Higher levels of narrativity lead to similar patterns of EEG activity across individuals

Hossein Dini[1¶*], Aline Simonetti[2¶], Enrique Bigne[2], Luis Emilio Bruni[1*]

[1] Aalborg University, The Augmented Cognition Lab, Copenhagen, 2450, Denmark

[2] University of Valencia, Department of Marketing and Market Research, Valencia, 46022, Spain

* Corresponding authors

E-mail: hdi@create.aau.dk (HD); leb@create.aau.dk (LB)

¶ These authors contributed equally to this work

Abstract

The focus of cognitive and psychological approaches to narrative has not so much been on the elucidation of important aspects of narrative, but rather on using narratives as tools for the investigation of higher order cognitive processes elicited by narratives (e.g., understanding, empathy, etc.). In this study, we work toward a scalar model of narrativity, which can provide testable criteria for selecting and classifying communication forms in their level of narrativity. We investigated whether being exposed to videos with different levels of narrativity modulates shared neural responses, measured by inter-subject correlation, and engagement levels. Thirty-two participants watched video advertisements with high-level and low-level of narrativity while their neural responses were measured through electroencephalogram. Additionally, participants' engagement levels were calculated based on the composite of their self-reported attention and immersion scores. Results demonstrated that both calculated inter-subject correlation and engagement scores for high-level video ads were significantly higher than those for low-level, suggesting that narrativity levels modulate inter-subject correlation and engagement. We believe that these findings are a step toward the elucidation of the viewers' way of processing and understanding a given communication artifact as a function of the narrative qualities expressed by the level of narrativity.

Introduction

Attention is involved in all cognitive and perceptual processes [1]. To some degree, an attentive state toward an external stimulus implies the silencing of internally oriented

mental processing [2]. Sufficiently strong attentive states can hamper conscious awareness of one's environment and oneself [3]. In contrast to neutral stimuli, emotional stimuli attract greater and more focused attention (see [4] for a review of the topic). As stories can generate and evoke strong feelings [5], the pathos implicit in narratives could be considered attention-seeking stimuli. In fact, people tend to engage emotionally with stories [5]. Attentional focus to a narrative stimulates complex processing [6], with narratives inducing "emotionally laden attention" [2]. In the narrative domain, attentional focus and the sense of being absorbed into the story are part of narrative engagement [3]. Indeed, researchers have suggested that narratives' inherent persuasiveness is related to the feelings of immersion they evoke, a phenomenon that has been termed "transportation effects" [7]. In this context, "transportation" indicates a combination of attention, feelings, and imagery where there is a convergent process of different perceptual, cognitive and affective systems and capacities to the narrative events [7]. During narrative comprehension, involuntary autobiographical memories triggered by the story in question appear to impair attention only momentarily [8]. This is not the case for either internally or externally generated daydreaming or distraction [8]. Thus, narrative engagement can be hampered when thoughts unrelated to the narrative arise [3]. In addition, the degree to which audiences engage with a narrative varies based on delivery modality (e.g., audio or visual) [9], during narrative comprehension [10], and across individuals [11,12]. Nonetheless, narratives are naturally engaging [12], and they reflect daily experiences, making them potentially useful devices for understanding cognitive processes such as engagement.

The use of simplified and abstract stimuli has been common in cognitive neuroscience for decades (e.g.: [13]). Although such stimuli enable highly controlled experiments and isolation of the study variables, they tend to lack ecological validity. Hence, recent

studies have employed stimuli that simulate real-life situations, including narratives, termed “naturalistic stimuli” (see [12] for a review of the topic). From this perspective, advertising is considered a naturalistic stimulus that is designed to be emotionally persuasive [12]. Video ads provide narrative content in short period of time (e.g., 30 seconds); hence, advertisers focus on delivering a key message through stories with different narrativity levels. In fact, companies use narrative-style ads because stories captivate, entertain, and involve consumers [14,15]. Researchers have found that, at the neural level, narratives (particularly more structured ones) tend to induce similar affective and cognitive states across viewers [2,10]. However, whether consumers perceive stories in similar ways is of great interest, as this might affect whether they ultimately engage with the advertisement as expected. To investigate the underlying cognitive processes generated by narratives, many studies have used traditional methods such as electroencephalogram (EEG) power analysis [16]. Though valid, the metrics provided by traditional methods may not be ideal when considering narrative as a continuous stimulus, given that such methods could generally require stimulus repetition. In addition, they do not capture how the same information is processed across individuals.

