Linguistic Information in Auditory Dynamic Events Contributes to the Detection of Fine, Not Coarse Event Boundaries

Frank Papenmeier[1], Annika E. Maurer[1], and Markus Huff[2, 1]

[1] Department of Psychology, Eberhard Karls Universität Tübingen, Tübingen, Germany

[2] Department of Research Infrastructures, German Institute for Adult Education, Bonn, Germany

ABSTRACT

Human observers (comprehenders) segment dynamic information into discrete events. That is, although there is continuous sensory information, comprehenders perceive boundaries between two meaningful units of information. In narrative comprehension, comprehenders use linguistic, non-linguistic , and physical cues for this event boundary perception. Yet, it is an open question – both from a theoretical and an empirical perspective – how linguistic and non-linguistic cues contribute to this process. The current study explores how linguistic cues contribute to the participants' ability to segment continuous auditory information into discrete, hierarchically structured events. Native speakers of German and non-native speakers, who neither spoke nor understood German, segmented a German audio drama into coarse and fine events. Whereas native participants could make use of linguistic, non-linguistic, and physical cues for segmentation, non-native participants could only use non-linguistic and physical cues. We analyzed segmentation behavior in terms of the ability to identify coarse and fine event boundaries and the resulting hierarchical structure. Non-native listeners identified almost identical coarse event boundaries as native listeners, but missed some of the fine event boundaries identified by the native listeners. Interestingly, hierarchical event perception (as measured by hierarchical alignment and enclosure) was comparable for native and non-native participants. In summary, linguistic cues contributed particularly to the identification of certain fine event boundaries. The results are discussed with regard to the current theories of event cognition.

KEYWORDS

event segmentation,
event cognition,
linguistic information,
event hierarchy

INTRODUCTION

When watching movies or listening to audio dramas, observers have to mentally represent the dynamically unfolding plot in order to understand the storyline. The central challenge for human information processing is thus coping with the complexity of the dynamic information resulting from the constantly changing sensory input. Although the sensory input is continuous, perception is not. Instead, the continuous stream of information is segmented into discrete events (Newtson, 1973); these events are separated from each other by *event boundaries*.

Recent research has shown that event boundary perception is triggered by semantic changes in the plot. Changes in one or more dimensions in the current situation model (Zwaan, Langston, & Graesser, 1995) are typically related to event boundary perception processes (Huff et al., 2018; Huff, Meitz, & Papenmeier, 2014; Zacks, Speer, & Reynolds, 2009). Yet, in auditory processing, there are at least two kinds of mean-

Corresponding author: Frank Papenmeier, University of Tübingen, Schleichstraße 4, D-72076 Tübingen, Germany. E-mail: frank.papenmeier@uni-tuebingen.de

ingful semantic information: linguistic (i.e., language based) and non-linguistic (i.e., sound information, such as a ringing phone). Moreover, purely perceptual changes of physical dimensions such as brightness or loudness can also influence segmentation behavior (Cutting, Brunick, & Candan, 2012; Sridharan, Levitin, Chafe, Berger, & Menon, 2007). In an event segmentation experiment, we investigated the contribution of linguistic semantic information to event segmentation in audio dramas. In particular, we compared the segmentation behavior of native speakers that could utilize both linguistic and non-linguistic semantic information, as well as physical cues for segmentation, with the segmentation behavior of non-native speakers, that could utilize only non-linguistic semantic information and physical cues, but not linguistic semantic information, for segmentation.

Event Segmentation

Newtson (1973; Newtson & Engquist, 1976) introduced a simple paradigm to study how humans unitize the continuous stream of information into discrete events. It involves asking participants to watch a dynamic scene (e.g., a video clip depicting an everyday action) and instructs them to press a button whenever they think that one meaningful event has ended and another meaningful event has begun. This results in a series of discrete events divided by event boundaries. Typically, agreement among participants in defining the ending and starting points of an event is remarkably high (Newtson & Engquist, 1976). This high intersubject agreement is probably because event boundaries often depict changes in the plot. These changes can be based on semantic information, such as intentions and goals of an actor (Speer, Zacks, & Reynolds, 2007), or on physical changes, such as changing brightness, changing object trajectories, or changing views of a scene (Baker & Levin, 2015; Cutting, DeLong, & Brunick, 2011; Hard, Tversky, & Lang, 2006; Zacks, 2004).

Several segmentation grains are reported in the literature. The most prominent ones are *fine-segmentation*, in which participants are instructed to identify the boundaries between the smallest meaningful events, and *coarse-segmentation*, in which participants are instructed to identify the boundaries between the largest meaningful events. Events are structured hierarchically, meaning that several fine events are grouped within one coarse event (Hard et al., 2006; Kurby & Zacks, 2011; Zacks, Tversky, & Iyer, 2001). The ability to structure events hierarchically is important for an appropriate encoding of the perceived action in memory (Swallow, Zacks, & Abrams, 2009; Zacks, Swallow, Vettel, & McAvoy, 2006; Zacks & Tversky, 2001).

The Basis of Event Boundary Perception

Previous research provides evidence for two possible triggers of event segmentation. People are able to segment an ongoing activity based on semantic (Zacks, Speer, Swallow, Braver, & Reynolds, 2007) or physical changes (Cutting et al., 2012).

Semantic changes are typically conceptualized as changes in the dimensions of the present situation model or event model (Radvansky & Zacks, 2011; Zwaan et al., 1995). Whereas research focusing on

text comprehension processes uses the dimensions of time, space, protagonist, intentionality, and causality (Zwaan et al., 1995), more recent work utilizing filmic stimulus material uses the dimensions of time, space, protagonist, and action to describe the unfolding narrative plot (Huff et al., 2014; Magliano & Zacks, 2011). For auditory dynamic events such as audio dramas, we suggest the existence of the additional dimensions of music and narrator (Huff et al., 2018). Importantly, naturalistic auditory events consist of a speech signal and a nonspeech signal. The speech signal is rich in information for the listeners, as it transmits linguistic information on multiple temporal scales (Poeppel, Idsardi, & van Wassenhove, 2008; Rosen, 1992). Within this project, we focus on the language-based (i.e., semantic and syntactic) aspects of linguistic information. In this context, linguistic information can introduce changes in the plot. For example, the narrator can introduce the appearance of a new character as follows: "Simon enters the living room." The nonspeech signal, on the other hand, transmits information that is not language-based, such as the sound accompanying the action of answering a ringing phone. We refer to this kind of auditory information as non-linguistic information.

