. (1964). *The Secret of Terror Castle*. Random House Childrens Books.
- Baggett, P. (1979). Structurally equivalent stories in movie and text and the effect of the medium on recall. *Journal of Verbal Learning and Verbal Behavior*, 18(3), 333–356.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
- Bavelier, D., & Neville, H. J. (2002). Cross-modal plasticity: where and how? *Nature Reviews. Neuroscience*, 3(6), 443–452. <https://doi.org/10.1038/nrn848>
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, 105(38), 14325–14329.
- Bull, R., Rathborn, H., & Clifford, B. R. (1983). The Voice-Recognition Accuracy of Blind Listeners. *Perception*, 12(2), 223–226. <https://doi.org/10.1068/p120223>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213. <https://doi.org/10.1037/h0026256>
- Cohen, L. G., Celnik, P., Pascual-Leone, A., Corwell, B., Falz, L., Dambrosia, J., ... Hallett, M. (1997). Functional relevance of cross-modal plasticity in blind humans. *Nature*, 389(6647), 180–183. <https://doi.org/10.1038/38278>

- Cohen, M. A., Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). Auditory and visual memory in musicians and nonmusicians. *Psychonomic Bulletin & Review*, 18(3), 586–591.
<https://doi.org/10.3758/s13423-011-0074-0>
- Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *PNAS*, (106), 6008–6010.
- DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General*, 142(4), 1277–1286.
<https://doi.org/10.1037/a0034024>
- Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. *Princeton, NJ: Educational Testing Service*. Retrieved from http://www.ets.org/Media/Research/pdf/Manual_for_Kit_of_Factor-Referenced_Cognitive_Tests.pdf
- Fox, J., & Weisberg, S. (2010). *An R companion to applied regression*. Sage.
- Graziano, W. G., Moore, J. S., & Collins, J. E. (1988). Social cognition as segmentation of the stream of behavior. *Developmental Psychology*, 24(4), 568–573. <https://doi.org/10.1037/0012-1649.24.4.568>
- Hard, B. M., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: building event schemas. *Memory & Cognition*, 34(6), 1221–35.
- Huff, M., Maurer, A. E., & Papenmeier, F. (revision under review). Linguistic information in auditory dynamic events is processed at fine, not coarse event boundaries. Retrieved from <https://osf.io/bvx3b/>
- Huff, M., Meitz, T. G. K., & 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), 1377–1388. <https://doi.org/10.1037/a0036780>
- Huff, M., Papenmeier, F., Maurer, A. E., Meitz, T. G. K., Garsoffky, B., & Schwan, S. (2017). Fandom biases retrospective judgments not perception. *Scientific Reports*, 7, 43083.
<https://doi.org/10.1038/srep43083>

- Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review*, 2(1), 105–112. <https://doi.org/10.3758/BF03214414>
- Körting, H. (2001). *Die drei Fragezeichen und das Gespensterschloss* [Audio CD]. Europa (Sony Music).
- Kurby, C. A., & Zacks, J. M. (2012). Starting from scratch and building brick by brick in comprehension. *Memory & Cognition*, 40(5), 812–26. <https://doi.org/10.3758/s13421-011-0179-8>
- Lamoris, A. (1956). *The red balloon*. Films Montsouris.
- Lessard, N., Paré, M., Lepore, F., & Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects. *Nature*, 395(6699), 278–280. <https://doi.org/10.1038/26228>
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.
- 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), 1489–517. <https://doi.org/10.1111/j.1551-6709.2011.01202.x>
- Meyerhoff, H. S., & Huff, M. (2015). Semantic congruency but not temporal synchrony enhances long-term memory performance for audio-visual scenes. *Memory & Cognition*, 1–13. <https://doi.org/10.3758/s13421-015-0575-6>
- Newton, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 28(1), 28–38. <https://doi.org/10.1037/h0035584>
- Newton, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, 12(5), 436–450. [https://doi.org/10.1016/0022-1031\(76\)90076-7](https://doi.org/10.1016/0022-1031(76)90076-7)
- Papenmeier, F. (2014). segmag: Determine Event Boundaries in Event Segmentation Experiments (Version 1.2.2). Retrieved from <http://cran.r-project.org/web/packages/segmag/index.html>

- Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1), 8–13.
- Perfect, T. J., Watson, E. L., & Wagstaff, G. F. (1993). Accuracy of confidence ratings associated with general knowledge and eyewitness memory. *Journal of Applied Psychology*, 78(1), 144–147. <https://doi.org/10.1037/0021-9010.78.1.144>
- Perleth, C., & Effinger, I. (2001). “Beiläufiges” Lernen bei Personen mit unterschiedlich starker Sehschädigung. *Unterrichtswissenschaft*, 29(2), 131–152.
- Radvansky, G. A. (2012). Across the event horizon. *Current Directions in Psychological Science*, 21(4), 269–272. <https://doi.org/10.1177/0963721412451274>
- Radvansky, G. A., & Copeland, D. E. (2010). Reading times and the detection of event shift processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1), 210–216. <https://doi.org/10.1037/a0017258>
- Rinck, M., & Weber, U. (2003). Who when where: an experimental test of the event-indexing model. *Memory & Cognition*, 31(8), 1284–1292.
- Roder, B., & Neville, H. (2003). Developmental functional plasticity. *Handbook of Neuropsychology*, 9, 231–270.
- Röder, B., Rösler, F., Hennighausen, E., & Näcker, F. (1996). Event-related potentials during auditory and somatosensory discrimination in sighted and blind human subjects. *Cognitive Brain Research*, 4(2), 77–93. [https://doi.org/10.1016/0926-6410\(96\)00024-9](https://doi.org/10.1016/0926-6410(96)00024-9)
- Röder, B., Rösler, F., & Neville, H. J. (2000). Event-related potentials during auditory language processing in congenitally blind and sighted people. *Neuropsychologia*, 38(11), 1482–1502. [https://doi.org/10.1016/S0028-3932\(00\)00057-9](https://doi.org/10.1016/S0028-3932(00)00057-9)
- Sebastian, K., Ghose, T., Zacks, J. M., & Huff, M. (2017). Understanding the cognitive potential of intellectually disabled people in workshops for adapted work. *Applied Cognitive Psychology*. <https://doi.org/10.1002/acp.3315>

- 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.
<https://doi.org/10.1016/j.jml.2005.02.009>
- Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science*, 18(5), 449–55. <https://doi.org/10.1111/j.1467-9280.2007.01920.x>
- Sridharan, D., Levitin, D. J., Chafe, C. H., Berger, J., & Menon, V. (2007). Neural Dynamics of Event Segmentation in Music: Converging Evidence for Dissociable Ventral and Dorsal Networks. *Neuron*, 55(3), 521–532. <https://doi.org/10.1016/j.neuron.2007.07.003>
- Standing, L. (1973). Learning 10000 pictures. *The Quarterly Journal of Experimental Psychology*, 25(2), 207–222.
- Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 19(2), 73–74.
<https://doi.org/10.3758/BF03337426>
- Swallow, K. M., Barch, D. M., Head, D., Maley, C. J., Holder, D., & Zacks, J. M. (2010). Changes in events alter how people remember recent information. *Journal of Cognitive Neuroscience*, 23(5), 1052–1064. <https://doi.org/10.1162/jocn.2010.21524>
- Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, 138(2), 236–257. <https://doi.org/10.1037/a0015631>
- Verde, M. F., & Rotello, C. M. (2003). Does familiarity change in the revelation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(5), 739.
- Whitney, C., Huber, W., Klann, J., Weis, S., Krach, S., & Kircher, T. (2009). Neural correlates of narrative shifts during auditory story comprehension. *NeuroImage*, 47(1), 360–366.
<https://doi.org/10.1016/j.neuroimage.2009.04.037>

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., ...

Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries.

Nature Neuroscience, 4(6), 651–655.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film

comprehension. *Journal of Experimental Psychology: General*, 138(2), 307–327.

<https://doi.org/10.1037/a0015305>

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception:

a mind-brain perspective. *Psychological Bulletin*, 133(2), 273–293.

<https://doi.org/10.1037/0033-2909.133.2.273>

Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. (2006). Visual motion and the neural

correlates of event perception. *Brain Research*, 1076(1), 150–162.

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.

<https://doi.org/10.1111/j.1467-9280.1995.tb00513.x>

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction

in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and*

Cognition, 21(2), 386–397. <https://doi.org/10.1037/0278-7393.21.2.386>

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory.