Inter-subject correlation (ISC) is an appropriate neural metric for investigating shared neural responses, especially when using naturalistic stimuli. This data-driven method assumes the occurrence of common brain reactions to a narrative, which improves the generalizability of the findings. By correlating neural data across individuals, this metric can identify localized neural activities that react to a narrative in a synchronous fashion (i.e., in a time-locked manner) [17]. ISC is well-suited to analysis of both functional magnetic resonance (fMRI) [18] and EEG [19,20] data. ISC obtained using naturalistic stimuli has been used to investigate episodic encoding and memory [10,21–23], social

interaction [24], audience preferences [25], information processing [26], narrative comprehension [27], and, similar to this study, attention and engagement [2,10,11,19,28,29].

ISC calculated from EEG data has been shown to predict levels of attentional engagement with auditory and audio-visual narratives [11,29]. Cohen et al. [29] found that neural engagement with narratives measured through ISC was positively correlated with behavioral measurements of engagement, including real-world engagement. Attention was also shown to modulate (attended) narrative processing at high levels of the cortical hierarchy [26]. Ki et al. [11] found that ISC was weaker when participants had to concurrently perform mental arithmetic and attend to a narrative than when participants only attended to the narrative. This was especially the case when participants attended to audiovisual narratives compared to auditory-only narratives. They also reported that the ISC difference between the two attentional states, as well as its magnitude, was more pronounced when the stimulus was a cohesive narrative than when it was a meaningless one. Moreover, they demonstrated that audiovisual narratives generated stronger ISC than audio-only narratives. In line with this finding, a recent fMRI study indicated that sustained attention network dynamics are correlated with engagement while attending audiovisual narratives. However, this is not the case for audio-only narratives. These results might suggest that the sustained attention network is visually specific [10]. Furthermore, the authors found higher interpersonal synchronization of default mode network activity during moments of higher self-reported engagement with both audio-visual and audio-only narratives [10]. Finally, ISC was observed to diminish when viewers watched video clips for the second time [2,19,28], which was supported by a study showing lower ISC when viewers were presented with a scrambled narrative for the second time compared to the first time [27].

In summary, while some previous studies have examined the relationship between ISC and narrative engagement (or its correlate, attention) [2,10,11,26,28,29], they assumed that "being a narrative," is an either-or quality. From the narratological perspective, there has been increased interest in "narrativity" as a scalar property; that is, artifacts (pictures, videos, stories, and other forms of representation) may have different degrees of narrativity. In other words, narrative may be a matter of more-or-less rather than either-or [30]. Therefore, how different levels of narrativity affect narrative engagement remains an open question. Moreover, whether narrativity levels lead to different degrees of ISC, and whether it is possible to predict narrativity level based on ISC values, has not yet been explored. Therefore, our main research question for this study is as follows: does narrativity level (high [HL] vs. low [LL]) modulate self-reported engagement and ISC? To answer this question, we asked 32 participants to watch 12 real video ads twice each (with sound removed) varying in their degree of narrativity. We defined "narrativity level" as the degree to which an artifact is perceived to evoke a complete narrative script [30]. Essentially, this refers the degree to which a story seems complete in all its structural elements, such as the presence of defined characters in a stipulated context, clear causal links between events, and a sense of closure resulting from the intertwining of these events [30]. While participants watched the videos, we recorded their EEG brain signals to investigate shared neural responses to the video ads across participants; that is, we investigated the interpersonal reliability of neural responses (expressed by ISC). We further assessed how narrativity level affected self-reported engagement with the ads. Thus, we asked participants to rate how much each ad captured their attention and how immersed in it they felt. Furthermore, we examined how engagement and ISC differ for video ads watched for the first time versus the second time. Based on the aforementioned studies, which explore how narratives

engage individuals and indicate that ISC levels are higher during moments of higher narrative engagement and, we postulate that HL video ads will have higher ISC than LL video ads and that self-reported engagement will be higher for HL ads than for LL ads. Moreover, we expect that both engagement and ISC will be higher the first time the videos are seen compared to the second time.

