

Transition-Related Potentials as Markers of Narrative Comprehension in Continuous EEG

Bálint Csanády^{a,*}, Péter Vedres^{b,c}, Kristóf Zsolt Makó^a,
Orsolya Papp-Zipernovszky^{a,d}, Márta Volosin^f, Dávid Apagyi^a,
András Lukács^a, András Bálint Kovács^a, Zoltan Nadasdy^{a,e,g,**}

^a*ELTE Eötvös Loránd University, Budapest, Hungary*

^b*Budapest University of Technology and Economics, Budapest, Hungary*

^c*HUN-REN Wigner Research Centre for Physics, Budapest, Hungary*

^d*Semmelweis University, Budapest, Hungary*

^e*University of Miskolc, Miskolc, Hungary*

^f*University of Szeged, Szeged, Hungary*

^g*The University of Texas at Austin, Austin, TX, USA*

Abstract

Harnessing the potential of electroencephalography (EEG) for brain research is fundamentally limited by intrinsic noise and the diffuse projection of brain-generated activity over the scalp. The standard event-related potential (ERP) paradigm addresses this limitation by relying on repeated independent trials, albeit at the cost of moving away from naturalistic experimental conditions. As a more naturalistic alternative, we collected continuous EEG while participants watched short films and extracted potentials aligned to sharp cinematic transitions (cuts). We demonstrate that such transition-related potentials (TRPs) exhibit canonical ERP-like temporal structure associated with significant information processing. By comparing coherent films with scene-scrambled versions containing matched post-cut sensory input, we find that these responses are systematically shaped by narrative context. We then show that the cut-related EEG signature can be recovered directly from group-averaged continuous recordings with a compact deep neural network (DNN). The detector generalized across films and subject groups, and the resulting TRPs reproduced the main context-dependent effects observed for manually annotated cuts. These results indicate that narrative context leaves

*Corresponding author: csbalint@protonmail.ch

**Corresponding author: zoltan@utexas.edu

a measurable signature in EEG responses, that this signature can be detected directly in continuous recordings, and that such detections provide a semi-automated framework for analyzing how viewers process and understand film narratives. We propose that the method outlined here can be adapted to parse EEG responses to other forms of continuous stimulation, providing a general tool for probing experimental conditions that are closer to natural human experience.

Keywords: EEG, ERP, deep learning, LSTM, cut detection, short films

1. Introduction

Our conscious experience is dominated by the parsing of continuous sensory streams into discrete, meaningful segments, through which we construct the underlying narrative structure of the world around us. This segmentation is often triggered by sensory events that elicit recursive perceptual cycles of feature extraction, scene segmentation, object recognition, and integration with prior knowledge (Stojić and Nadasdy, 2024). Understanding the neuronal processes underlying such events remains a challenge in cognitive neuroscience, especially when relying on non-invasive and readily available tools such as electroencephalography (EEG). What triggers a cognitive event in one individual may not trigger the same event in another, and even when an event is shared, the timing of the corresponding cognitive EEG markers may differ across people. This variability is compounded by the fact that scalp-recorded activity has a low signal-to-noise ratio (SNR) and reflects the simultaneous activation, interference, and superposition of multiple neural sources. Moreover, the mapping from neural sources to scalp potentials is not uniquely invertible, so recovering the underlying neural events from continuous EEG constitutes an ill-posed inverse problem (Nunez and Srinivasan, 2006).

A standard way to separate stimulus-related activity from noise is to improve the SNR by averaging EEG segments aligned to repeated triggers. In the event-related potential (ERP) paradigm, this trigger is usually a stimulus onset. EEG segments aligned to repeated presentations of the same stimulus or stimulus category allow related activity to combine constructively, while activity unrelated to the stimulus is attenuated (Luck, 2014; Kappenman and Luck, 2016; Boudewyn et al., 2018). This yields reproducible ERP waveforms and has become a central method for tracking cognitive processes with EEG.

However, the method typically relies on a trial-by-trial structure in which stimuli are repeatedly presented in isolation, usually in a randomized order, so that trials can be treated as approximately independent. This requirement is in tension with cognition in its natural form, where each event unfolds in relation to those preceding it. In this way, the continuous experience that motivates the question is partly sacrificed for methodological control.

In contrast, allowing subjects to watch short films as a continuous, task-free visual experience provides a more naturalistic window into brain activity (Sonkusare et al., 2019). Although professionally produced films differ markedly from the trial-by-trial stimulus presentations typical of ERP studies, they are far from unstructured stimuli. Films contain precisely planned transitions (cuts), and salient yet unpredictable events that closely emulate real-life scenarios.

We recorded EEG from participants while they watched a film, treated the film as a continuous naturalistic stimulus, and used the cuts separating shots as event triggers. The method preserves the continuity and contextual dependence of naturalistic experience, but introduces transitions that are discrete, frame-accurate, and shared across viewers. While individual differences in interpretation may introduce substantial inter-individual EEG variance (Mars et al., 2008), cuts serve as precisely timed markers around which average waveforms can be computed, approximating the kinds of event boundaries that are thought to structure ongoing experience (Zacks et al., 2007, 2010; Kurby and Zacks, 2008; Magliano and Zacks, 2011). Related event-boundary and cinematic-transition studies have shown that such transitions can elicit characteristic ERP-like responses (Silva et al., 2019; Sanz-Aznar et al., 2023). As such, cuts provide temporal anchors for EEG analysis through transition-related potentials (TRPs), marking moments at which attention may shift and new information may be processed.

If cuts can serve as temporal anchors for naturalistic EEG, the next question is what kind of neural response they elicit. The EEG responses to cuts themselves have been characterized in some detail. Cut-locked responses contain both early components and later components (Matran-Fernandez and Poli, 2015); they are sensitive to whether a cut joins related or unrelated material (Francuz and Zabielska-Mendyk, 2013), and they vary with the editing technique used (Heimann et al., 2017). What has not yet been examined is whether and how these responses depend on narrative coherence, that is, on the extent to which transitions are supported by the preceding narrative context. Furthermore, to date, these responses have primarily been measured

by averaging around cuts whose timing was already known. It remained an open question whether cut-related responses can be detected within the continuous EEG recordings without prior knowledge of the cut timings.

In this study, we investigate both questions. First, we characterize the effect of narrative coherence on cut-locked TRP signals. Second, we propose a population-level deep-learning model for detecting cut-related EEG responses in continuous recordings. In addition to the model’s ability to identify cuts with great precision, we demonstrate that the detected timestamps are suitable substitutes for the original cut annotations, in the sense that the resulting TRPs still exhibit the same characteristics in relation to narrative coherence as their hand-annotated counterparts. Finally, we compare our cut-detection approach with alternative methods to better understand the factors driving model performance. Together, these analyses provide a semi-automated framework for using EEG to study how viewers segment, update, and comprehend film narratives during naturalistic viewing.

2. Experimental Setup

The experiment was conducted at the Institute of Psychology, University of Szeged, between October 2022 and May 2023. Participants sat in a comfortable chair in a dimly lit, sound-attenuated, and electrically shielded room. A 20-inch LCD monitor (LG Flatron L206WTQ-SF; LG Electronics) and two speakers were placed on a table in front of them. Participants were presented with one of two short romantic films with minimal dialogue and a clear narrative structure: *The Art of Love* (Farinella, 2013) or *City Lights* (Wiles, 2016). Hereafter, we refer to them as *Art* and *City*, respectively.

In addition to the original film with a coherent narrative, an incoherent version was created by a professional editor, who cut the film at scene boundaries and randomly reordered the scenes. Because simple randomization could leave two consecutive takes in place, we enforced a scrambling procedure that preserved no consecutive takes (Figure 1). Each participant viewed the coherent version of one film and the incoherent version of the other. Thus, participants watched either the coherent version of *Art* and the incoherent version of *City*, or vice versa. The order of film presentation (film title and narrative coherence) was counterbalanced across participants. After each film, participants completed the validated Hungarian version of the Narrative Engagement Scale (Papp-Zipernovszky et al., 2024). The Narrative Engagement Scale is a 12-item self-report measure of audience engage-

ment with narrative content, originally developed by Busselle and Bilandzic (2009). It contains four subscales (Narrative Understanding, Attentional Focus, Emotional Engagement, and Narrative Presence), and each response is given on a 7-point Likert scale.

Coherent narrative (original)

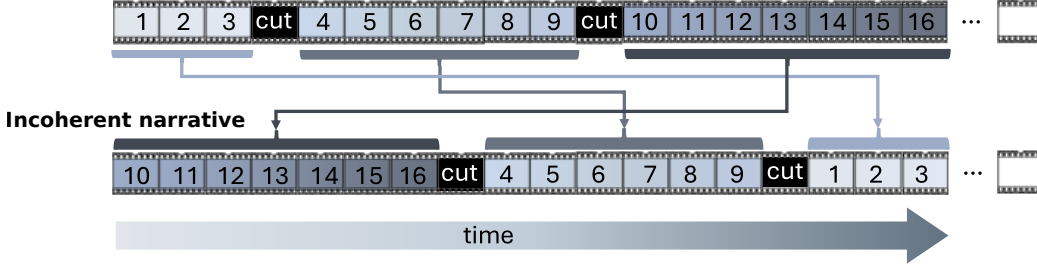


Figure 1: Editing the films with incoherent narratives. The incoherent versions were created by shuffling the takes separated by cuts. Each participant watched one version of each film, one with a coherent narrative and one with an incoherent narrative.

2.1. Participants

The experiment included 34 university students aged between 18 and 47 years (10 men, 24 women; mean age: 22.82 years, SD: 5.48; 5 left-handed). Participants were recruited through university courses and mailing lists, and received no financial compensation. All reported normal or corrected-to-normal vision, normal hearing, and no history of psychiatric or neurological disorders. None were studying cinematography or related fields. All participants provided written informed consent after receiving a full explanation of the experimental procedure. The study was approved by the United Review Committee for Research in Psychology (Hungary, EPKEB – number 2021-123), and was conducted in accordance with the Declaration of Helsinki (World Medical Association, 2025).

2.2. EEG Recording

Continuous EEG was recorded at a sampling rate of 512 Hz using a BioSemi ActiveTwo system (BioSemi B.V., 2001). A total of 34 active electrodes (32 scalp electrodes and 2 additional electrodes) were mounted on an elastic cap according to the 10/20 system (Nuwer et al., 1998). Electrode impedances were kept below 30 k Ω . Stimulus presentation was controlled

using MATLAB (Version 9.0.0, R2016a; The MathWorks Inc.) and Psychophysics Toolbox Version 3.0.18 (Brainard, 1997; Kleiner et al., 2007), running under the Windows 10 operating system.

2.3. EEG Preprocessing

The EEG data were processed in MNE-Python (Gramfort et al., 2013) using the BioSemi 32-channel montage. Cuts were manually annotated at frame accuracy and stored as millisecond timestamps corresponding to the start of the first frame following each cut. Bad channels were automatically identified in MNE on the basis of a local outlier factor criterion and subsequently reconstructed by spatial interpolation. Line noise was removed by a 50 Hz notch filter, and a 0.1–60 Hz band-pass filter was applied to reduce low-frequency drift and high-frequency noise. Extracted EEG segments were normalized per channel by dividing by the standard deviation computed over the full epoch window (−200 to 1200 ms), then baseline-corrected by subtracting the mean amplitude of the −200 to 0 ms pre-stimulus interval.