Psychological Bulletin, 123(2), 162–185.

Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing

multidimensional situation models during reading. *Scientific Studies of Reading*, 2(3), 199–

220. https://doi.org/10.1207/s1532799xssr0203_2

Table 1: Examples of target and matched distractor items used in this experiment .English translation by the authors. Stimulus material was presented in German.

Target item	Distractor item
Justus: I'll answer the phone! <i>Telephone rings;</i> Switch on the amplifier, Bob. Bob: Alright. Justus: Hello, Justus Jonas	Peter: Do you come? <i>Telephone rings;</i> He Bob, Just, what's up? Justus: Oh!
Morton: For Sure! Justus: Mh, thanks! Let's get out. <i>Car door is opened; chirping; steps</i>	Justus: Yes, Morton. Please wait here. Morton: Of course! <i>Crickets chirping</i>

Table 2: Intercorrelation matrix of dimension changes at event boundaries.

	Time	Location	Person	Action	Narrator	Music
Time	1.00	.43**	.37*	.18	.55***	.69***
Location		1.00	.20	.33*	.23	.29
Person			1.00	.15	.39**	.33*
Action				1.00	-.09	.17
Narrator					1.00	.38**
Music						1.00

Note. Without the last event boundary. * $p < .05$, ** $p < .01$, *** $p < .001$

Figure 1. Mean receiver-operator-characteristic (ROC) curves obtained from the combination of the yes/no and confidence ratings in Experiment 1a (A) and Experiment 1b (B), as a function of event boundary (event boundary vs. non-event boundary).

Figure 2. Mean receiver-operator-characteristic (ROC) curve obtained from the combination of the yes/no and confidence ratings in Experiment 2a (A) and sensitivity d_a as a function of participants' expertise (B). Error interval represents the standard error of the mean (*SEM*).

Figure 3. Mean receiver-operator-characteristic (ROC) curve obtained from the combination of the yes/no and confidence ratings in Experiment 2b.

Figure 4. Exemplary presentation of the reading time method used in Experiment 3. Participants read the story sentence-by-sentence by pressing the space bar to proceed to the next sentence. Person information was indicated by bold and noise information by italic letters. Grey shaded items represent sentences belonging to an event boundary. Please note that in the experiment, all sentences were presented on a white background. English translation by the authors. Stimulus material was presented in German.

Figure 5. A: Mean reading time per word at an event boundary as a function of the number of dimension changes of the respective event boundary. Error interval represents the standard error of the mean (*SEM*). B: Mean receiver-operator-characteristic (ROC) curves obtained from the combination of the yes/no and confidence ratings in recognition test of Experiment 3 as a function of number of dimension changes.

Appendix A

Item statistics for the expertise scale with item-total correlation r , means and standard deviations. Item-total correlation r was calculated with part-whole correction.

Experiment 1a	r	M	SD
audio drama in general	.59	4.45	2.31
audio drama “The Three Investigators”	.78	2.45	2.58
number of episodes listened to	.64	1.58	1.20
Experiment 1b	r	M	SD
audio drama in general	.45	3.47	2.37
audio drama “The Three Investigators”	.65	0.88	1.24
number of episodes listened to	.58	1.00	1.11
Experiment 2a	r	M	SD
audio drama in general	.69	4.15	2.49
audio drama “The Three Investigators”	.82	2.26	2.63
number of episodes listened to	.73	1.66	1.15
Experiment 2b	r	M	SD
audio drama in general	.60	3.16	2.03
audio drama “The Three Investigators”	.75	1.80	1.61
number of episodes listened to	.61	1.40	1.04
Experiment 3	r	M	SD
audio drama in general	.77	5.52	3.28
audio drama “The Three Investigators”	.83	3.16	3.05
number of episodes listened to	.87	1.43	1.24

Appendix B

Coding of event boundaries for changes in the dimensions: time, location, person, action, narrator and music, number of dimension changes, their segmentation magnitude, if they were also segmented in the coarse condition of Huff et al. (revision under review) and the experiments they were shown in. Naming of boundaries results from their position in the audio drama, in seconds from beginning.