Results

Online validation test

Of the 156 participants who completed the online test, we considered 124 responses to be valid as the remaining participants answered too fast or failed to answer the attention question correctly. Results of the online test confirmed that the general public also perceived ads in the HL category as having a high narrativity level (M_{HL} = 70.95, SD_{HL} = 10.79) and adds in the LL category as having a low narrativity level (M_{LL} = 47.71, SD_{LL} = 13.52; see Fig 5-e in the materials and methods section). All six of the HL ads received higher scores than any of the six LL ads. An independent sample t test confirmed that the difference between the two categories was statistically significant ($t(61) = 7.554$, $p < 0.001$).

Narrativity level modulates self-reported engagement

High levels of focused attention and high levels of immersion (feelings of transportation) can indicate high levels of narrative engagement. Thus, we calculated our engagement metric by averaging attention and immersion scores. We first report our descriptive analysis of the attention, immersion, and engagement scores for each of the

four conditions (HL–PC, HL–VR, LL–PC, and LL–VR; see "materials and methods" section for abbreviations), then report results of the statistical analysis conducted on engagement scores. Attention scores for the PC conditions were as follows: $M_{HL\text{-}PC}$ was 57.03 ($SD_{HL\text{-}PC}$ = 14.76), and $M_{LL\text{-}PC}$ was 44.16 ($SD_{LL\text{-}PC}$ = 13.92). The attention scores for the VR conditions were as follows: $M_{HL\text{-}VR}$ was 56.26 ($SD_{HL\text{-}VR}$ = 14.19), and $M_{LL\text{-}VR}$ was 46.28 ($SD_{LL\text{-}VR}$ = 14.41). Fig 1-a shows the average attention scores in each of the four conditions. Immersion scores for the PC conditions were as follows: $M_{HL\text{-}PC}$ was 54.26 ($SD_{HL\text{-}PC}$ = 13.28), and $M_{LL\text{-}PC}$ was 42.53 ($SD_{LL\text{-}PC}$ = 15.09). Immersion scores for the VR conditions were as follows: $M_{HL\text{-}VR}$ was 53.67 ($SD_{HL\text{-}VR}$ = 10.81), and $M_{LL\text{-}VR}$ was 44.56 ($SD_{LL\text{-}VR}$ = 14.15). Fig 1-b shows the average immersion scores in each of the four conditions. Engagement scores for the PC conditions were as follows: $M_{HL\text{-}PC}$ was 55.65 ($SD_{HL\text{-}PC}$ = 13.54), and $M_{LL\text{-}PC}$ was 43.35 ($SD_{LL\text{-}PC}$ = 14.04). Engagement scores for the VR conditions were as follows: $M_{HL\text{-}VR}$ was 54.97 ($SD_{HL\text{-}VR}$ = 12.16), and $M_{LL\text{-}VR}$ was 45.42 ($SD_{LL\text{-}VR}$ = 14.00). Results of the 2 × 2 repeated measures ANOVA showed a main effect of narrativity level ($F(1,28) = 19.779$, $p < .001$). Post-hoc analysis showed that participants engaged more with HL ads than with LL ads in both PC ($F(28,1) = 17.03$, $p < .001$) and VR ($F(28,1) = 20.320$, $p < .001$) conditions. See Fig 1-c.

Fig 1 about here

Fig 1: Participants' self-reported scores. Each bar is the average of the scores across participants and across corresponding video ads. The colored dots represent the score for each participant averaged across the corresponding videos. **(a)** Self-reported attention scores. **(b)** Self-reported immersion scores. **(c)** Engagement scores (average of

attention and immersion scores). This shows that the HL engagement scores were significantly higher than the LL engagement scores in both PC ($p < .001$) and VR ($p < .001$) conditions.