To visualize the spatial distribution of the cut-TRP response, electrode-wise grand-average waveforms were computed separately for each film by averaging across participants. Figure 2 shows the resulting butterfly plots for the coherent versions of the films, together with the corresponding anterior and posterior averages. A separation between the anterior (negative) and posterior (positive) electrode groups is clearly visible, motivating the use of anterior and posterior electrode-group averages in the TRP analyses below.

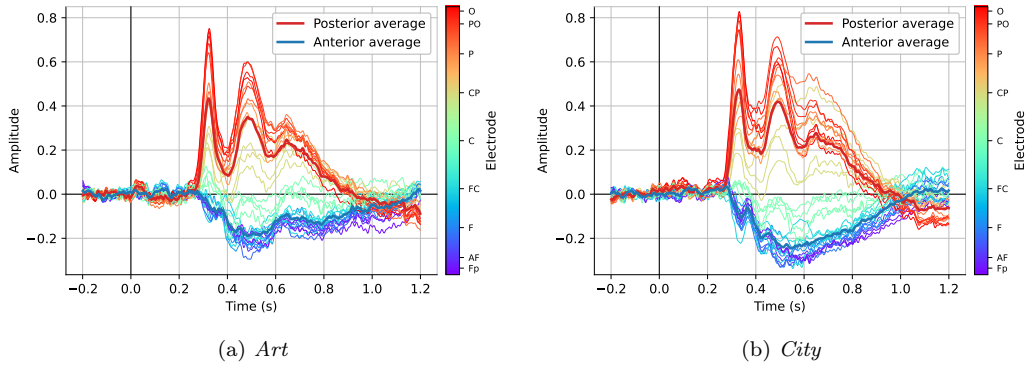


Figure 2: Butterfly plots of the average TRP signals aligned to cuts in the coherent versions of the films, by individual electrodes and anterior/posterior averages. The color of the electrode traces is based on their anteriority.

3. Results

3.1. TRP Analysis

Rather than partitioning the experiment into discrete trials, we annotated the timing of the cuts in the films and treated them as naturally occurring temporal markers in the continuous EEG recordings. Because cinematic cuts can take several forms, ranging from gradual dimming or translucent transitions to instantaneous frame changes without an intervening blank frame, we defined each cut as the earliest detectable transition between two scenes. If cuts function as event boundaries at which significant new information is introduced, the corresponding cut-locked TRPs should contain ERP-like components, interpreted within the standard ERP framework (Luck, 2014; Picton et al., 2000).

As anticipated, the TRP displayed in Figure 2 reproduced a prototypical posterior positivity in the P2 range (~ 250 – 400 ms), followed by prominent late components including the P3/P300 (~ 400 – 600 ms) and LPC/P600 ($\gtrsim 600$ ms) over the posterior domain. This was complemented over the anterior electrodes by a modest N270-like deflection and a sustained late negativity between 400 and 1000 ms. One striking feature of the TRPs was a divergence between anterior and posterior activity. This divergence began with a steep increase in occipital positivity at approximately 300 ms after cut onset, leading to the peak formation of P2, and was accompanied by the simultaneous evolution of anterior negativity reaching the minimum at ~ 500 ms, followed by a convergence complete at ~ 1 s. Conspicuously, the earlier exogenous visual components, such as P50 and N100, were not expressed. This pattern is consistent with previous ERP studies using dynamic video stimuli, in which early visual components were weak or absent under continuous stimulus presentation, whereas later components related to scene comprehension, semantic integration, and post-cut updating remained robustly detectable (Sitnikova et al., 2003, 2008; Sanz-Aznar et al., 2023). These early sensory components are likely attenuated, refractory, or temporally smeared in continuous film viewing, because the visual system is already engaged by ongoing stimulation and because the exact perceptual registration of a cut may vary across edits and observers. Despite the possible arbitrariness of onset detection, the ERP-like components were reproduced with temporal precision characteristic of event-related responses.

To investigate whether and how the cut-locked response was shaped by the coherence of the film’s narrative, we compared TRPs derived from the

coherent and incoherent versions of each film. Since the cut-locked TRPs demonstrate a clear separation of anterior and posterior activity, we compared the anterior and posterior average TRP signals (Figure 3). Clear statistical differences were observed between TRP amplitudes following coherent vs. incoherent cuts, highlighted by green shaded intervals.

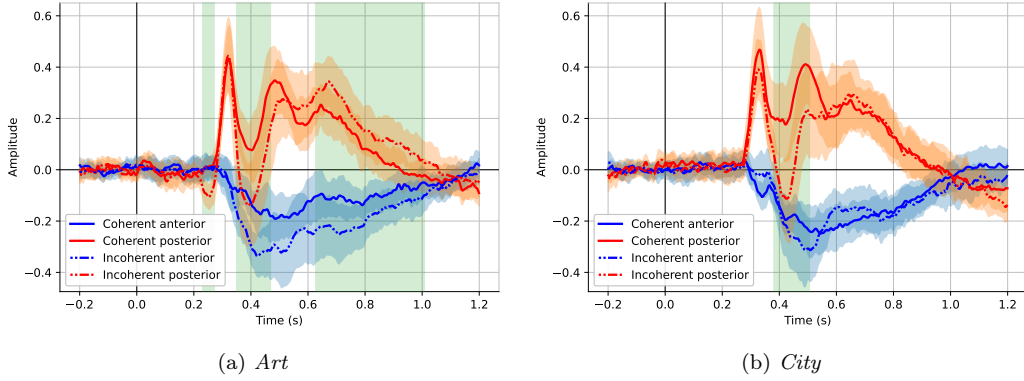


Figure 3: Anterior and posterior TRPs of the cuts within coherent and incoherent narrative versions of the same film. Cuts are matched one-to-one. Green bars indicate significant clusters.

Statistical significance was assessed using a spatio-temporal cluster permutation test across subjects (Maris and Oostenveld, 2007). Candidate clusters were formed across adjacent time samples and neighboring EEG sensors using a one-way F-statistic with a cluster-forming threshold of $p < 0.0027$ (3σ). Clusters were considered significant at a threshold of $p < 0.0456$ (2σ), estimated using 10 000 permutations. This conservative cluster-forming threshold was chosen to favor focal, high-confidence effects over broad clusters driven by weaker distributed differences. In both films, the posterior average showed a stronger late negative deflection when the narrative was incoherent. Additionally, in *Art*, there was a statistically significant negative bump in the incoherent posterior average, and the amplitude attenuated more slowly during cuts related to an incoherent narrative context. The exact details of the clusters are shown in Table 1.

It is worth pointing out that in these comparisons the film segments following the cuts were paired one-to-one with frame precision. This means that the TRPs compared aggregate responses from time windows in which the immediate post-cut audiovisual information presented to participants was identical across the coherent and incoherent versions. As noted above,

Film	$t_{\min}$	$t_{\max}$	p -value	Channels
<i>Art</i>	0.230	0.271	0.0106	CP1, CP2, CP5, CP6, O1, O2, Oz, P3, P4, P7, P8, P03, P04, Pz
<i>Art</i>	0.350	0.469	0.0018	AF3, AF4, C4, CP5, CP6, F3, F7, F8, FC5, FC6, Fp1, Fp2, P4, P7, T7, T8
<i>Art</i>	0.627	1.008	0.0011	CP1, CP2, Cz, F3, F7, FC1, FC5, P3, P4, P03, Pz, T7
<i>City</i>	0.381	0.506	0.0042	CP1, CP2, CP5, CP6, F8, FC6, O1, Oz, P3, P4, P7, P8, P03, P04, Pz, T7, T8

Table 1: Significant clusters for the *Art* and *City* films.

coherence assignment was counterbalanced across films, so the observed effects cannot be attributed to a fixed difference between participant groups. Consequently, the measured effects reflect the influence of preceding context, including narrative comprehension.

To assess whether the differences observed over posterior and anterior electrode groups were related to participants’ engagement with the narrative, we analyzed the post-experiment questionnaire completed by each participant. Table 2 shows the aggregate Narrative Engagement scores from the questionnaire completed after each film. At the individual level, coherent film viewing was strongly correlated with higher engagement scores (*Art*: 0.77; *City*: 0.73). Overall, the coherent versions of the two films received significantly higher engagement scores than their incoherent versions, providing a plausible behavioral correlate of the TRP differences. Moreover, people who watched the incoherent version of *Art* scored somewhat lower on the scale, which is consistent with the stronger TRP difference; however, the confidence intervals overlap, so this effect cannot be taken as conclusive.

Narrative Film	Coherent		Incoherent	
	<i>Art</i>	<i>City</i>	<i>Art</i>	<i>City</i>
Mean score	64.2	66.3	42.6	45.3
CI _{95%}	4.8	3.1	4.5	7.5

Table 2: Narrative Engagement mean scores and confidence intervals by film.

To elucidate the main factors underlying the modulation of TRP amplitudes, we performed a temporal Principal Component Analysis (PCA) on cut-locked TRP waveforms (Picton et al., 2000). For a given film (e.g. *Art*), we extracted the coherent and incoherent TRP channel averages from 200 to 1200 ms. This yielded 64 observations in a 512-dimensional temporal feature space. PCA on these observations produced 512-dimensional temporal eigenvectors and principal component (PC) scores consisting of 32 + 32 weights,

one for each channel in each condition. Figure 4 summarizes the results of this analysis, performed on *Art* and *City* independently. Notably, the resulting eigenvectors were similar across films, and the first PC scores highlight the main anterior–posterior polarity separation we have already remarked on. This evidence supports the anterior–posterior polarity separation as the strongest topographic pattern associated with the processing of cuts in both coherent and incoherent narratives. The second PC, in contrast, captured a posterior topographic pattern related to narrative coherence (Figure 4). We can see that PC2 modulates the extensive temporal-parietal negativity of the occipito-parieto-temporal dipole for the incoherent narratives for both films while it was completely absent during watching the coherent films. A spatial variant of the PCA analysis is available in the Appendix (Figure B.13).