According to current theories of event comprehension, perception is guided by event models, which represent the ongoing observed behavior in working memory (Radvansky & Zacks, 2011; Zacks, Speer, et al., 2009). These event models can be conceived as situation models from text comprehension research (van Dijk & Kintsch, 1983; Zacks, Speer, et al., 2009; Zwaan et al., 1995) and they represent the current state of the situation in working memory. Humans predict the future state of an action or situation based on this event model. An error detection mechanism monitors deviations between the predicted state of the observed action and its actual state. In case of significant deviations, the event model is updated and event boundaries are perceived.

Event segmentation is hierarchical and operates at multiple segmentation grains, such as the aforementioned fine-segmentation and coarse-segmentation. Yet, processing of coarse and fine boundaries differs in some respects. First, as revealed by an fMRI study, compared to fine event boundaries, coarse event boundaries elicit significantly larger responses in regions of the brain typically associated with narrative event boundaries (Speer et al., 2007). Second, there is evidence that the inability to process semantic information harms coarse-grained segmentation more than it does fine-grained segmentation. Zalla, Pradat-Diehl, and Sirigu (2003) examined the ability to identify coarse and fine events in patients with prefrontal lesions. Compared to the healthy control group, the patients had more difficulties in detecting coarse-grained event boundaries, but showed a comparable performance in detecting fine-grained event boundaries. Nonetheless, both groups identified significantly more fine than coarse event boundaries–which is a prerequisite for hierarchical event perception. Further evidence for the role of semantic information in the detection of coarse event boundaries comes from studies examining segmentation behavior in patients with schizophrenia, which is also characterized by an impairment in semantic and phonological fluency (Gourovitch, Goldberg, & Weinberger, 1996). These patients also have trouble identifying the coarse events within an observed action (Zalla, Verlut,

Franck, Puzenat, & Sirigu, 2004) and they perform poorly in tasks that are sensitive to lesions of the frontal regions (Kolb & Whishaw, 1983; Liddle & Morris, 1991).

Further evidence for the interplay of semantic information and event boundary perception comes from research on the cognitive processing of sign language. Malaia, Ranaweera, Wilbur, and Talavage (2012) studied the neural activation of deaf signers and hearing non-signers while observing American Sign Language, with special attention put on the discrimination between telic and atelic verb signs. A telic verb "predicates denoting events with an inherent end-point representing a change of state (*break*, *appear*)" (Malaia et al., 2012, p. 4094). In contrast, an atelic verb describes events which do not directly implicate an end and a starting point (e.g., *swim*, *sew*; Malaia et al., 2012). For deaf signers, but not for hearing nonsigners, telic verbs, as compared to atelic verbs, elicited an increased activation in the brain regions (especially the precuneus) that are also involved in event boundary perception (Malaia et al., 2012; Speer et al., 2007; Zacks, Braver, et al., 2001; Zacks, Swallow, et al., 2006). Because telic and atelic signs also differ in their physical properties such as peak velocity and deceleration (Malaia & Wilbur, 2012), both the physical properties of the signal or the semantic processing might contribute to event boundary perception.

These findings are consistent with research studying neural processing of telic and atelic verbs during the comprehension of written texts. Research using event-related potentials suggests that telic verbs (in contrast to atelic verbs) facilitate online processing of sentences by means of consolidating event schemas in episodic memory (Malaia & Newman, 2015). Thus, semantic information by means of verbal telicity might provide information that is necessary for event boundary perception.

In sum, there is evidence that semantic cues influence event boundary perception. Linguistic information, in particular within an audio drama, provides the main part of semantic information of a story line. We therefore assume that manipulating the processing of linguistic semantic information of an audio drama should have similar effects as was observed in the studies described above. Thus, we predict a stronger contribution of linguistic semantic information on the perception of coarse event boundaries than fine event boundaries.

Semantic and Non-Semantic Cues in a Single Experiment

The reviewed literature reveals evidence for both semantic and non-semantic influences on event boundary perception. Zacks, Kumar, Abrams, and Mehta (2009) recorded everyday activities using a magnetic motion tracking system and designed abstract versions of the videos. Colored spheres that were connected by solid lines represented the heads and hands of the actors. One half of the participants were provided with semantic information on the origin of the movement pattern (live video preview of a person performing the respective activity), while the other half remained uninformed. Surprisingly, segmentation behavior did not differ between these two experimental

conditions. Thus, this study adds further support for the importance of low-level, bottom-up processes (see also Hard et al., 2006).

For several reasons, however, investigating the contribution of semantic cues on processes of event segmentation is difficult when using pictorial stimuli: First, when studying segmentation behavior during the perception of abstracted stimuli (such as balls representing the recorded movements of a person doing some everyday tasks), physical features (such as motion changes) dominate event boundary perception, which is then unaffected by prior semantic information on the origin of those motion patterns (Zacks, Kumar, et al., 2009). Thus, results obtained during the study of abstracted stimuli might not generalize to naturalistic scenes. Second, when using naturalistic visual or audiovisual stimulus material (such as edited movies), participants can extract the gist of the scene in less than 100 ms (Oliva, 2005; Potter, 1976). Thus, it is difficult to manipulate the occurrence of semantic processing when viewing naturalistic pictorial stimuli.

Auditory stimulus material such as audio dramas offer a step towards solving this problem. The processing of parts of the semantic cues (linguistic cues) can be hindered by inviting participants who do not understand and speak the language of the audio drama and their performance can be compared with those participants who fully understand the language. This makes it possible to study the contribution of linguistic semantic cues to event segmentation processes without using stimulus material that is too artificial.

Experimental Overview and Hypotheses

Our focus in the current study is the contribution of linguistic semantic information to event segmentation processes. We invited two groups of participants. One group spoke and understood the language of the audio drama and was thus able to process the linguistic information. We refer to this group as *native* participants. The second group of participants neither spoke nor understood the language of the audio drama, or any other language of the Germanic language family (Gray & Atkinson, 2003). Therefore, these *non-native* participants were not able to process the linguistic semantic information of the auditory signal. The latter group could therefore only use non-linguistic information and changes in physical attributes to segment the plot into distinct events. The only difference between the native and non-native participants was whether they could use linguistic semantic information or not.