Narrativity level modulates inter-subject correlation

We applied the same statistical analysis used for the self-reported data to assess whether our neurological metric (ISC) changed based on narrativity level. To do so, we used summed ISC over the first two strongest components (see Materials and Methods section) as a representative of neural reliability. Results of the 2 × 2 repeated measures ANOVA showed a main effect of narrativity level, where ISC was higher when participants watched the HL ads than when they watched the LL ads ($F\ (1,28) = 4.467$, $p = .044$). Although the interaction effect was not significant, we then conducted post-hoc analysis to determine whether this difference occurred in the PC or VR condition. Results of the post-hoc analysis showed that ISC was significantly higher for the HL ads ($M_{HL} = 5.47 \times 10^{-3}$, $SD_{HL} = 2.97 \times 10^{-3}$) than for the LL ads ($M_{LL} = 3.48 \times 10^{-3}$, $SD_{LL} = 2.51 \times 10^{-3}$) in the PC condition ($F\ (28,1) = 5.266$, $p = .017$). In the VR condition, the ISC of HL ads ($M_{HL} = 8.72 \times 10^{-3}$, $SD_{HL} = 4.62 \times 10^{-3}$) was still higher than that of LL ads ($M_{LL} = 8.35 \times 10^{-3}$, $SD_{LL} = 3.92 \times 10^{-3}$), but this difference was not significant ($F\ (1,28) = 0.226$, $p = .638$). See Fig 1-a.

Furthermore, to investigate which factors affect neural reliability, we compared the ISC of HL video ads to that of LL video ads separately for each component. We corrected the p-values because we repeated our calculations twice for the two components, and results can be seen in the top part of Fig 2-b. Results showed that, for the first component, the main effect of narrativity is not significant ($F\ (1,28) = 1.145$, *corrected* $p = .294$). However, for the second component, the main effect of narrativity was

significantly higher in HL video ads than in LL video ads (*F (1,28)* = 5.412, *corrected p* = .045). Post-hoc analysis of the second component revealed that the ISC of HL ads (M_{HL} = 2.65 × 10^{-3}, SD_{HL} = 3.34 × 10^{-4}) was significantly higher than that of LL ads (M_{LL} = 9.61 × 10^{-4}, SD_{LL} =2.99 × 10^{-4}) in the PC condition (*F (1,28)* = 13.676, *corrected p* = .002.); however, there was no significant difference between HL and LL ads in the VR condition (*F (1,28)* = 1.056, *corrected p* = .313). In addition, different correlated components had different patterns ofscalp activities (Fig 2-b, bottom part). For the first component, activity was distributed throughout the frontal and occipital regions, while in the second component, activity was concentrated in the occipital region. Overall, the second component showed the strongest narrativity level modulation in response to the HL narrative video ads.

Fig 2 about here

Fig 2: (a) Summed ISC over the first two strongest components in four conditions. This shows that, in the PC condition, the ISC of HL video ads was significantly higher than that of LL video ads (*p* = .029), and in the VR condition, the ISC of HL video ads was still higher than that of LL video ads, but not significantly so (*p* = .638). **(b-top)** Calculated ISC separately for each component and within each condition. This shows that, for the first component, the difference between HL and LL video ads is not significant in either the PC or the VR condition. However, in the second component, the ISC of HL video ads was significantly higher than that of LL video ads (*p* = .002) in PC condition, but not in the VR condition (*p* = .313). The gray bars show the average calculated chance level for each condition **(b-bottom)** Scalp activity of first two

strongest components. Activity in component 1 is throughout the frontal occipital area, while in component 2, activity is concentrated in the occipital area. In both figures, the colored dots represent the calculated ISC values for each participant.

Viewing order modulates self-reported engagement but not inter-subject correlation

To evaluate whether viewing order affects engagement scores and ISC, we implemented the abovementioned statistical procedure with narrativity level and viewing order as independent variables. Participants watched each video twice: once in one medium and once in the other medium. The medium that was used first was counterbalanced across participants. For the analysis, we created a group for the first viewing and a group for the second viewing, regardless of the medium used. In both of these groups, participants watched the same videos, including both HL and LL video ads. We first report the resulting engagement scores and ISC scores.