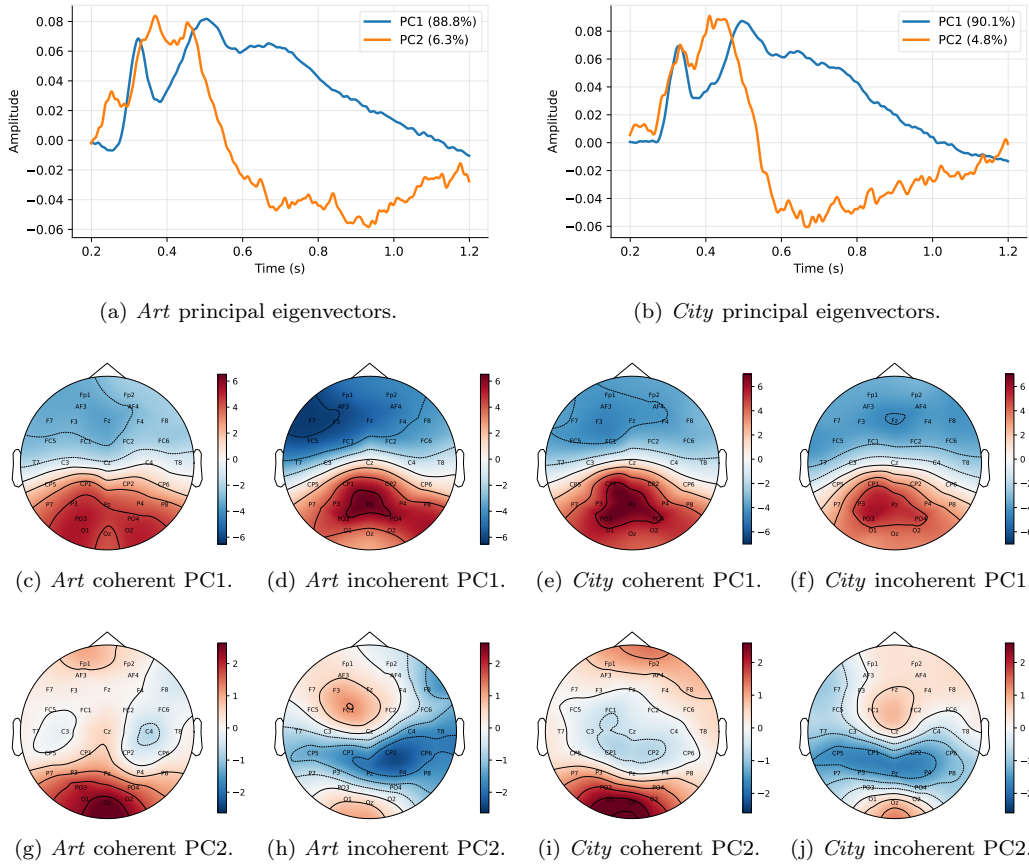


Figure 4: Temporal PCA analysis of the TRPs.

3.2. *Cut Detection*

We have shown that TRPs are sensitive markers of cognitive events that not only reproduce critical features of ERPs, but also reveal important aspects of film narrative processing without participants being explicitly instructed to pay attention to them. This finding motivated us to use deep learning to detect significant cognitive events that occur naturally during the processing of the complex sensory stream typical of film viewing. More specifically, we trained a deep neural network (DNN) to detect manually annotated cinematic cuts from EEG signals alone and then applied the model to recover cut-like EEG patterns, including patterns not necessarily elicited by cinematic cuts.

DNNs are well established for detecting events in EEG, most prominently for seizure detection (Roy et al., 2019) and sleep-stage scoring (Supratak et al., 2017), and compact architectures developed for brain-computer interfaces have shown that such models can generalize across tasks and subjects (Lawhern et al., 2018; Schirrneister et al., 2017). These events, however, are defined by the neural state itself: seizures and sleep stages can be identified from EEG by professionals, which guarantees the presence of a detectable neural signature. In contrast, cinematic cuts are defined objectively by the film and independently of brain activity. In addition, cuts are audiovisual transitions that may elicit sensory and cognitive processes related to the closure of the previous event and the beginning of a new one, consistent with event-segmentation accounts of film comprehension and cinematic continuity (Zacks et al., 2007, 2010; Magliano and Zacks, 2011; Smith, 2012), thereby modulating late ERP-like components. Studies using naturalistic stimuli have applied inter-subject correlation to identify stimulus-driven EEG responses shared across viewers (Hasson et al., 2004). Our analysis also relied on common components of EEG signals, but focused on a more complex objective. We asked not only whether the events could be detected with accuracy exceeding correlation-based methods, but also whether their onsets could be localized with sufficient temporal precision to recover high-quality TRPs.

To fulfill these requirements, we trained a DNN to recover the timing of cinematic cuts directly from the synchronized EEG signals. The population-level detector operated on short sliding windows of the averaged EEG signal, producing a continuous cut-likelihood score that was converted into discrete cut predictions by detecting peaks in the score trace (Figure 5). To clarify the

detector, we first outline the DNN model used to compute the cut-likelihood score.

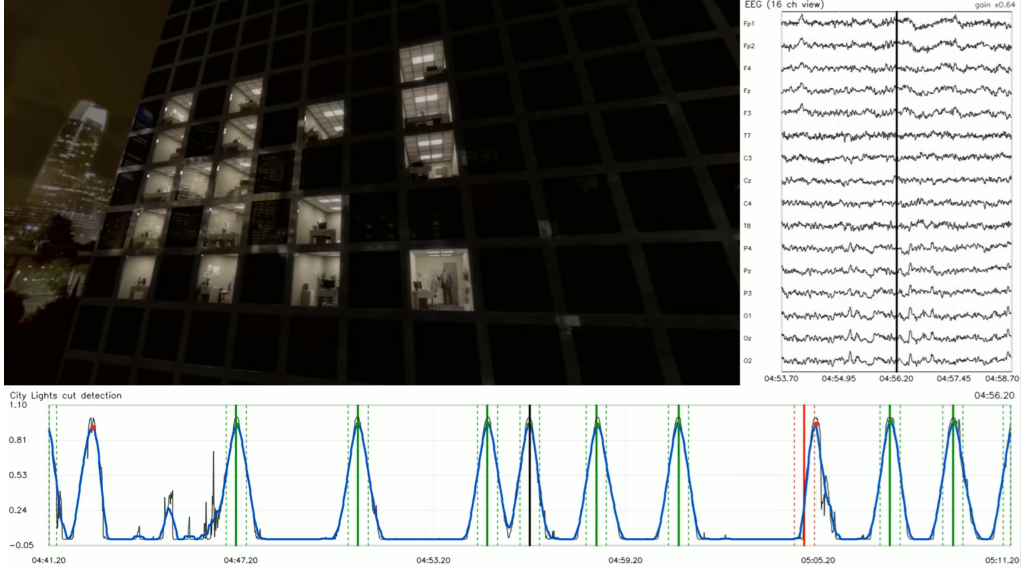


Figure 5: Snapshot from the EEG cut detection on the film *City*.

3.2.1. DNN Model

Let $X \in \mathbb{R}^{S \times 32 \times T}$ denote the cut-synchronized EEG recordings for one version of one film, where S is the number of subjects, 32 is the number of channels, and T is the number of samples. To ensure comparable signal contribution from each participant, we first normalized each response using a causal rolling baseline Z-score at the participant level. Specifically, each EEG sample was standardized channel by channel relative to the mean and standard deviation of the preceding 200 ms of the same signal. The normalized signals were then averaged across subjects to obtain $\hat{X} \in \mathbb{R}^{32 \times T}$. This across-subject averaging improved the signal-to-noise ratio by emphasizing activity that was temporally aligned and shared across subjects. $\hat{X}$ was divided into overlapping windows $W_i \in \mathbb{R}^{32 \times 512}$ of 512 samples (1 s).

Each extracted window was assigned a ground-truth label $y(W_i)$ based on its temporal proximity to annotated events. For each window onset t_i and nearest cut annotation t_i^* (both measured in samples), we defined $y(W_i)$ as:

$$y(W_i) = \min \left(\max \left(1.1 - \frac{|t_i - t_i^*|}{320}, 0 \right), 1 \right).$$

Architecture. We built the cut detector as a compact classifier with recurrent convolutional architecture. A 1D convolutional embedding first compressed the input EEG window to a shorter temporal representation. We employed an embedding consisting of two convolutional stages with output channel sizes 32 and 64, using kernel size and stride of 3 in the first and 4 in the second layer. The embedded sequence was then processed by a two-layer bidirectional residual LSTM with 32 hidden units. The processed sequence was average-pooled resulting in a single 64 dimensional vector. Finally, this vector was processed by a dense classification head with 16 hidden neurons, producing a scalar logit for each input window. The convolutional layers captured short-timescale temporal patterns, while the LSTM examined how these patterns evolved across the window.

Training. The models were trained on randomly sampled mini-batches of size 64, with binary cross-entropy loss on logits with the continuous target value $y(W_i)$, and AdamW optimization (learning rate: 0.0005, weight decay: 0.001). For sampling purposes, windows where $y(W_i) > 0$ were considered positive, and windows with $y(W_i) = 0$ were considered negative. To balance training, negative windows were randomly under-sampled in each epoch to match the number of positive windows. For regularization, we trained the networks using batch normalization and dropout (0.25). To augment training, we added random white noise (SD: 0.1) to each sample before presenting it to the network. Unless otherwise stated, models were trained for 20 epochs.

Inference Methodology. At inference time, the model was evaluated at every eighth window, providing a time series of model predictions throughout the film. This model score trace was smoothed using a 20-sample moving average window. Cuts were predicted by peak finding on the smoothed score trace. A peak was accepted only if it satisfied the chosen minimum height (0.5), minimum width (3), and minimum inter-peak distance (0.625 s).

A detected cut was classified as a true positive (TP) if it fell within an acceptance window of 0.625 s centered on a true cut. In other words a detected peak corresponding to EEG window W_i was accepted if $y(W_i) \geq 0.575$. Similarly, detected cuts were classified as false positives (FPs) if they fell outside all acceptance windows, and acceptance windows with no detected peak were counted as false negatives (FNs). Double-counting of true positives was prevented by setting the minimum inter-peak distance equal to the acceptance-window width.

3.2.2. Results

In the following analysis, we compared different ways of partitioning the four available datasets (2 films, 2 narrative coherence modes) into training and evaluation sets, as summarized in Table 3. Each row defines one generalization setup. The DNN was trained on Group A and evaluated on Group B, and the same comparison was also performed in the reverse direction. These splits tested whether the detector generalized across film identity, participant group, narrative coherence, or the joint change of film identity and participant group.

Generalization	Group A	Group B
Film	<i>Art</i> , <u><i>Art</i></u>	<i>City</i> , <u><i>City</i></u>
Subjects	<i>Art</i> , <u><i>City</i></u>	<u><i>Art</i></u> , <i>City</i>
Coherence	<i>Art</i> , <i>City</i>	<u><i>Art</i></u> , <u><i>City</i></u>
Film & subjects	<i>Art</i> <u><i>Art</i></u>	<i>City</i> <u><i>City</i></u>

Table 3: Dataset splits used for DNN generalization tests. Models were trained on one group and evaluated on the other, with both directions tested. The wavy underline marks the incoherent version of a film (e.g. *Art*).

We performed five independent training sessions for each setup. For each training session, we summarized performance by averaging the F1 scores for detected peaks across epochs 10 to 20. Figure 6 shows the box plot of the results. As expected, the most challenging training setup was the one in which the model had to generalize across both film and subject group. This was likely due to both the amount of training data being halved and the harder generalization task. However, this setup prevented the detector from exploiting film- or subject-specific idiosyncrasies and tested whether the learned EEG signature of cuts transferred across stimulus material. The remaining experimental results in this section are based on this training setup. The difference in performance between the two films was also notable. Although all training setups achieved higher F1 scores when tested on the incoherent version of *Art*, this pattern did not hold for *City*.