If event segmentation is only based on physical cues (Cutting et al., 2012; Hard et al., 2006), we should not to find a difference in segmentation behavior between native and non-native participants. In contrast, if linguistic semantic information contributes to the detection of event boundaries, we should observe differences in segmentation behavior between the two experimental groups. To be more specific, if linguistic cues are important for the perception of a hierarchically structured event, we expect that non-native participants' event segmentation behavior is less hierarchically structured as compared to native participants' segmentation behavior (Hypothesis 1). In Hypothesis 2, we expect that this lower hierarchy can be traced back to the fact that non-

native participants have difficulties in perceiving coarse event boundaries because in order to identify them, linguistic semantic changes are crucial (Zalla et al., 2004).

METHOD

Participants

We recruited 68 students from the University of Tübingen for participation in this study. Participants received monetary compensation or course credit for participation. No participant reported any hearing difficulties. Thirty-four participants were German native speakers (31 female; M_{age} = 21.82 years, SD_{age} = 4.65, range 18–42 years). Thirty-four participants neither spoke nor understood German (23 female; M_{age} = 24.88 years, SD_{age} = 4.06, range 20–36 years). The non-native speakers' mother tongues were: Spanish (n = 3), Turkish (n = 4), Mandarin (n = 1), Hindi (n = 5), Korean (n = 4), Sindhi (n = 1), Croatian (n = 1), Chinese (n = 2), Polish (n = 1), Finnish (n = 2), Tamil (n = 1), Slovak (n = 1), Romanian (n = 1), Italian (n = 2), Arabic (n = 1), and English (n = 4). Data of the English native speakers were collected by mistake. Originally, we did not plan to collect data from native speakers of any language that is part of the Germanic language family (Gray & Atkinson, 2003). Thus, we excluded the data of the four English native speakers before data analysis.

In addition, due to technical problems, we lost one fine segmentation dataset in the native condition, two fine, and four coarse segmentation datasets in the non-native condition. Thus, the final dataset included data of 33 fine and 34 coarse datasets in the native condition as well as 28 fine and 26 coarse datasets in the non-native condition. This dataset is available as open data at https://osf.io/rp5bg/.

To ensure that the non-native participants actually did not understand German, they completed a language comprehension test. The participants were asked to describe 20 terms that were part of the audio drama, such as Telefonbuch (telephone book). An answer received one point if the participants knew the meaning of the word and two points if the participants additionally knew its location in the storyline. Thus, the score of this test ranged from 0 (i.e., no words understood) to 40 (i.e., all words, including the narrative, understood). The results showed that non-native-participants' understanding was poor, M = 8.20, SD = 3.58.

The research was conducted in accordance with APA standards for the ethical treatment of participants and we gained informed consent from the participants.

Material and Procedure

After giving their informed consent, the participants were seated in front of a notebook computer (Lenovo R400 7439-CTO) with headphones attached. On this notebook, we presented the instruction for the segmentation task as well as the audio drama. The experiment was programmed in PsychoPy (Peirce, 2007). As experimental stimulus material, we used the episode *Die drei Fragezeichen und das Gespensterschloss* [The Three Investigators and the Secret of Terror Castle]. As an audio drama, this stimulus consists not only of spoken language by different people, but also of environmental sounds, such as car horns, footsteps, dogs barking, or the ringing of a telephone. In addition, the original stimulus included passages in which a narrator repeated something, sometimes accompanied by music. Since those passages are not important for the comprehension of the plot, we edited the episode such that it contained no long music passages and only short segments in which the narrator repeated the previous plot. These remaining segments lasted from about three to five seconds. The resulting audio drama was 44 min 8 s long.

Participants were instructed to segment the audio drama into the smallest and largest units that are meaningful to them. In one session, each participant segmented the audio drama into the smallest meaningful units (fine-grained segmentation) and, in a second session, into the largest meaningful units (coarse-grained segmentation). The order of segmentation grain was counterbalanced across the participants. The delay between the two sessions was at least 24 hours.

The procedure was similar for both the native and the non-native participants. The native participants received the informed consent form and the instructions in German. After having ensured that the non-native participants did not understand German, they received the same informed consent form and the instructions in English. Because the non-native participants were international students at the University of Tübingen, their understanding of English was adequate. Afterwards, the participants listened to the same audio drama as the native participants.

Event Segmentation Measures

We employed four measures that focused on different aspects of event segmentation behavior. The number of event boundaries measure focused on the raw number of button presses, irrespective of their timing. The hierarchical alignment and the hierarchical enclosure measures examined the temporal relation of fine and coarse boundaries at those moments in time where the participants perceived coarse event boundaries. The segmentation agreement measure focused on the relation of the individual participants' button presses to a respective group norm.

NUMBER OF EVENT BOUNDARIES

For each language and segmentation grain condition, we counted the number of key presses.

HIERARCHICAL ALIGNMENT

Hierarchical alignment is a standard measure of event perception research indicating whether the temporal alignment of coarse and fine event boundaries is closer than what would be expected by chance (Hard, Recchia, & Tversky, 2011; Hard et al., 2006; Sargent et al., 2013; Zacks, Tversky, et al., 2001). We first computed the observed average distance between all coarse boundaries and their nearest fine boundaries for each participant. Then, we computed the expected average distance between coarse boundaries and their nearest fine boundaries, given chance as indicated by the null model prediction given

by Equation 3 described in Zacks, Tversky, et al. (2001). In order to compare hierarchical alignment across language groups, we calculated the difference between the expected and observed average distances and divided those differences by the expected average distance for each participant. Thus, we accounted for varying numbers of perceived boundaries for each participant (see Hard et al., 2006).

HIERARCHICAL ENCLOSURE

Hierarchical enclosure is another standard measure of event perception used to investigate the hierarchical structure between coarse and fine event boundaries (Hard et al., 2011, 2006; Sargent et al., 2013). If event boundaries are perceived hierarchically, coarse event boundaries should enclose and thus follow (rather than precede) fine event boundaries. For each participant, we first identified the nearest fine boundary for each coarse boundary. Then, we calculated enclosure by computing the proportion of coarse boundaries that followed their nearest fine boundary. Typically, enclosure is compared against 0.5 (chance level).