Event boundary id	Time	Location	Person	Action	Nar-rator	Music	No. dim. change	Segmen-tation magn.	Coarse boundar y	Shown in Exp.
EG0019	No	No	No	Yes	No	No	1	17.47	No	all
EG0025	No	No	Yes	Yes	No	No	2	10.75	Yes	2a, 2b, 3
EG0090	Yes	Yes	Yes	Yes	Yes	Yes	6	23.68	Yes	all
EG0114	Yes	Yes	Yes	Yes	Yes	No	5	18.64	No	all
EG0201	No	No	No	Yes	No	No	1	10.62	No	2a, 2b, 3
EG0244	Yes	Yes	No	Yes	No	Yes	4	23.42	Yes	all
EG0263	No	Yes	Yes	Yes	No	No	3	12.49	No	all
EG0464	No	Yes	Yes	Yes	Yes	Yes	5	16.58	Yes	all
EG0530	No	No	No	No	No	No	0	10.46	No	2a, 2b, 3
EG0539	No	No	No	Yes	No	No	1	10.55	No	2a, 2b, 3
EG0550	No	No	Yes	Yes	No	No	2	11.53	No	2a, 2b, 3
EG0578	No	No	Yes	Yes	No	No	2	10.66	No	2a, 2b, 3
EG0590	Yes	No	Yes	Yes	Yes	Yes	5	23.47	Yes	all
EG0684	No	No	No	Yes	No	No	1	12.89	No	all
EG0706	Yes	Yes	No	Yes	Yes	Yes	5	19.88	Yes	all
EG0721	No	Yes	Yes	Yes	Yes	No	4	13.78	No	all
EG0755	No	Yes	Yes	Yes	No	No	3	9.05	No	2a, 2b, 3
EG0780	No	No	No	Yes	No	No	1	8.71	No	2a, 2b, 3
EG0822	No	No	No	Yes	No	No	1	12.08	No	all
EG0975	Yes	Yes	Yes	Yes	No	Yes	5	15.66	Yes	all
EG1003	No	Yes	Yes	Yes	No	No	3	11.75	No	2a, 2b, 3
EG1040	No	Yes	No	Yes	No	No	2	10.55	No	2a, 3
EG1171	No	No	No	Yes	No	No	1	9.35	No	2a, 3
EG1228	Yes	Yes	Yes	Yes	Yes	Yes	6	22.24	Yes	all
EG1239	No	No	Yes	Yes	Yes	No	3	11.95	No	2a, 2b, 3
EG1327	No	No	Yes	Yes	No	Yes	3	10.00	Yes	2a, 2b, 3
EG1381	No	No	Yes	No	No	No	1	12.70	No	all
EG1448	Yes	Yes	Yes	Yes	Yes	Yes	6	12.59	Yes	all
EG1467	No	Yes	No	Yes	No	No	2	9.90	No	2a, 2b, 3
EG1581	No	No	No	Yes	No	No	1	8.98	No	2a, 3

EG1651	No	Yes	No	Yes	No	No	2	9.35	No	2a, 2b, 3
EG1714	Yes	Yes	Yes	Yes	Yes	Yes	6	17.19	Yes	all
EG1744	Yes	Yes	Yes	Yes	Yes	No	5	18.11	No	all
EG1769	No	Yes	No	Yes	No	No	2	12.94	No	1a, 1b, 2a, 3
EG1863	No	No	No	Yes	No	No	1	11.23	No	2a, 2b, 3
EG1924	No	No	No	No	Yes	No	1	10.03	No	2a, 2b, 3
EG1950	No	No	No	Yes	Yes	No	2	13.41	No	all
EG2003	No	Yes	No	Yes	No	No	2	12.06	No	all
EG2034	No	No	Yes	Yes	No	No	2	8.88	No	2a, 2b, 3
EG2166	No	Yes	Yes	Yes	Yes	No	4	9.15	Yes	2a, 2b, 3
EG2188	No	No	No	No	Yes	No	1	13.40	No	all
EG2248	No	Yes	No	Yes	No	No	2	11.92	No	2a, 2b, 3
EG2321	No	Yes	No	Yes	No	No	2	10.16	No	2a, 2b, 3
EG2342	No	Yes	No	Yes	No	No	2	10.80	No	2a, 2b, 3
EG2611	Yes	Yes	Yes	Yes	Yes	No	5	13.39	Yes	all
EG2646	No	No	No	No	Yes	Yes	2	15.83	Yes	all