Error Analysis. The primary goal of EEG-based cut detection was to test whether TRPs triggered by visual and cognitive events are robust enough relative to noise to support cut identification. At the same time, detection errors, especially false positives, can reveal the types of patterns and narrative events that share EEG characteristics with cinematic cuts. Therefore,

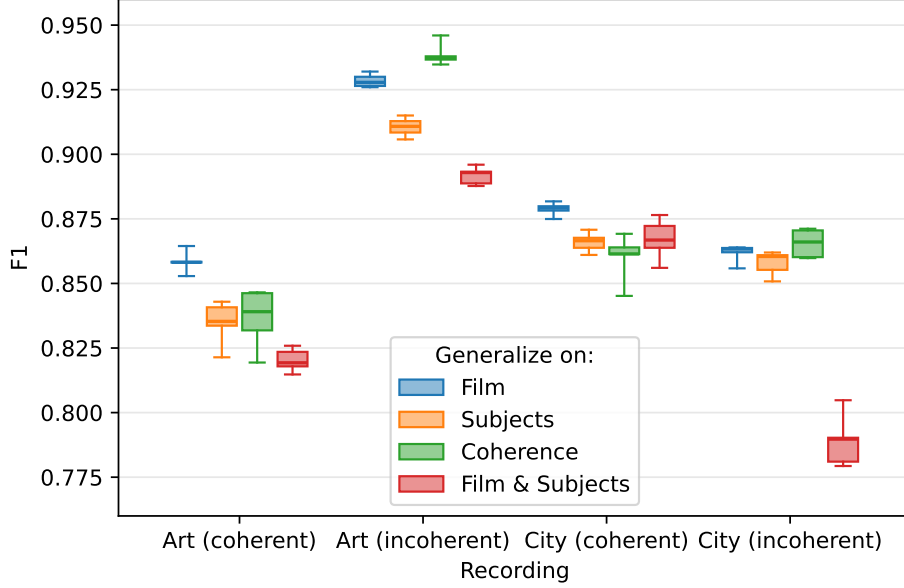


Figure 6: Box plot of the neural model’s performance on the different EEG recordings, by the generalization objective.

these errors may help to clarify what kinds of stimuli TRPs capture from the EEG. Upon initial inspection, false negatives did not appear to form a random subset: at least some corresponded to cuts in which the audiovisual transitions were too subtle, or in which transients occurred in rapid succession, making the signal difficult to resolve. For example, cuts corresponding to camera zoom or angle shifts seemed hard to notice. Likewise, false positives did not seem to appear at random either: many were auditory and/or visual transitions in the film, such as a light switching on or off, a picture being shown on a computer monitor, or a change of framing where a narratively important object or person appears. These may share common sensory or narrative features with true cuts.

In order to analyze the errors by the above aspects, we categorized them manually (Table 4). This error analysis used all annotated cinematic cuts, rather than only the matched subset used for the TRP comparisons. False positives were classified as *transients* when an intense auditory or visual transient was present other than a cinematic cut, as *significant events* when the detection also coincided with a narratively relevant event, or as *undetermined* otherwise. False negatives were classified as *weak transients* when the audio-

Film Narrative	<i>Art</i>		<i>City</i>		%
	Coherent	Incoherent	Coherent	Incoherent	
True positive	123	158	147	118	
False positive	4	28	17	5	
Transient	0	12	8	5	46.3%
Significant event	4	13	5	0	40.7%
Undetermined	0	3	4	0	13.0%
False negative	44	7	28	59	
Weak transient	7	0	8	10	18.1%
Sub-window	12	5	16	28	44.2%
Undetermined	25	2	4	21	37.7%

Table 4: Cut detection outcomes by EEG recording, manual error assessment.

visual transient associated with the cut was subtle (e.g. cut involved only a change of camera angle), as *sub-window* when a neighboring cut was present within a 1 s window, as *undetermined* when neither of these applied. Overall, the error structure suggests that the detector learned a physiologically meaningful cut-related response rather than an arbitrary mapping from EEG fluctuations to annotations. The majority of false negatives (62.3%) occurred when the cinematic transition was perceptually weak or temporally crowded by neighboring cuts, whereas false positives mostly coincided (87%) with strong audiovisual changes or narratively meaningful events. This pattern supports the interpretation that the detector was sensitive to EEG activity resembling event-boundary processing, even when such processing was not always tied to a formally annotated cinematic cut.

TRP Validation. To validate the efficacy of the model outputs, we repeated the TRP analysis from Section 3.1 using the detected cuts (both true and false positives). The accuracy of the recovered TRPs is remarkable, considering that the model was blind to the cut-segmentation of the films. Figure 7 shows the resulting TRPs, and Table 5 lists the cluster permutation test results. Most importantly, similar clusters were recovered as in the hand-annotated TRPs (Figure 3 and Table 1) with marginally smaller confidence (larger p -values). These results not only corroborate the model’s accuracy in finding cuts, but also highlight its temporal precision, which was essential for obtaining TRP waveforms with clear ERP-like features. Additionally, temporal PCA based on the detected cut TRPs yielded similar results to PCA based on the original cuts (Figure B.14 in the Appendix).

Film	$t_{\min}$	$t_{\max}$	p -value	Channels
<i>Art</i>	0.240	0.293	0.0128	CP1, O1, O2, Oz, P3, P4, P7, P8, P03, P04, Pz
<i>Art</i>	0.369	0.479	0.0045	AF3, AF4, C3, CP1, CP5, F3, F7, FC5, Fp1, Fp2, P3, P7, T7
<i>Art</i>	0.703	1.070	0.0025	CP1, CP2, Cz, F3, F7, FC1, FC5, P3, P4, P04, Pz
<i>City</i>	0.367	0.502	0.0053	CP1, CP5, CP6, F8, FC6, O1, O2, Oz, P3, P4, P7, P8, P03, P04, T7, T8

Table 5: Significant clusters for the detected cuts in *Art* and *City*.

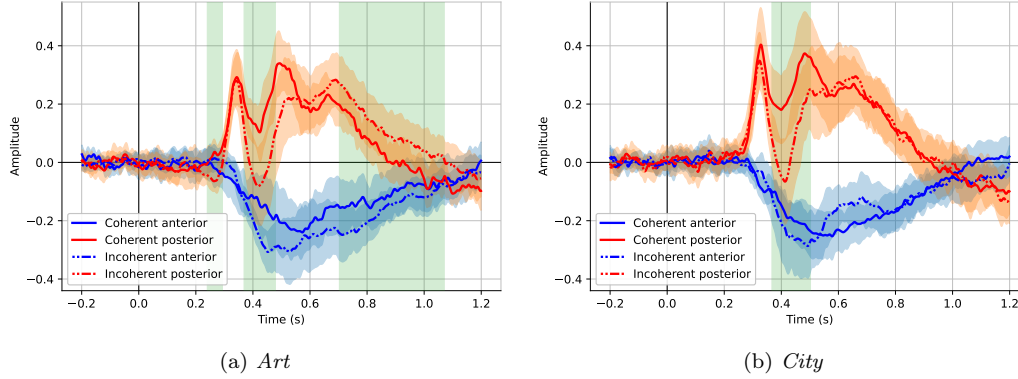


Figure 7: Anterior and posterior average TRPs of the detected cuts within the coherent and incoherent narrative versions of the same film. Green bars indicate significant clusters.

Classifying Narrative Coherence. Finally, we asked whether narrative coherence could be decoded from single-subject EEG responses. To address this question, we computed subject-level TRPs separately for each participant and each film condition by averaging the cut-locked EEG segments of that participant. Each subject-level TRP was then cropped to the 200–1000 ms post-cut interval, and the resulting 32×409 channel-by-time matrix was flattened into a single feature vector. We used these feature vectors to classify whether the corresponding TRP came from the coherent or incoherent version of the film. Classification was evaluated with leave-one-subject-out cross-validation, so that the subject being tested was never included in the training set. Because each subject contributed only one TRP per film, the classifier could not rely on repeated samples from the same subject within a film-specific dataset. Thus, successful classification would indicate that subject-level TRPs retained information about narrative coherence, rather than merely reflecting subject-specific idiosyncrasies. We repeated the analysis using both the original manually annotated cuts and the cut events detected by the DNN model. In the manually annotated version, cuts were

matched one-to-one across coherent and incoherent conditions. This ensured that the classifiers could not rely on systematic differences in the immediate post-cut sensory input.

Film Cut Event	<i>Art</i>		<i>City</i>	
	Original	Detected	Original	Detected
Gradient Boost	0.71	0.74	0.82	0.74
LDA	0.85	0.74	0.74	0.77
Logistic Regression	0.82	0.85	0.85	0.77
Random Forest	0.88	0.85	0.74	0.71
Ridge Classifier	0.79	0.91	0.77	0.77
SVM	0.82	0.85	0.82	0.77
XGBoost	0.82	0.74	0.77	0.79
Average	0.815	0.811	0.786	0.756

Table 6: Classification accuracy scores of narrative coherence based on true cuts and DNN-detected cut-like events. The numbers represent accuracy scores achieved by different ML models using leave-one-subject-out cross-validation.

Narrative coherence was classified using different machine learning (ML) approaches. As Table 6 indicates, average classifier performance decreased only marginally when detected cuts were used instead of original cut annotations. The classifiers correctly identified whether subject-level TRPs were derived from coherent or incoherent narratives for the majority of held-out subjects. Although the average accuracy of the ML models was $> 75\%$, substantially above the empirical chance level, the main point of this analysis was not to maximize classification performance. Instead, the aim was to demonstrate the efficacy of this method to extract subject-level cognitive information from EEG, and to further corroborate the quality of the DNN-based cut detections.

Ablations. A baseline comparison is reported in Appendix A. In brief, the main DNN detector outperformed the anterior-posterior divergence, TRP-template-fit, and correlation-based baselines in the film-generalization setting. Importantly, the DNN localized cut onsets more accurately: on average, true positives were within one frame of the true cut onsets, while simpler peak-detection baselines approached the timing error expected from random detections within the acceptance window. Corroborating that temporal precision was critical for recovering high-quality TRPs from detected events.

4. Discussion

EEG is the most accessible noninvasive tool for studying the brain’s macroscopic electric potentials, which correlate with cognitive functions at millisecond precision. To unlock EEG’s full potential in diverse experimental settings (Wang et al., 2018; Sonkusare et al., 2019; Lau-Zhu et al., 2019; Kaushik et al., 2022) and access deeper layers of information processing in the brain we need to improve the sensitivity and specificity of the data processing method. Even when extrinsic noise is managed, the biggest challenge is the large-amplitude intrinsic noise frequently obscuring the relationship between neuronal responses, particularly when stimuli are continuous and events occur at irregular intervals. Conventional ERP analysis addresses this by averaging time-locked responses across repeated stimulus presentations (Luck, 2014), but this repetition renders most experiments unnatural and contrived. One conventional solution developed to overcome this limitation was single-trial analysis, replacing waveform averaging with spectral or spatial filters and classifiers (Blankertz et al., 2011). Instead, we developed an alternative approach to extract ERP-like transition-related potentials (TRPs) from continuous EEG recorded while subjects watched short films, time-locked to well-defined sensory events, such as cinematic cuts.