SEGMENTATION AGREEMENT

Segmentation agreement is a measure indicating how closely a participant's segmentation responses resemble the event structure identified by a group of participants (Kurby & Zacks, 2011; Sargent et al., 2013; Zacks, Speer, Vettel, & Jacoby, 2006). Segmentation agreement is computed using the following steps. First, the stimulus is divided into 1-second bins. Second, for each participant, it is determined whether at least one segmentation response occurred or not within each bin. Third, the group norm is defined as the proportion of participants segmenting in each bin. Fourth, each participants' responses are correlated with the group norm. Fifth, in order to account for the fact that this correlation depends on the number of segmentation responses of a participant, segmentation agreement is defined as $(r_{obs} - r_{min}) / (r_{max} - r_{min})$, with r_{obs} being the observed correlation and r_{min} and r_{max} being the minimum and maximum correlation possible given the number of button presses (Kurby & Zacks, 2011; Sargent et al., 2013). In contrast to previous research, we applied two changes to the computation of segmentation agreement in order to adopt this measure to our design. First, we computed two distinct reference group norms, one based on the responses of the native participants and another one based on the responses of the non-native participants. This was done for two reasons. First, it should not be assumed a priori that the event structure identified by non-native participants, who cannot use the linguistic semantic information of the audio drama, comes from the same population as the event structure identified by native participants. Second, we can compute two segmentation agreement scores for each reference group norm, one for the native participants and another one for the non-native participants. This provides an index of how closely the event boundaries identified by the two groups match. Second, we excluded each participant's own responses from the respective reference group norm when calculating the correlation. This was done to prevent the issue of spurious self-correlations that would otherwise occur when computing segmentation agreement against the reference group norm of the participants' own language group. Similar to previous research, we ran the segmentation agreement analysis for fine and coarse event boundaries separately.

RESULTS

Prior to the analysis, we excluded all responses that occurred more than 2 s after the end of the audio drama (5 responses, 0.07% of data). Because some of our measures (hierarchical alignment and hierarchical enclosure) are calculated based on fine and coarse segmentation responses, we excluded the data of all participants where either the fine or coarse segmentation data was missing due to technical problems (one native participant and four non-native participants). Finally, we excluded one non-native participant that produced more coarse responses than fine responses. Thus, the final data set entering our analysis contained the data of 33 native participants and 24 non-native participants.

Segmentation Behavior Analysis

NUMBER OF EVENT BOUNDARIES

We first analyzed the number of identified event boundaries by fitting a Poisson generalized linear mixed using the glmer function of the lme4 package in R (Bates et al., 2014). This model included Number of Event Boundaries (i.e,. key presses) as the dependent variable as well as the main effects for Grain (coarse, fine) and Language (native, non-native) and their interaction as fixed effects. We used the maximal random effects structure for this model (Barr, Levy, Scheepers, & Tily, 2013), that is, we included a random intercept and grain effect for participant to the model. We submitted the model to a type II analysis of variance (ANOVA) using the Anova function of the car package (Fox & Weisberg, 2010) in R. As depicted in Figure 1, we found a main effect of segmentation grain, indicating that the participants identified more fine than coarse event boundaries, $\chi2(1) = 247.29$, $p < .001$. The main effect for language was not significant, $\chi2(1) = 1.85$, $p = .173$. Importantly, there was also a significant interaction of grain and language, $\chi2(1) = 15.05$, $p < .001$. We used two-sample Wilcoxon rank sum tests to further investigate the interaction. While native participants perceived significantly more fine event boundaries than did non-native participants, $W = 541.0$, $p = .019$, there was no significant difference between the two groups of participants for coarse event boundaries, $W = 373.5$, $p = .722$. Thus, although this first analysis provided no information on the location of the segmentation responses, it gives a first indication that any differences in segmentation behavior might be observed at fine segmentation grain.

HIERARCHICAL ALIGNMENT

For both, native and non-native participants, the observed distance of coarse boundaries to their closest fine boundaries (native: $M = 5.10$ s, $SD = 7.77$ s, non-native: $M = 14.82$ s, $SD = 20.35$ s) was, on average, closer than expected by chance (null model; native: $M = 19.13$ s, $SD = 19.16$ s, non-native: $M = 29.33$ s,

SD = 18.94 s). This difference between observed and expected coarse to fine distances was significant for native participants, $t(32) = 5.11$, $p < .001$, and non-native participants, $t(23) = 4.00$, $p = .001$. Thus, native and non-native participants structured coarse and fine event boundaries hierarchically.

In order to compare hierarchical alignment across native and non-native participants, we corrected the alignment scores for the number of perceived boundaries (see Method section). Hierarchical alignment between native participants ($M = 0.66$, $SD = 0.25$) and non-native participants ($M = 0.49$, $SD = 0.59$) did not differ significantly as indicated by an unequal variances *t* test, $t(29.37) = 1.34$, $p = .192$. Thus, the degree to which non-native participants hierarchically aligned coarse and fine event boundaries was comparable to native participants, even though non-native participants could not utilize the linguistic semantic cues available in the audio drama.

HIERARCHICAL ENCLOSURE

Both native ($M = 0.74$, $SD = 0.17$) and non-native ($M = 0.70$, $SD = 0.20$) participants' enclosure measures were significantly higher than 0.5 (chance level), $t(32) = 8.21$, $p < .001$, and $t(23) = 4.77$, $p < .001$, respectively. Again, the hierarchical enclosure did not differ across native and non-native participants, $t(55) = 0.80$, $p = .426$.

In summary, the two measures exploring the hierarchical structure of segmentation behavior (alignment and enclosure) showed no differences between the native and non-native participants. However, it is important to note that these measures only included those fine event boundaries that are next to a coarse event boundary. As we have already shown in the Number of event boundaries analysis section, linguistic information probably contributes to the perception of

a subset of fine event boundaries that were not perceived by the non-native participants. Based on the results of the alignment and enclosure analyses, we presume that these might be those event boundaries that are perceived in the midst of an event (and do not have a neighboring coarse event boundary). Thus, in the next step, we analyzed the segmentation agreement in order to overcome the shortcomings of the hierarchical analyses.