Regarding engagement scores, the average and standard deviation for the first viewing group were as follows: M_{HL} was 57.71 (SD_{HL} = 2.10), and M_{LL} was 45.42 (SD_{LL} =2.65). For the second viewing group, these figures were as follows: M_{HL} was 52.90 (SD_{HL} = 2.57), and M_{LL} was 43.36 (SD_{LL} =2.56; see Fig 3-a). Results of this statistical analysis revealed a significant main effect of viewing order. Further post-hoc analysis indicated that this main effect was driven by the HL category, for which videos viewed first received significantly higher engagement scores than those viewed second ($F(1,28)$ = 12.784, p = .001). However, in the LL condition, there was no significant difference between the first and second viewing groups ($F(1,28)$ = 1.686, p = .205). Moreover, there was a main effect of narrativity level: of videos in the first viewing group, HL videos received significantly higher engagement scores than LL videos ($F(1,28)$ =

23.047, $p < .001$); this was also the case for videos in the second viewing group ($F(1,28) = 13.391$, $p = .001$).

Regarding ISC, the average and standard deviation for videos in the first viewing group were as follows: M_{HL} was 6.92×10^{-3} ($SD_{HL} = 0.74 \times 10^{-3}$), and M_{LL} was 5.97×10^{-3} ($SD_{LL} = 0.81 \times 10^{-3}$). For the second viewing group, these figures were as follows: M_{HL} was 7.27×10^{-3} ($SD_{HL} = 0.82 \times 10^{-3}$), and M_{LL} was 5.88×10^{-3} ($SD_{LL} = 0.71 \times 10^{-3}$; see Fig 3-b). Results of this statistical analysis showed no significant main effect of viewing order. Further post-hoc analysis showed that the HL video ads received higher ISC scores than LL video ads upon second viewing ($F(1,28) = 4.411$, $p = .045$), but not upon first viewing ($F(1,28) = 1.144$, $p = .294$).

Fig 3 about here

Fig 3: Results of viewing order (first and second) considering narrativity level (HL and LL). **(a)** Results of engagement scores, where for videos in the HL condition, engagement scores upon first viewing were significantly higher than those upon second viewing. Moreover, HL videos received significantly higher engagement scores than LL videos upon both first and second viewings. (**b)** Results for calculated ISC of viewing order, where the only significant difference was between the second viewing of HL videos and that of LL videos.

ISC-HL

We calculated ISC-HL to test whether it was possible to determine whether a single subject was watching an HL video ad or an LL video ad based on their neural activity.

We calculated the projections based only on data from the HL group. These projections represent the overall brain activity of all participants attending to HL videos. Based on these projections, we calculated the ISC-HL for participants from each group (HL and LL) separately. We hypothesized that the brain activity of participants from the HL group would be more similar to the overall activity of all participants in the HL group (meaning higher ISC-HL) and that the brain activity of participants from the LL group would be less similar to the overall activity of all participants in the HL group (meaning lower ISC-HL). Fig 4-a and Fig 4-b display the calculated ISC-HL in both VR and PC conditions. As expected, the neural responses of 23 (76.6%) participants were much more similar to those of the HL group (had higher ISC-HL) when they attended to HL videos than when they attended to LL videos in the PC condition (Fig 4-a). In the VR condition, the neural responses of 15 (50%) participants showed higher ISC-HL while attending to HL videos than while attending to LL videos (Fig 4-b). To assess the classification performance of the support vector machine classifier with 10-fold cross-validation for the two groups (attended to HL videos vs. attended to LL videos), we calculated the area under the curve (AUC). To do so, we first calculated the actual AUC value by calculating the projections using actual labels, then randomly shuffled the labels 1,000 times. For each iteration, we calculated new projections and computed the AUC. The average AUC over 1,000 iterations was the chance-level AUC. Given ISC-HL as the predictor of narrativity level, the classifier shows above-chance performance in the PC condition but not in the VR condition (Figs 4-c and 4-d). In the PC condition, the actual AUC was .637, while the chance-level AUC was .548. In the VR condition, the actual AUC was .507, while the chance-level AUC was .543. Therefore, in the PC condition, the ISC-HL was able to predict the narrativity level based on neural responses (measured via EEG) to the stimuli. This means that the model significantly

predicted whether the participants were attending to HL or LL narrativity. The level of significance was not as strong as the previously reported accuracy for predicting attentional states [11]; however, this could be due to explicit differences between experimental conditions (participants were made to count backward during the task to diminish their attentional state). This finding is the first step toward predicting exposure to different levels of narrativity based solely on the ISC of neural activity evoked by stimuli, which leads to a better comprehension of the brain mechanism that processes narrativity levels. Further studies with different designs should be conducted with a more specific focus on predicting narrativity level.