The first part of the study demonstrated that TRPs reproduce essential cognitive components of ERPs, specifically P2, P3 and N4 (Figure 2). Among those, we showed that TRPs are systematically modulated by narrative context. Group-level TRPs drawn from EEGs recorded while participants watched scene-scrambled versions of films, exhibited reproducibly different characteristics relative to the TRPs computed from the EEGs while watching the original version. Specifically, we observed significant modulation of P2 and P3 in the posterior and N400 amplitudes in the anterior electrodes consistent with the coherent-incoherent versions (Figure 3). These effects were controlled by the counterbalanced allocation of participants across films and by one-to-one scene matching, which ensured identical audiovisual input in the post-cut TRP source-windows. Furthermore, the TRP differences aligned with participants’ subjective assessments, as scene-scrambled films were consistently rated as less narratively engaging. The observed differences were further corroborated by temporal PCA, which separated two main components underlying the modulation of TRPs (Figure 4). The first and strongest principal component, explaining $\approx 90\%$ of signal variance, captured a broadly distributed posterior positivity: identified with the P2

and P3 components. It also confirmed that the large topographic separation between anterior negativity and posterior positivity was the dominant characteristic of cut-elicited TRPs in general, and was not specific to the narrative differences between the films. In contrast, the second principal component displayed a pronounced difference between coherent and incoherent narratives, with maximal expression over parietal and temporal areas. Together, these findings support two conclusions: first, that cuts set a cascade of cognitive components into motion, originating over occipital cortex and followed by the evolution of a sustained frontal negativity; and second, that the modulation of these components by visual transients carries the fingerprints of highly abstract cognitive processing, specifically narrative coherence, a key determinant of film comprehension and, more broadly, of how meaning is constructed during continuous real-world experience.

It is worth to contextualize these findings within the ERP literature. Regarding the positive components of the TRPs, both P2 and P3 are associated with the reduction of subjective uncertainty. P2 (Luck and Hillyard, 1994) is commonly evoked in auditory oddball and visual priming paradigms, and is most pronounced centrally in auditory speech perception tasks (Signoret et al., 2013), with a more attenuated expression in visual paradigms (Capilla et al., 2016). Its manifestation in our experiment was nonetheless evident: given that P2 is generally associated with novelty (Barcelo and Knight, 2007), it was expected to be most prominent following a cut when the new scene is less predictable. Likewise, the P3 component is sensitive to the information-theoretic surprise value of a stimulus, with P3a over frontal and central areas reflecting attentional engagement (Polich, 2003) and P3b, with its posterior maximum, indexing stimulus unexpectedness (Duncan-Johnson and Donchin, 1977).

Two competing hypotheses can be articulated regarding the P3b in our experiment. The first predicts a more prominent P3b in incoherent-narrative EEGs, since the randomized sequence of shots produces a higher rate of unexpected transitions. The second, however, draws on the relationship between P3b amplitude and information-theoretic unexpectedness, as captured by Rényi entropy (Rényi, 1961), and leads to the opposite prediction. In a coherent film, predictable transitions predominate, meaning that the rare unpredictable cut carries greater surprise and may therefore elicit a larger P3b than any individual transition in an incoherent film, where all cuts are equally unexpected and no strong prior exists to be violated. This dependence on baseline stimulus probability is precisely what Rényi entropy for-

malizes, and trial-by-trial P3 modulation by such prior probabilities has been demonstrated in Bayesian paradigms (Mars et al., 2008).

In addition to the positive components, we observed a prominent anterior negativity spanning from 250 to 1000 ms, which we associate with the cognitive process of anticipation resolution. This frontally-dominant negative potential is consistent with feedback-related negativity documented in the literature (Maruyama et al., 2026). Although N400 is most commonly associated with linguistic tasks, it is not strictly a semantic component; it arises in any task involving the processing of meaning or conceptual information (Ganis et al., 1996). This has been demonstrated in image-based sentence completion paradigms, where the final word is replaced by a predictable or unpredictable image rendering the sentence semantically congruous or anomalous, respectively (Nigam et al., 1992; Ganis et al., 1996). To our knowledge, the only prior study to elicit an N400-like response using video stimuli is that of Sitnikova et al. (2003), who presented contextually congruous and incongruous film endings. They found that incongruous endings evoked an enhanced anterior negativity resembling N400, as well as a more positive late positivity relative to congruous endings of short video presentations of human interactions with objects. This pattern is perfectly concordant with our own finding that cuts demarcating incoherent narratives elicit larger N400 deflections and increased late P3b positivity. Notably, the above referenced study called for exactly the paradigm we employed (p. 163): *“To get a better estimate of the timing of semantic integration in videos, it might be useful to collect ERPs to critical items that have a clear point of appearance in the scene (e.g., at a scene change).”* Our experiment directly answers that call.

Next we turned the task around and approached it from the opposite direction. Leveraging the power of sequence-processing artificial neural networks, we trained DNNs to exploit the characteristic EEG signature of TRPs and localize cinematic cuts in continuous EEG based solely on their electrographic signatures (Figure 5). More specifically, we developed a DNN-based cut-detection methodology that located the vast majority of cinematic cuts from group-averaged EEG signals with a low error rate, and most importantly, with a sub-frame onset accuracy (< 40 ms). With the exception of a single outlier, cut detection achieved consistently strong performance ($0.8 < F1 < 0.95$; Figure 6); even in the most challenging data partitioning scenario, when generalizing across both, film and participant group, using only one group-averaged EEG recording for training. Nevertheless,

performance differences across all generalization conditions remained within $\Delta F1 = 0.05$, underscoring the robustness of the learned representations. Although this approach has practical value as a method for simply detecting cinematic cuts only to a limited extent, failures of the detector may point to events with cognitive properties similar to those of cuts.

Thus, investigating the detection errors proved highly informative. The majority of false negatives (62.3%) were accounted for by cuts involving insufficiently salient visual transients (such as transparent transition), or by cuts occurring in rapid succession within the 1-second window during which multiple detections were suppressed. False positives, by contrast, were of considerable interest, as they shed light on the cognitive underpinnings of cut detection. Conceptually, a false positive arises when an EEG pattern shares the electrographic fingerprint of a TRP-template without being preceded by an actual cut. Analyzing the false positives revealed a general pattern. Some false positives coincided with audiovisual transients that were very similar to cuts while technically not being cinematic cuts, such as a light in the room being switched on/off, or a camera shutter effect. Naturally, the transients associated with false positives (87%), were on a spectrum in terms of intensity. A general trend we found was that some of the weaker transients were also associated with narratively important moments, and this association was a likely factor in their detection. For example, in one scene, on his computer monitor, the protagonist reviews photos he took earlier, and the transitions between these photos resemble cinematic cuts, but impact only a portion of the frame. Out of the four pictures, the one that coincided with a false positive detection was the most narratively important, a portrait of the girl the protagonist is interested in. While such events may lack a strong audiovisual transient and thus the early sensory components might have been too weak, their late cognitive components – P2, P3, and N4 – were nonetheless strong enough to be triggered by narrative events similar to cuts with respect to demanding an update to the viewer’s internal model of the story. The coincidence of false positives with such narrative events suggests that the DNN had learned to generalize TRP features beyond literal cuts, effectively detecting cognitively salient moments in the continuous EEG.

The timing precision of the detector was further demonstrated by the fact that detected events could be used directly to generate TRP waveforms that retained the main narrative-context-dependent differences observed for the original cut annotations (Figure 7). In addition, a participant-level classification of narrative coherence (inferring from individual TRPs if a participant

watched the coherent or the incoherent version of the film), suffered only a marginal decrease in accuracy when TRPs were generated from the algorithmically detected cuts rather than from the original annotations. These results appear even more compelling in light of the errors produced by the detector: many false negatives occurred in challenging cases, such as fast-paced cuts or weak audiovisual changes, whereas many false positives coincided with strong audiovisual transients, often with narrative significance. Taken together, these results provide strong evidence that the DNN learned to identify narratively significant cognitive events from EEG waveform signatures derived from cinematic cuts.

5. Conclusion

Our findings collectively demonstrate that narrative comprehension leaves a measurable electrographic signature at moments of event transition, manifest as transition-related potentials (TRPs). Moreover, by leveraging the structure of TRPs, a DNN can be trained to extract analogous signatures from continuous EEG group-averages by capturing events that share physical or cognitive attributes with cinematic cuts or other well-defined stimulus boundaries. Furthermore, with the appropriate algorithmic approach, these detections are sufficiently accurate to automatically reproduce high-quality TRP waveforms. Importantly, these results indicate, that a robust mapping exists between continuous EEG dynamics and the timing of externally defined narrative transitions. Such electrographic signatures offer a principled means of parsing continuous EEG for cognitive events representing episodes of information updating, uncertainty reduction, or the resolution of suspense: hallmarks of the dramatic architecture of produced films and, more broadly, of natural experience (Dini et al., 2023). This discovery opens the door to detecting meaningful transitions in other continuous stimuli, where event boundaries may not be directly observable or cannot be annotated by hand with comparable precision. Such applications would extend the TRP framework beyond cinematic cuts and enable the analysis of information processing across a broader range of continuous, naturalistic experiences. Future efforts will be devoted to test the generalization of TRPs and to parse real-time individual EEG.

6. Limitations

Several limitations should be considered when interpreting these results. First, the experiment used only two short films with clear narrative structure and relatively sparse dialogue. The generality of the findings should therefore be tested on a wider range of films, genres, editing styles, and audiovisual materials. Second, the DNN detector operated on subject-averaged EEG. It should therefore be interpreted as a population-level detector rather than a real-time single-subject detector. Future work should test whether subject-adaptive models can recover comparable event timing from individual EEG recordings. Third, the interpretation of false positives and false negatives relied partly on manual categorization. Combining EEG-based detection with richer stimulus annotations, including auditory changes, object appearances, semantic shifts, and narrative turning points were out of scope of this study, and would be subject of further investigation. Fourth, due to the small number of experimental subjects, the TRP results are based on signals averaged across both individual cuts and participants. Future studies aiming to analyze responses to individual cuts should therefore use substantially larger participant samples. Finally, the analyses reported here are scalp-level analyses. They identify reproducible temporal and topographic patterns in the EEG, but they do not localize the neural generators of these effects. Source-level analyses or multimodal recordings would be needed to determine the neural systems underlying the observed TRP components.

Ethics Statement

All participants provided written informed consent after receiving a full explanation of the experimental procedure. The study was approved by the United Review Committee for Research in Psychology (Hungary, EPKEB 2021-123). The study was conducted in accordance with the Declaration of Helsinki. Participants received no financial compensation.

Data and Code Availability Statement

The EEG dataset is available at Mendeley Data (Csanády et al., 2026). Code is available at GitHub (Csanády, 2026).