SEGMENTATION AGREEMENT

We analyzed the segmentation agreement (see Figure 2) using a mixed ANOVA with the factors of Segmentation Grain (fine, coarse; within), Reference Group Norm (native group norm, non-native group norm; within), and Language (native, non-native; between). We observed a significant three-way interaction of grain, reference group norm, and language, $F(1, 55) = 4.14$, $p = .047$, $\eta_p^2 = .07$. Bonferroni corrected post hoc t tests with an adjusted α level of .0125 showed that native and non-native participants differed significantly in their segmentation agreement against the native group norm for fine boundaries, $p = .012$, but not for coarse boundaries, $p = .133$. When computing the segmentation agreement against the non-native group norm, there was no significant difference between native and non-native participants for both fine boundaries, $p = .281$, and coarse boundaries, $p = .816$. The comparable segmentation agreement of native and non-native participants with the group norm based on non-native responses indicated that the event boundaries identified by non-native participants were a subset of the event boundaries identified by native participants. Thus, it seems that non-native participants were well able to extract the event structure of the audio drama despite not understanding the linguistic semantic information provided. For the native group norm, in contrast,
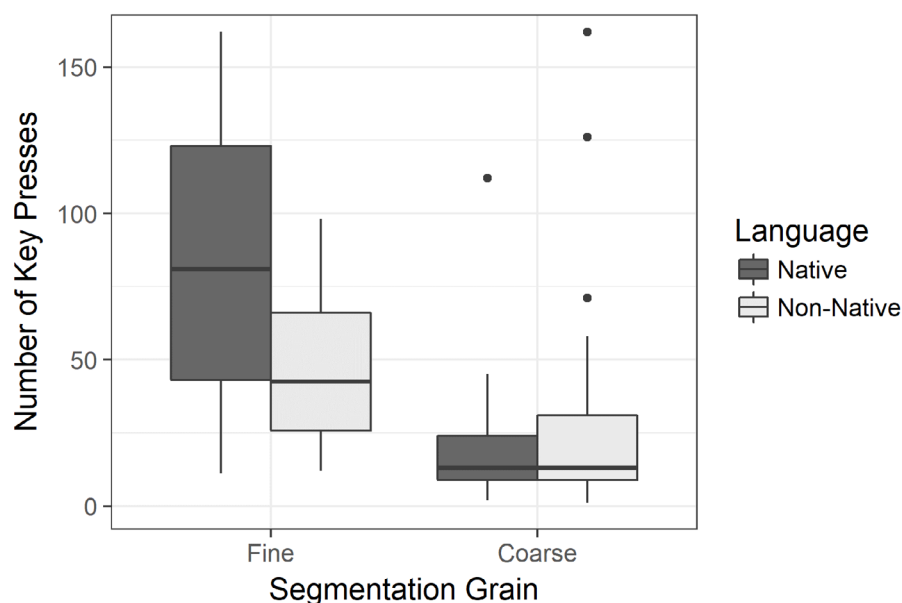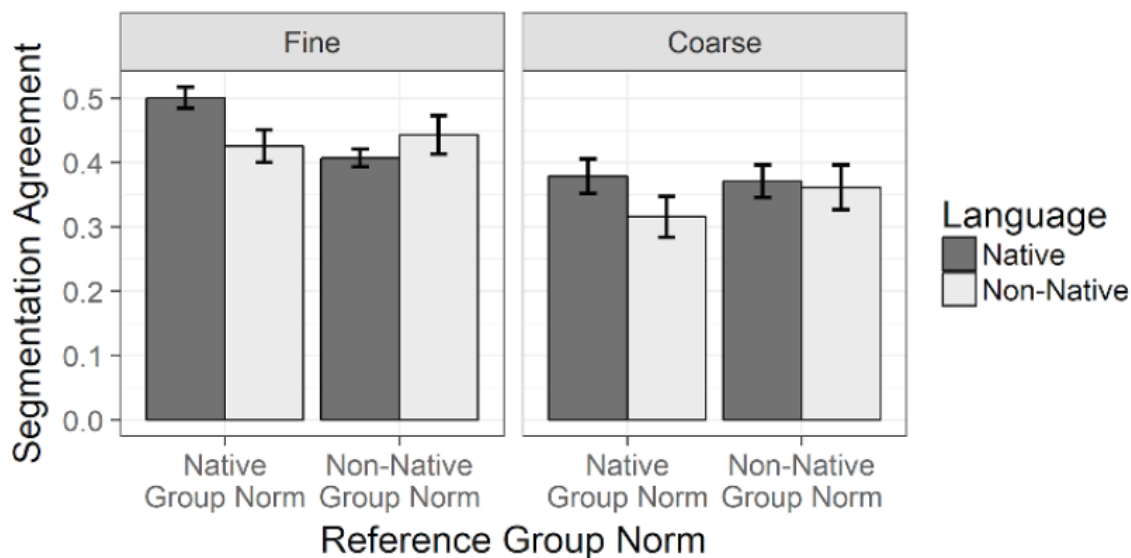

FIGURE 1.

Boxplots depicting the number of identified event boundaries (i.e., key presses) for native and non-native participants as a function of segmentation grain (fine, coarse). The following outliers are not visible in the figure: native fine 313, 378, 561, and 655; non-native fine 169, 310, and 387; native coarse 280.

FIGURE 2.

The segmentation agreement as a function of segmentation grain (left: fine boundaries; right: coarse boundaries), language and reference group norm. Error bars represent the SEM

the native participants showed a higher agreement than did non-native participants for the fine, but not coarse, boundaries. This conforms with our number of event boundaries analysis presented above and shows that there is an additional fine-grained event structure in the audio drama that could only be captured with the access to linguistic semantic information. The other effects of the ANOVA were as follows. There was a significant effect of grain, $F(1, 55) = 28.88$, $p < .001$, $\eta_p^2 = .34$, an interaction of language and reference group norm, $F(1,55)=24.00$, $p<.001$, $\eta_p^2=.30$, and of grain and reference group norm, $F(1, 55) = 19.26$, $p < .001$, $\eta_p^2 = .26$. The main effects of language, $F(1,55)=0.85$, $p=.360$, $\eta_p^2=.02$, and reference group norm, $F(1,55)=3.71$, $p = .059$, $\eta_p^2 = .06$, as well as the interaction of language and grain, $F(1, 55) = 0.28$, $p = .602$, $\eta_p^2 < .01$, were not significant.

DISCUSSION

Event segmentation is a means to study how humans perceive dynamic events (Newtson, 1973). Recent research provided evidence that physical and semantic changes are points in time at which humans perceive an event boundary (Huff et al., 2014; Magliano & Zacks, 2011; Zacks, 2004). Yet in auditory dynamic events, there are at least two kinds of semantic information: linguistic (i.e., language-based, spoken information) and non-linguistic (e.g., a ringing phone). We assumed that in an audio drama, linguistic information constitutes a main part of the semantic content. Based on this, the present study investigated the importance of linguistic semantic information for perceiving a dynamic event by studying the contribution of linguistic cues to the event segmentation process. We compared native (German) speaking and non-native (non-German) speaking participants' segmentation behavior. Participants listened to a German audio drama and were asked to indicate the temporal location of fine and coarse event boundaries.

Whereas native participants could use linguistic, non-linguistic, and physical cues, non-native participants could only use the latter two (i.e., non-linguistic and physical) cues.

We tested two hypotheses: First, the hierarchical structure of event perception (i.e., participants perceiving multiple fine event boundaries within one coarse event) should be less pronounced if linguistic semantic information cannot be processed. However, two classical measures of hierarchical event perception – hierarchical alignment and hierarchical enclosure – did not show any differences between native and non-native participants' event segmentation behavior. Thus, contrary to our expectations, the absence of linguistic semantic information did not impair the perception of the hierarchical event structure. This indicates that listeners extracted the hierarchical event structure of the audio drama based on non-linguistic semantic information and physical cues rather than linguistic semantic information.

In the second hypothesis, we expected that reduced hierarchical event perception in non-native participants can be traced back to difficulties in perceiving coarse event boundaries. This assumption was based on the idea that linguistic information (semantic and syntactic) within an audio drama provides the main part of semantic content, and is, therefore, important in perceiving coarse rather than fine event boundaries. Earlier studies on semantic understanding for event segmentation have shown an influence on coarse-grained segmentation (e.g., Zalla et al., 2003). In contrast to our expectation, the analyses showed that linguistic information is crucial for the perception of fine but not coarse event boundaries. This was true for the number of identified event boundaries and segmentation agreement. Because the measures of hierarchical alignment and hierarchical enclosure – which, by definition, disregard any fine event boundaries in the midst of an event (i.e., without a neighboring coarse event boundary) – were not affected by our manipulation, we conclude that linguistic semantic

information contributes to the detection of fine event boundaries in the midst of an event. We can only speculate about the reasons for this unexpected finding. It is possible that the coarse boundaries were more prominent and accompanied by respective non-linguistic semantic and physical cues, thus making the linguistic semantic information obsolete for the detection of coarse but not fine boundaries occurring in the midst of an event. Nonetheless, it is important to note that our results show that linguistic semantic information is important for the detection of at least some kinds of fine boundaries. Further research manipulating all three types of information (linguistic, non-linguistic and physical) is required to obtain further insights on how the sum or interaction of all three types of information contributes to the segmentation of audio dramas into fine and coarse events.

This study also relates to research on top-down effects on event segmentation (Huff, Papenmeier, Maurer, Meitz, Garsoffky, & Schwan, 2017). Previous research reported that participants identified fewer event boundaries when they were familiar with an activity (Hard et al., 2006) or when the action goals were predictable (Wilder, 1978; Zacks, 2004). Obviously, native participants are more familiar with the language of the audio drama than non-native participants, who were not able to process the semantic and syntactic information of the auditory event. Thus, native participants' language-based predictions are presumably more reliable than those of the non-native participants. One might have therefore expected that native participants should have identified fewer event boundaries as compared to non-native participants. However, we observed the opposite: Native participants identified more fine event boundaries than did non-native participants. One needs to take into account, however, that the unfamiliarity of non-natives with the language of the audio drama prevented them not only from making predictions about upcoming information, but also from detecting semantic changes in the speech stream, thus possibly resulting in the detection of fewer instead of more boundaries.

Our results are related to findings from (psycho-)linguistic research on second language (L2) learning, which has shown that non-native speakers, as compared to native speakers, are less sensitive to syntactic information and rely more on lexical-semantic and other nonsyntactic cues (Marinis, Roberts, Felser, & Clahsen, 2005). In a reading time experiment that explored the processing and comprehension of wh-phrases, native, but not non-native, speakers made use of intermediate syntactic gaps (Marinis et al., 2005). Further, listeners were less able to notice speaker changes and identify speaker identities in a foreign language (Perrachione, Chiao, & Wong, 2010; Perrachione & Wong, 2007; Wester, 2012). These findings might explain the findings of the present study – compared to native speakers, non-native speakers, who were not able to process semantic and syntactic information, showed different event segmentation behavior for fine event boundaries. It is important to note, however, that studies L2 learning typically study participants who are well able to understand the language of the subject matter (Marinis et al., 2005; Perrachione et al., 2010). In contrast, in the present experiment, we studied participants who neither spoke nor understood the language of the subject matter. A promising direction for future research would be to combine the approaches and to

study segmentation behavior of non-native listeners with different levels of proficiency. Such a study would provide the missing link between event cognition and L2 learning research.

Our results have implications for the event segmentation theory (EST; Zacks et al., 2007). According to EST, event perception is hierarchically structured and guided by event models which represent the current state of the situation. These event models guide perception by facilitating predictions about the near future of the action. If the predictions no longer match the actual state of the plot, either on the fine or coarse level, participants perceive an event boundary. Apart from the number of key presses, which does not consider the temporal dimension of the underlying event, there are two measures reported in the literature, namely, hierarchical alignment (Zacks, Tversky, et al., 2001) and hierarchical enclosure (Hard et al., 2006), that examine the hierarchy of the event structure by considering the temporal dimension. Both measures focus on the relation between fine and coarse event boundaries at the time points when participants perceive coarse event boundaries, thus disregarding any fine event boundaries in the midst of an event (i.e., without a neighboring coarse event boundary). The segmentation agreement, in contrast, takes the timing of all event boundaries into account. Using the combination of these measures, we conclude that the perception of subevents, which are separated by fine event boundaries, is triggered by linguistic cues.