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Credit Authorship Contribution Statement

Bálint Csanády: *conceptualization, methodology, software, formal analysis, investigation, writing – original draft, data visualization*. Péter Vedres: *conceptualization, methodology, formal analysis, investigation*. Kristóf Zsolt Makó: *software, investigation*. Orsolya Papp-Zipernovszky: *research design, psychometric assessment*. Márta Volosin: *research design, investigation, EEG data recording*. Dávid Apagyí: *software, investigation*. András Lukács: *supervision, methodology*. András Bálint Kovács: *conceptualization, research design, methodology*. Zoltán Nádasdy: *conceptualization, methodology, supervision, writing – review and editing*.

Acknowledgments

This research was supported by the EKÖP-25-4-I-ELTE-1093 University Research Fellowship Program of the Ministry for Culture and Innovation from the Source of the National Research, Development and Innovation Fund. The research was additionally supported by the Hungarian National Research, Development and Innovation Office within the framework of the Thematic Excellence Program 2021 – National Research Sub programme: “Artificial intelligence, large networks, data security: mathematical foundation and applications” and the Artificial Intelligence National Laboratory Program (MILAB). This research was also funded by the OTKA research funding program under project number NKFI-138108.

References

- Barcelo, F., Knight, R.T., 2007. An Information-Theoretical Approach to Contextual Processing in the Human Brain: Evidence from Prefrontal Lesions. *Cerebral Cortex* 17, i51–i60. doi:10.1093/cercor/bhm111.
- BioSemi B.V., 2001. ActiveTwo/Active2.5: 280-Channel, DC Amplifier, 24-Bit Resolution Biopotential Measurement System with Active Electrodes. https://www.biosemi.com/Products_ActiveTwo.htm. Accessed: 2026-06-02.
- Blankertz, B., Lemm, S., Treder, M., Haufe, S., Müller, K.R., 2011. Single-trial analysis and classification of ERP components – a tutorial. *NeuroImage* 56, 814–825. doi:10.1016/j.neuroimage.2010.06.048. multivariate Decoding and Brain Reading.

- Boudewyn, M.A., Luck, S.J., Farrens, J.L., Kappenman, E.S., 2018. How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology* 55, e13049. doi:10.1111/psyp.13049.
- Brainard, D.H., 1997. The Psychophysics Toolbox. *Spatial Vision* 10, 433–436. doi:10.1163/156856897X00357.
- Busselle, R., Bilandzic, H., 2009. Measuring Narrative Engagement. *Media Psychology* 12, 321–347. doi:10.1080/15213260903287259.
- Capilla, A., Melcón, M., Kessel, D., Calderón, R., Pazo-Álvarez, P., Carretié, L., 2016. Retinotopic mapping of visual event-related potentials. *Biological Psychology* 118, 114–125. doi:10.1016/j.biopsycho.2016.05.009.
- Csanády, B., 2026. Analysis of Transition-Related Potentials in EEG During Film Viewing. URL: https://github.com/aielte-research/EEG_cuts.
- Csanády, B., Volosin, M., Zipernovszky, O.P., Kovács, A.B., Vedres, P., 2026. Continuous EEG Recordings During Coherent and Scene-Scrambled Film Viewing. doi:10.17632/3kd62t4dt3.1.
- Dini, H., Simonetti, A., Bruni, L.E., 2023. Exploring the Neural Processes behind Narrative Engagement: An EEG Study. *eNeuro* 10, ENEURO.0484–22.2023. doi:10.1523/ENEURO.0484–22.2023.
- Dmochowski, J.P., Sajda, P., Dias, J., Parra, L.C., 2012. Correlated Components of Ongoing EEG Point to Emotionally Laden Attention – A Possible Marker of Engagement? *Frontiers in Human Neuroscience* 6, 112. doi:10.3389/fnhum.2012.00112.
- Duncan-Johnson, C.C., Donchin, E., 1977. On Quantifying Surprise: The Variation of Event-Related Potentials With Subjective Probability. *Psychophysiology* 14, 456–467. doi:10.1111/j.1469-8986.1977.tb01312.x.
- Farinella, L., 2013. The Art of Love. Short film. URL: <https://www.imdb.com/title/tt26250136/>.
- Francuz, P., Zabielska-Mendyk, E., 2013. Does the Brain Differentiate Between Related and Unrelated Cuts When Processing Audiovisual Messages? An ERP Study. *Media Psychology* 16, 461–475. doi:10.1080/15213269.2013.831394.

- Ganis, G., Kutas, M., Sereno, M.I., 1996. The Search for “Common Sense”: An Electrophysiological Study of the Comprehension of Words and Pictures in Reading. *Journal of Cognitive Neuroscience* 8, 89–106. doi:10.1162/jocn.1996.8.2.89.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., Hämäläinen, M.S., 2013. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience* 7, 1–13. doi:10.3389/fnins.2013.00267.
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject Synchronization of Cortical Activity During Natural Vision. *Science* 303, 1634–1640. doi:10.1126/science.1089506.
- Heimann, K.S., Uithol, S., Calbi, M., Umiltà, M.A., Guerra, M., Gallese, V., 2017. “Cuts in Action”: A High-Density EEG Study Investigating the Neural Correlates of Different Editing Techniques in Film. *Cognitive Science* 41, 1555–1588. doi:10.1111/cogs.12439.
- Kappenman, E.S., Luck, S.J., 2016. Best Practices for Event-Related Potential Research in Clinical Populations. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 1, 110–115. doi:10.1016/j.bpsc.2015.11.007.
- Kaushik, P., Moye, A., van Vugt, M., Roy, P.P., 2022. Decoding the cognitive states of attention and distraction in a real-life setting using EEG. *Scientific Reports* 12, 20649. doi:10.1038/s41598-022-24417-w.
- Ki, J.J., Kelly, S.P., Parra, L.C., 2016. Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. *Journal of Neuroscience* 36, 3092–3101. doi:10.1523/JNEUROSCI.2942-15.2016.
- Kleiner, M., Brainard, D., Pelli, D., 2007. What’s new in Psychtoolbox-3? *Perception* 36, 14. doi:10.1177/03010066070360S101.
- Kurby, C.A., Zacks, J.M., 2008. Segmentation in the perception and memory of events. *Trends in Cognitive Sciences* 12, 72–79. doi:10.1016/j.tics.2007.11.004.

- Lau-Zhu, A., Lau, M.P.H., McLoughlin, G., 2019. Mobile EEG in research on neurodevelopmental disorders: Opportunities and challenges. *Developmental Cognitive Neuroscience* 36, 100635. doi:10.1016/j.dcn.2019.100635.
- Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J., 2018. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *Journal of Neural Engineering* 15, 056013. doi:10.1088/1741-2552/aace8c.
- Luck, S.J., 2014. *An Introduction to the Event-Related Potential Technique*. 2 ed., MIT Press, Cambridge, MA.
- Luck, S.J., Hillyard, S.A., 1994. Electrophysiological correlates of feature analysis during visual search. *Psychophysiology* 31, 291–308. doi:10.1111/j.1469-8986.1994.tb02218.x.
- 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.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* 164, 177–190. doi:10.1016/j.jneumeth.2007.03.024.
- Mars, R.B., Debener, S., Gladwin, T.E., Harrison, L.M., Haggard, P., Rothwell, J.C., Bestmann, S., 2008. Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *Journal of Neuroscience* 28, 12539–12545. doi:10.1523/JNEUROSCI.2925-08.2008.
- Maruyama, Y., Aoyama, K., Fukayama, O., 2026. Modulation of feedback-related negativity by objective and subjective response correctness. *Scientific Reports* 16, 13574. doi:10.1038/s41598-026-43451-6.
- Matran-Fernandez, A., Poli, R., 2015. Event-Related Potentials induced by cuts in feature movies and their exploitation for understanding cut efficacy, in: *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, IEEE. pp. 74–77. doi:10.1109/NER.2015.7146563.

- Nastase, S.A., Gazzola, V., Hasson, U., Keysers, C., 2019. Measuring shared responses across subjects using intersubject correlation. *Social Cognitive and Affective Neuroscience* 14, 667–685. doi:10.1093/scan/nsz037.
- Nigam, A., Hoffman, J.E., Simons, R.F., 1992. N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience* 4, 15–22. doi:10.1162/jocn.1992.4.1.15.
- Nunez, P.L., Srinivasan, R., 2006. *Electric Fields of the Brain: The neurophysics of EEG*. 2 ed., Oxford University Press, New York. doi:10.1093/acprof:oso/9780195050387.001.0001.
- Nuwer, M.R., Comi, G., Emerson, R., Fuglsang-Frederiksen, A., Guérit, J.M., Hinrichs, H., Ikeda, A., Luccas, F.J.C., Rappelsburger, P., 1998. IFCN standards for digital recording of clinical EEG. *Electroencephalography and Clinical Neurophysiology* 106, 259–261. doi:10.1016/S0013-4694(97)00106-5.
- Papp-Zipernovszky, O., Deák-Kovács, T., Bálint, K.É., Kovács, A.B., 2024. A Narratív Bevonódás Skála magyar nyelvű adaptációja. *Alkalmazott Pszichológia* 26, 29–50. doi:10.17627/ALKPSZICH.2024.3.29.
- Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, R., Miller, G.A., Ritter, W., Ruchkin, D., Rugg, M.D., Taylor, M.J., 2000. Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology* 37, 127–152. doi:10.1111/1469-8986.3720127.
- Polich, J., 2003. Theoretical Overview of P3a and P3b, in: *Detection of Change: Event-Related Potential and fMRI Findings*. Springer US, Boston, MA, pp. 83–98. doi:10.1007/978-1-4615-0294-4_5.
- Rényi, A., 1961. On Measures of Entropy and Information, in: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, CA. pp. 547–561.
- Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T.H., Faubert, J., 2019. Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering* 16, 051001. doi:10.1088/1741-2552/ab260c.

- Sanz-Aznar, J., Bruni, L.E., Soto-Faraco, S., 2023. Cinematographic continuity edits across shot scales and camera angles: an ERP analysis. *Frontiers in Neuroscience* 17, 1173704. doi:10.3389/fnins.2023.1173704.
- Schirrneister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., Ball, T., 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping* 38, 5391–5420. doi:10.1002/hbm.23730.
- Signoret, C., Gaudrain, E., Perrin, F., 2013. Similarities in the neural signature for the processing of behaviorally categorized and uncategorized speech sounds. *European Journal of Neuroscience* 37, 777–785. doi:10.1111/ejn.12097.
- Silva, M., Baldassano, C., Fuentemilla, L., 2019. Rapid memory reactivation at movie event boundaries promotes episodic encoding. *Journal of Neuroscience* 39, 8538–8548. doi:10.1523/JNEUROSCI.0360-19.2019.
- Sitnikova, T., Holcomb, P.J., Kiyonaga, K.A., Kuperberg, G.R., 2008. Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience* 20, 2037–2057. doi:10.1162/jocn.2008.20143.
- Sitnikova, T., Kuperberg, G., Holcomb, P.J., 2003. Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology* 40, 160–164. doi:10.1111/1469-8986.00016.
- Smith, T.J., 2012. The Attentional Theory of Cinematic Continuity. *Projections* 6, 1–27. doi:10.3167/proj.2012.060102.
- Sonkusare, S., Breakspear, M., Guo, C., 2019. Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends in Cognitive Sciences* 23, 699–714. doi:10.1016/j.tics.2019.05.004.
- Stojić, S., Nadasdy, Z., 2024. Event as the central construal of psychological time in humans. *Frontiers in Psychology* 15. doi:10.3389/fpsyg.2024.1402903.
- Supratak, A., Dong, H., Wu, C., Guo, Y., 2017. DeepSleepNet: a Model for Automatic Sleep Stage Scoring based on Raw Single-Channel EEG.