Despite the strong impact of our language manipulation on top-down processing, event segmentation was surprisingly robust. Our results suggest that linguistic semantic information contributed neither to the perception of coarse event boundaries nor to the extraction of the hierarchical event structure as measured by hierarchical alignment and hierarchical enclosure. This finding adds to the recent discussion on the relative contribution of top-down and bottom-up processes to event perception (Brockhoff, Huff, Maurer, & Papenmeier, 2016; Cutting et al., 2012; Huff & Papenmeier, 2017; Huff, et al., 2017; Zacks, Kumar, et al., 2009; Zacks et al., 2007; Zacks, Tversky, et al., 2001). Based on our results, one might argue that bottom-up processes contribute more to event segmentation than do top-down processes. In order to further support this view, however, it would be important to also remove the non-linguistic semantic cues from the auditory event, for example by using an audio book instead of an audio drama as stimulus material.

Limitations and Further Research

In the present study, we used an audio drama as stimulus material to study the influences of linguistic cues on event boundary perception processes. Using (audio-) visual stimulus material, it is not possible to manipulate the processing of semantics to a high degree – unless testing neurological or psychiatric patients, as done in the studies by Zalla et al. (2003). Therefore, the present study is an important step in event perception research, as it allowed us to determine processes related to event perception in healthy participants. One important point to consider when interpreting these data is that we used just one single stimulus to explore how linguistic information contributes to event perception processes. We used this specific audio drama because it has

many dimension changes and it has just one narrative thread with no jumps in time (i.e., flash backs). From this point of view, this audio drama fulfills most of the criteria which made *The Red Balloon* (Lamorisse, 1956) very popular in cognitive research (Baggett, 1979; Zacks, Speer, et al., 2009). More research using a broader range of stimulus materials (such as audio books or recorded conversations) is needed to explore the boundary conditions of the effects described in this project.

An important point concerns the question whether the native and non-native participants actually segmented at the same grain. Although the segmentation instructions were matched for content, native participants were instructed in their mother tongue and non-native participants were instructed in English, their L2. Thus, task difficulty might have been different for native and non-native participants, and we cannot rule out that native and non-native participants actually segmented at different grains as a consequence. One way to overcome this problem in future research is to establish a practice phase right before the actual experiment. The goal of the practice phase would be to equate segmentation across the experimental groups (see also Sebastian, Ghose, Zacks, & Huff, 2017).

Further, it is worth considering that the relative importance of linguistic and non-linguistic semantic information might be different for native and non-native participants. In particular, it seems reasonable to assume that non-linguistic semantic information is more important for non-native participants because they are not able to process semantic and syntactic information. Future research is necessary to study this question. One approach to study the influence of the relative importance of linguistic and non-linguistic semantic cues on event segmentation behavior is to ask participants to segment the different parts of an audio drama (soundtrack with conversations only, soundtrack with ambient noise only) separately. If the relative importance of processing non-linguistic semantic information is different for native and non-native participants, one might expect different segmentation patterns for the soundtrack with ambient noise parts. If, however, the relative importance of linguistic and non-linguistic semantic information is not different between native and non-native participants, one might expect differences only in the segmentation patterns of the soundtrack with the conversation parts.

Finally, this first study on the role of linguistic information for comprehending dynamic auditory events opens new research opportunities at the intersection of event segmentation and linguistics. A possible starting point is the observation that native participants identified more fine event boundaries than did non-native participants and the possible follow-up question about the exact location of those event boundaries. Did non-native participants segment mostly between two words or did they also identify event boundaries between two syllables within a single word? If this effect is due to language-based differences in signal processing between native and non-native participants, we expect pronounced differences in a measure that includes the relative differences of native and non-native participants identifying word boundaries (i.e., event boundaries that are placed between two words) versus syllable boundaries (i.e., event boundaries that are placed between two syllables within a single word). Yet, the present

stimulus material prevented such a detailed analysis of the location of the boundaries, because the speech signal is mixed with ambient noise and music. Thus, it is not possible to link a specific event boundary with a single word or even syllable. Future research should use audio books as stimulus material, because there is no competing auditory information. Such a study would allow for tracing back the observed differences to semantic, language-based knowledge, while processing dynamic auditory events. In addition, it is important to note that the dependent measures used in the studies were based on the event segmentation task and did not account for the syntactic hierarchy. A further direction in future research could be the study of syntactic hierarchy and its contribution to event perception processes.

CONCLUSION

When listening to auditory narratives such as audio dramas, humans use linguistic, non-linguistic, and physical cues for event boundary perception. Surprisingly, listeners extracted the hierarchical structure of the audio drama also in the absence of linguistic semantic information. However, if linguistic semantic information could not be processed, fine, but not coarse, event boundary perception was impaired. Thus, linguistic semantic information contributes to the detection of certain fine event boundaries, presumably the ones located in the middle of a (coarse) event.

AUTHORS' CONTRIBUTIONS

A. Maurer developed the study concept and was responsible for testing and data collection. A. Maurer and M. Huff contributed to the study design. F. Papenmeier performed the data analysis. All authors were responsible for the interpretation of the results. All authors equally contributed to the preparation of the manuscript. All authors approved the final version of the manuscript for submission.

REFERENCES

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*, 333–356. doi:10.1016/S0022-5371(79)90191-9

Baker, L. J., & Levin, D. T. (2015). The role of relational triggers in event perception. *Cognition, 136*, 14–29. doi:10.1016/j.cognition.2014.11.030

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*, 255–278. doi:10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Dai, B. (2014). lme4: Linear mixed-effects models using Eigen and S4 (Version 1.1-7). Retrieved from http://

cran.r-project.org/web/packages/lme4/index.html

Brockhoff, A., Huff, M., Maurer, A., & Papenmeier, F. (2016). Seeing the unseen? Illusory causal filling in FIFA referees, players, and novices. *Cognitive Research: Principles and Implications, 1*, 1–12. doi:10.1186/s41235-016-0008-5

Cutting, J. E., Brunick, K. L., & Candan, A. (2012). Perceiving event dynamics and parsing Hollywood films. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 1–15. doi:10.1037/a0027737

Cutting, J. E., DeLong, J. E., & Brunick, K. L. (2011). Visual activity in Hollywood film: 1935 to 2005 and beyond. *Psychology of Aesthetics, Creativity, and the Arts, 5*, 115-125. doi:10.1037/a0020995

Fox, J., & Weisberg, S. (2010). *An R companion to applied regression*. Thousand Oaks, CA: Sage.