- IEEE Transactions on Neural Systems and Rehabilitation Engineering 25, 1998–2008. doi:10.1109/TNSRE.2017.2721116.
- Wang, M.Y., Luan, P., Zhang, J., Xiang, Y.T., Niu, H., Yuan, Z., 2018. Concurrent mapping of brain activation from multiple subjects during social interaction by hyperscanning: a mini-review. *Quantitative Imaging in Medicine and Surgery* 8, 819–837. doi:10.21037/qims.2018.09.07.
- Wiles, E., 2016. *City Lights*. Short film. URL: <https://www.imdb.com/title/tt4660164/>.
- World Medical Association, 2025. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants. *JAMA* 333, 71–74. doi:10.1001/jama.2024.21972.
- 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., Swallow, K.M., Maley, C.J., 2010. The Brain’s Cutting-Room Floor: Segmentation of Narrative Cinema. *Frontiers in Human Neuroscience* 4, 168. doi:10.3389/fnhum.2010.00168.

Appendix A. Ablation Studies

To gain insight into which aspects of the cut-related EEG waveforms the DNN cut detector is sensitive to, we compared it with alternative basic cut-detection approaches informed by the observed TRP waveforms and the previous literature. These detectors were designed to capture different properties of the EEG response expected around cuts. They served as interpretable reference points for the DL approach and helped illustrate why accurate, temporally precise event detection is important for TRP analysis.

The TRP waveforms derived from the DNN model’s false-positive and false-negative detections revealed the following patterns (Figure A.8). First, anterior–posterior separation between 400 and 800 ms appeared significantly less prominent in both the false positive and false negative TRPs, than in the TRPs generated from the original (gold) cuts. Second, the false positive TRP lacks the distinct components such as P2 in the posterior average, while the false negative TRP still resembles the expected TRP waveform. Regarding the anterior–posterior divergence, when we shift the mean-correction interval from 0–200 ms to 100–300 ms, the false positive TRP shows stronger anterior–posterior separation.

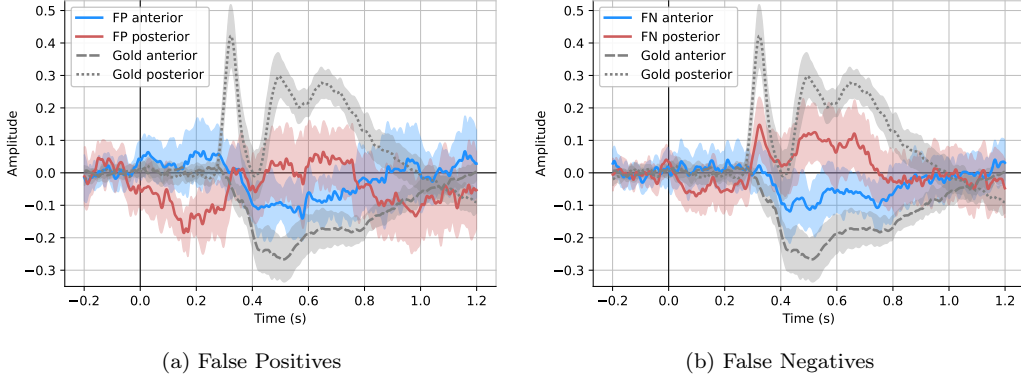


Figure A.8: Anterior and posterior TRPs drawn from false positive and false negative triggers produced by the DNN model in the film-generalization setting. Signals averaged over all films.

These differences between the false positive and false negative TRPs may arise because false positives are model-detected events and may therefore be less precisely time-locked, whereas false negatives are a subset of the true cut annotations. Nevertheless, these patterns motivate two basic cut-detection

strategies. First, if the DNN model is less likely to detect cuts when the anterior–posterior separation is less prominent, whereas its false positives still display such a separation, then a detector based on anterior–posterior divergence may provide a useful baseline for identifying film events that trigger cognitive updating. In other words, by such a detector we asked to what extent the DNN model’s performance could be explained by simply detecting anterior–posterior separation. Second, if the model misses cuts that still express ERP-like features, while producing false positives that do not, then a detector based on faithful TRP-template matching is worth investigating as a potential alternative.

We defined the anterior–posterior (A–P) divergence detector as follows. Electrodes were grouped into anterior¹ and posterior² regions, and the signal within each region was averaged across channels. The metric was computed within normalized (1 s) windows as the signed root mean square difference between the anterior and posterior regional averages. This score takes high values during periods in which the two regions diverge with the polarity pattern expected for cut-related responses. We then applied peak detection to the resulting score signal. Because the peaks corresponding to the strongest anterior–posterior separation occurred at a consistent latency after cut onset, we corrected the detected peak times by shifting them backward by a fixed temporal offset.

The second basic approach, the TRP-fit detector was calculated by fitting the cut-locked TRP template on the continuous EEG signal. We used cross-validation generalizing on films: the TRP template calculated from *Art* was fitted to the EEG signal of *City*, and vice versa. As in the A–P divergence cut-detector, the signals were first averaged within the anterior and posterior electrode groups. Separate TRP templates were then fitted to the anterior and posterior regional averages. The resulting fit score was evaluated over time on the test signal. Cut predictions were then obtained from this metric by peak-detection, as in the A–P divergence detector. To make the comparison less sensitive to slow drifts, the matching was performed on the discrete difference of the signals rather than on the raw waveform.

¹Fp1, Fp2, AF3, AF4, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, C3, C4.

²T7, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, P03, P04, O1, Oz, O2.

Appendix A.1. Correlation-Based Cut-Detectors

The correlation-based cut detectors were motivated by the assumption that salient events such as cinematic cuts, which may represent event boundaries in the film’s narrative (Zacks et al., 2010; Magliano and Zacks, 2011), evoke synchronized responses across viewers (Hasson et al., 2004; Dmochowski et al., 2012; Ki et al., 2016; Nastase et al., 2019). Under this assumption, cut-related time periods should be accompanied by increased similarity between the EEG signals of different subjects. The detectors therefore measure inter-subject similarity over time and use this signal to identify candidate cut locations. We compared multiple variants of the correlation-based cut detection.

- First, we measured the average inter-subject Pearson correlation within 0.5 s rolling windows, applied peak detection to the resulting time series, and shifted the peaks to the left by a constant amount, similar to the A–P divergence cut-detector.

Additionally, we trained the same peak-detection DNN architecture as discussed in Section 3.2, but instead of training on averaged EEG signals, we trained on correlation-derived signals:

- We trained one version on a 32-channel signal in which each channel contained the inter-subject Pearson correlation of the corresponding EEG channel within a 1 s moving window.
- Moreover, we trained a detector on a two-channel signal in which the two channels corresponded to the average correlations within the anterior and posterior electrode groups.

Appendix A.2. Ablation Results

We compared the basic cut detectors to the main DNN cut detector. All detectors were evaluated in the film-generalization cross validation scenario, meaning that where applicable, training was performed on both versions of one film and evaluation on both versions of the other film. As a reference, we also report a no-skill baseline, which provides an upper bound for a random detector constrained by the same minimum inter-peak distance. Because the minimum inter-peak distance was fixed to 0.625 s, and the acceptance window also spanned 0.625 s in total, this baseline distributed detections uniformly every 0.625 s throughout the recording. This construction achieves perfect

Model	F1	TP	FP	FN	Avg. distance
DNN	0.863	139	16	28	40 ms
Correlation DNN	0.610	111	86	56	90 ms
Correlation DNN AP	0.465	91	133	76	132 ms
Correlation	0.472	94	137	73	151 ms
A-P divergence	0.442	74	94	93	147 ms
TRP-fit	0.413	99	214	68	153 ms
No Skill	<i>0.278</i>	<i>167</i>	<i>868</i>	<i>0</i>	<i>156</i> ms

Table A.7: Evaluation metrics on *Art coherent* by model.

Model	F1	TP	FP	FN	Avg. distance
DNN	0.891	151	14	23	23 ms
Correlation DNN	0.707	116	37	59	74 ms
Correlation DNN AP	0.555	93	67	82	107 ms
Correlation	0.559	109	106	66	139 ms
A-P divergence	0.519	87	73	88	155 ms
TRP-fit	0.481	117	195	58	150 ms
No Skill	<i>0.355</i>	<i>175</i>	<i>636</i>	<i>0</i>	<i>156</i> ms

Table A.8: Evaluation metrics on *City coherent* by model.

Model	F1	TP	FP	FN	Avg. distance
DNN	0.953	161	12	4	29 ms
Correlation DNN	0.712	142	92	23	60 ms
Correlation DNN AP	0.509	110	157	55	122 ms
Correlation	0.490	106	162	59	144 ms
A-P divergence	0.516	89	91	76	164 ms
TRP-fit	0.397	104	255	61	163 ms
No Skill	<i>0.277</i>	<i>165</i>	<i>860</i>	<i>0</i>	<i>156</i> ms

Table A.9: Evaluation metrics on *Art incoherent* by model.

Model	F1	TP	FP	FN	Avg. distance
DNN	0.868	145	11	33	22 ms
Correlation DNN	0.710	114	29	64	66 ms
Correlation DNN AP	0.550	90	59	88	110 ms
Correlation	0.529	106	117	72	133 ms
A-P divergence	0.460	74	70	104	154 ms
TRP-fit	0.508	113	154	65	147 ms
No Skill	<i>0.381</i>	<i>178</i>	<i>575</i>	<i>0</i>	<i>156</i> ms

Table A.10: Evaluation metrics on *City incoherent* by model.

recall and yields maximal F1 score among no-skill detectors whose successive detections must be at least 0.625 s apart. Its purpose was to define a simple reference level against which the data-driven methods could be compared.

Tables A.7 to A.10 summarize the model performance across films and narrative versions. Among the models compared, the main DNN model detected cuts most accurately, achieving the highest F1 scores and the fewest errors across all conditions. The average distance (temporal error) of the detected true positive cuts by the DNN model was within one frame (40 ms at 25 frames per second). The simple peak-detection-based models performed better than random (no-skill baseline), however, their average temporal errors were close to 156.25 ms, the value expected if detections were distributed within the acceptance windows uniformly at random.

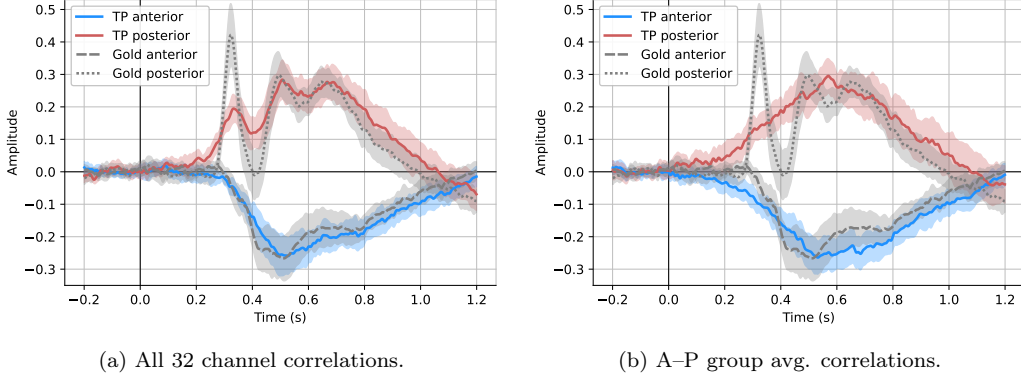


Figure A.9: Anterior and posterior TRPs drawn from the true positive triggers by the DNN models trained on channel correlations. Models generalized on the film, signals averaged over all films.

Frame-accurate cut detection is crucial for reliably reproducing TRP waveforms. Figure A.9 shows the TRPs generated from true positive detections of the DNN models trained on correlation data. The TRP generated from true positive detections of the 32-channel correlation DNN model is already heavily attenuated at an average temporal error of 71.67 ms, which is less than two frames (Figure A.9a). Moreover, at an average temporal error of less than three frames (117.95 ms), the anterior–posterior correlation-based DNN model already produces a true positive TRP waveform with no noticeable ERP-like features apart from basic anterior–posterior divergence (Figure A.9b).

Appendix A.3. Error Analysis

Film Narrative	<i>Art</i>		<i>City</i>		%
	Coherent	Incoherent	Coherent	Incoherent	
True positive	139	161	151	145	
False positive	16	12	14	11	
Transient	5	5	13	7	56.6%
Significant event	9	6	1	2	34.0%
Undetermined	2	1	0	2	9.4%
False negative	28	4	23	33	
Weak transient	2	0	7	8	19.3%
Sub-window	9	3	14	20	52.3%
Undetermined	17	1	2	5	28.4%

Table A.11: DNN cut detection outcomes by EEG recording in the film-generalization setting.

Film Narrative	<i>Art</i>		<i>City</i>		%
	Coherent	Incoherent	Coherent	Incoherent	
True positive	111	142	116	114	
False positive	85	92	37	29	
Transient	19	32	14	9	30.5%
Significant event	51	39	8	10	44.4%
Undetermined	15	21	15	10	25.1%
False negative	55	23	59	64	
Weak transient	4	0	9	13	12.9%
Sub-window	15	10	27	28	39.8%
Undetermined	36	13	23	23	47.3%

Table A.12: Cut detection outcomes by EEG recording for the 32-channel correlation DNN model in the film-generalization setting.

Although the correlation-based cut detectors performed worse than the main DNN model and operated with lower temporal precision, we hypothesized that they might identify more narratively relevant false positives. Tables A.11 and A.12 compare the manual error-analysis results for the main DNN model and the 32-channel correlation-based DNN model. As anticipated, the correlation-based model identified a larger number of narratively relevant events. However, these events cannot be directly used in the standard TRP framework, because they do not exhibit strong ERP-like features

when averaged around the detected peaks (Figure A.10). This suggests that correlation-based alternatives may be useful for discovering potentially interesting moments in the stimulus, but not directly for recovering meaningful event-related waveforms. Leveraging the additional potential of these significant false positive detections, would, therefore, require analysis methods beyond the scope of this study.

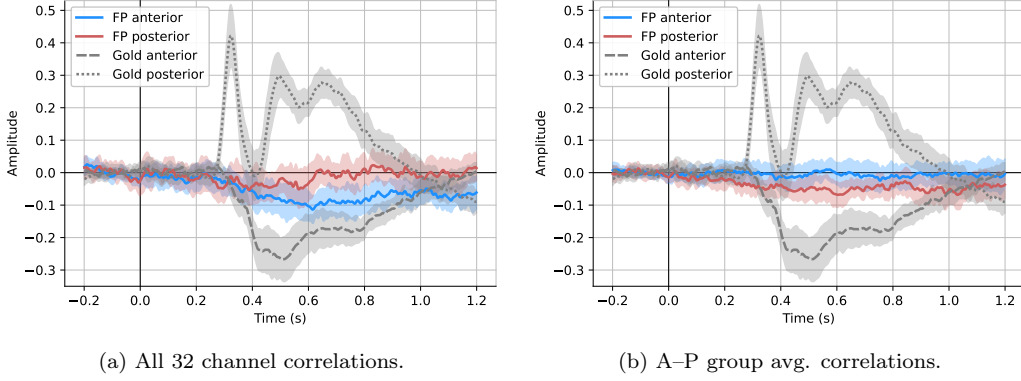


Figure A.10: Anterior and posterior TRPs drawn from the false positive triggers by the DNN models trained on channel correlations. Models generalized on the film, signals averaged over all films.

Appendix B. Additional Figures

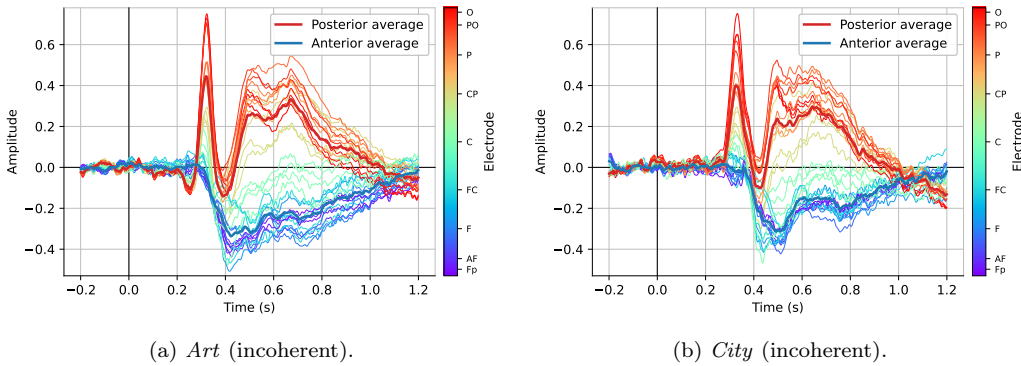


Figure B.11: Butterfly plots of the average anterior and posterior TRP signals aligned to all cuts in the incoherent versions of the films.

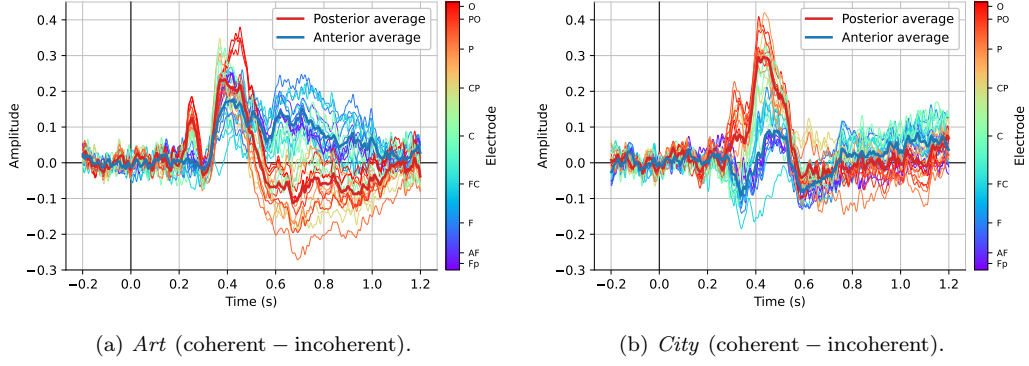


Figure B.12: Butterfly plots of the coherent – incoherent average TRP signals.

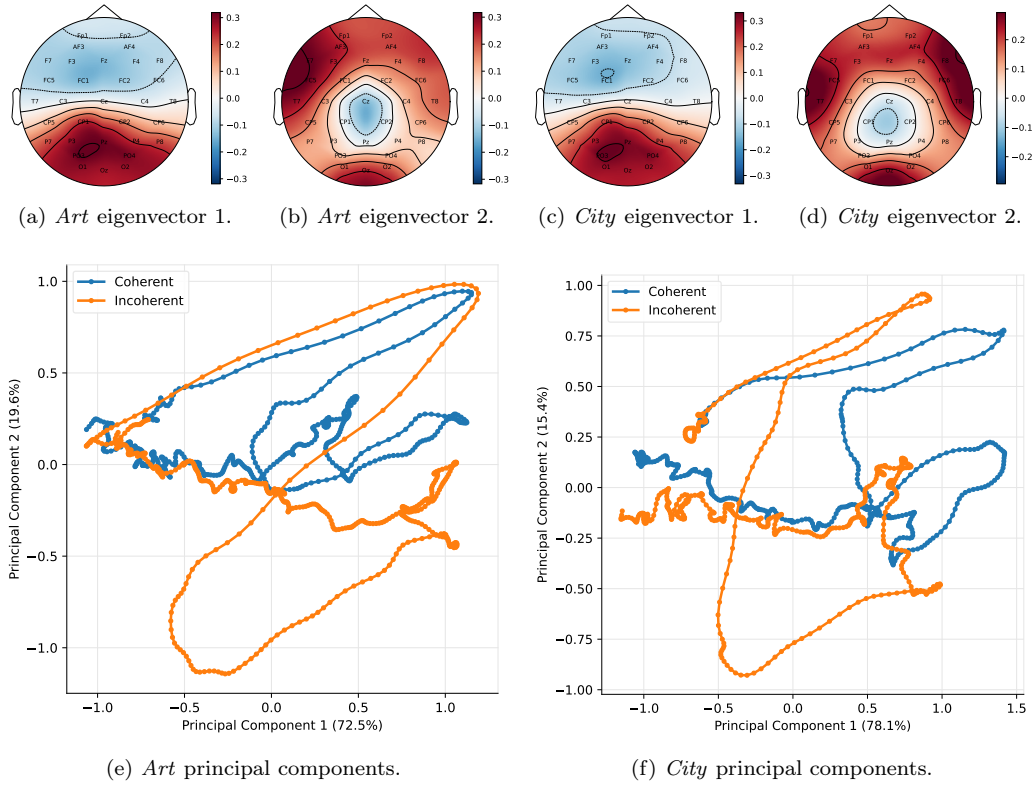
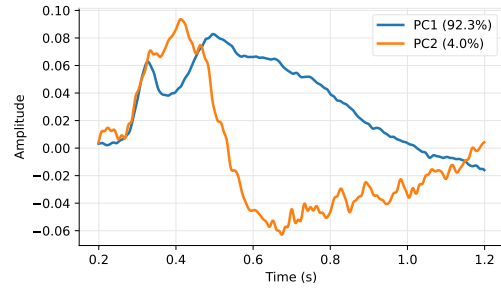


Figure B.13: Spatial PCA analysis of the scene-matched hand-annotated cut-TRPs.



(b) *City* principal eigenvectors.