Gourovitch, M. L., Goldberg, T. E., & Weinberger, D. R. (1996). Verbal fluency deficits in patients with schizophrenia: Semantic fluency is differentially impaired as compared with phonologic fluency. *Neuropsychology, 10*, 573–577. doi:10.1037/0894-4105.10.4.573

Gray, R. D., & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature, 426*, 435–439. doi:10.1038/nature02029

Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General, 140*, 586–604. doi:10.1037/a0024310

Hard, B. M., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: Building event schemas. *Memory & Cognition, 34*, 1221–1235. doi:10.3758/BF03193267

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*, 307–320. doi:10.1037/xlm0000482

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*, 1377–1388. doi:10.1037/a0036780

Huff, M., & Papenmeier, F. (2017). Event perception: From event boundaries to ongoing events. *Journal of Applied Research in Memory and Cognition, 6*, 129–132. doi:10.1016/j.jarmac.2017.01.003

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. doi:10.1038/srep43083

Kolb, B., & Whishaw, I. Q. (1983). Performance of schizophrenic patients on tests sensitive to left or right frontal, temporal, or parietal function in neurological patients. *The Journal of Nervous and Mental Disease, 171*, 435–443. doi:10.1097/00005053-198307000-00008

Kurby, C. A., & Zacks, J. M. (2011). Age differences in the perception of hierarchical structure in events. *Memory & Cognition, 39*, 75–91. doi:10.3758/s13421-010-0027-2

Lamorisse, A. (1956). *The Red Balloon*. Films Montsouris.

Liddle, P. F., & Morris, D. L. (1991). Schizophrenic syndromes and frontal lobe performance. *The British Journal of Psychiatry, 158*, 340–345. doi:10.1192/bjp.158.3.340

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

Malaia, E., & Newman, S. (2015). Neural bases of event knowledge and syntax integration in comprehension of complex sentences. *Neurocase, 21*, 753–766. doi:10.1080/13554794.2014.989859

Malaia, E., Ranaweera, R., Wilbur, R. B., & Talavage, T. M. (2012). Event segmentation in a visual language: Neural bases of processing American Sign Language predicates. *NeuroImage, 59*, 4094–4101. doi:10.1016/j.neuroimage.2011.10.034

Malaia, E., & Wilbur, R. B. (2012). Kinematic signatures of telic and atelic events in ASL predicates. *Language and Speech, 55*, 407–421. doi:10.1177/0023830911422201

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. S*tudies in Second Language Acquisition, 27*, 53–78. doi:10.1017/S0272263105050035

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology, 28*, 28–38. doi:10.1037/h0035584

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology, 12*, 436–450. doi:10.1016/0022-1031(76)90076-7

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.)., *Neurobiology of Attention*, (pp. 251–257). Cambridge, MA: Academic Press. doi:10.1016/B978-012375731-9/50045-8

Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162*, 8–13. doi:10.1016/j.jneumeth.2006.11.017

Perrachione, T. K., Chiao, J. Y., & Wong, P. C. M. (2010). Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. *Cognition, 114*, 42–55. doi:10.1016/j.cognition.2009.08.012

Perrachione, T. K., & Wong, P. C. M. (2007). Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia, 45*, 1899–1910. doi:10.1016/j.neuropsychologia.2006.11.015

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*, 1071–1086. doi:10.1098/rstb.2007.2160

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 509-522. doi:10.1037/0278-7393.2.5.509

Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley*

Interdisciplinary Reviews: Cognitive Science, 2, 608–620. doi:10.1002/wcs.133

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 336*, 367–373. doi:10.1098/rstb.1992.0070

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., … Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition, 129*, 241–255. doi:10.1016/j.cognition.2013.07.002

Sebastian, K., Ghose, T., Zacks, J. M., & Huff, M. (2017). Understanding the individual cognitive potential of intellectually disabled people in workshops for adapted work. *Applied Cognitive Psychology, 31*, 175–186. doi:10.1002/acp.3315

Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science, 18*, 449–455. doi: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*, 521–532. doi:10.1016/j.neuron.2007.07.003

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*, 236–257. doi:10.1037/a0015631

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York City, NY: Academic Press.

Wester, M. (2012). Talker discrimination across languages. *Speech Communication, 54*, 781–790. doi:10.1016/j.specom.2012.01.006

Wilder, D. A. (1978). Predictability of behaviors, goals, and unit of perception. *Personality and Social Psychology Bulletin, 4*, 604–607. doi: 10.1177/014616727800400422

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science, 28*, 979–1008. doi:10.1207/s15516709cog2806_5

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*, 651–655. doi:10.1038/88486

Zacks, J. M., Kumar, S., Abrams, R. A., & Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition, 112*, 201–216. doi:10.1016/j.cognition.2009.03.007

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General, 138*, 307–327. doi: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*, 273–293. doi:10.1037/0033-2909.133.2.273

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*, 466–482. doi:10.1037/0882-7974.21.3.466

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*, 150–162. doi:10.1016/j.brainres.2005.12.122

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*, 3–21. doi:10.1037/0033-2909.127.1.3

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

Zalla, T., Pradat-Diehl, P., & Sirigu, A. (2003). Perception of action boundaries in patients with frontal lobe damage. *Neuropsychologia, 41*, 1619–1627. doi:10.1016/S0028-3932(03)00098-8

Zalla, T., Verlut, I., Franck, N., Puzenat, D., & Sirigu, A. (2004). Perception of dynamic action in patients with schizophrenia. *Psychiatry Research, 128*, 39–51. doi:10.1016/j.psychres.2003.12.026

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*, 292–297. doi:10.1111/j.1467-9280.1995.tb00513.x