Fig 4 about here

Fig 4: (a) and (b) Results for calculated ISC-HL for each participant in the PC and VR conditions. Lines and markers of each color show the ISC-HL of one participant in the HL and LL conditions. **(a)** This shows that 76.6% of participants had higher ISC-HL when they attended to HL video ads than when they attended to LL video ads. This shows that most of the participants who were exposed to HL video ads showed higher similarity to the trained model (which is derived from data of HL condition) than those who are exposed to LL video ads. **(b)** This relates to the same evaluation as (a), but in the VR condition, where 50% of participants had higher ISC-HL. **(c)** and **(d)** AUC of classification performance for predicting exposure to HL and LL video ads in PC and VR conditions, respectively. The gray area shows the chance-level performance. **(c)** shows that the AUC of classification based on participants' neural activity in the PC

condition is higher than chance level, while **(d)** shows that the AUC of classification based on participants' neural activity in VR condition is below chance level.

Discussion

In this study, we tested whether different levels of narrativity (HL vs. LL) lead to differences in information processing reliability (represented by ISC), specifically while watching video ads. Furthermore, we evaluated whether different levels of narrativity cause differences in self-reported engagement ratings. To this aim, we presented HL and LL video ads to 32 participants while collecting their EEG signals. In addition, for each video ad, participants self-reported their levels of attention and immersion, which we considered two core factors of engagement. We calculated the ISC of each participant by calculating the similarity of their correlated components to those of the participant pool. One advantage of ISC analysis is that it does not require stimulus repetition; this is advantageous because such repetition causes decreased attention and engagement [2,11]. As expected, our results showed that both calculated ISC and engagement scores of HL video ads were significantly higher than those for LL video ads, suggesting that narrativity level modulates ISC and engagement.

Previous studies have evaluated the relationship between narrative engagement—or its core component, attention—and ISC. Cohen et al. measured self-reported engagement scores and ISC while participants were exposed to naturalistic videos [29]. They found that more engaging videos were processed uniformly in participants' brains, leading to higher ISC. Consistent with their results, Poulsen et al. reported that a lack of engagement manifests an unreliable neural response (meaning lower ISC), and they introduced EEG-ISC as a marker of engagement [28]. Song et al. investigated whether

engagement ratings modulate ISC by evaluating the relationship between continuous self-reported engagement ratings and ISC using continuous naturalistic stimuli [10]. They reported higher ISC during highly narrative engaging moments and concluded that ISC reflects engagement levels. In another study, Dmochowski et al. used short video clips to evaluate attention and emotion using ISC [2]. They found a close correspondence between expected engagement and neural correlation, suggesting that extracting maximally correlated components (ISC) reflects cortical processing of attention or emotion. Finally, Ki et al. investigated whether attentional states modulate ISC for audio and audiovisual narratives, and they concluded that higher attention leads to higher neural reliability across subjects [11]. Inspired by previous studies [2,3,31], we measured narrative engagement by averaging two of its important components, attention and immersion.

Our results indicated that, using ISC (in the PC condition), we were able to significantly discriminate levels of narrativity: the ISC levels of HL video ads were significantly higher than those of LL video ads (Fig 2-a). Consistent with our narrativity level discrimination, self-reported engagement ratings were significantly higher for HL video ads than for LL ones (Fig 1-c).

In summary, this study aligns with previous findings: it indicates that more engaging video ads lead to more similar neural activity across participants, represented by higher ISC. This ISC–engagement level relationship contributes to a better understanding of the underlying reasons for the possible association between narrativity level and shared neural responses.

Several studies compared ISC scores of interrupted or scrambled narratives with those of non-scrambled narratives. Dmochowski et al. reported higher ISC for non-scrambled narratives than for disrupted or scrambled narratives [2]. In addition, Ki et al. and

Poulsen et al. found higher ISC scores for a cohesive narrative than for a meaningless, scrambled narrative [11,28]. These findings demonstrate that scrambled narratives are linked to lower ISC scores. Song et al. used scrambled video clips to identify the moments in which narrative comprehension occurs [27], and they reported that story comprehension occurs when events are causally related to each other. Moreover, they showed that, in such moments, the underlying brain states were mostly correlated across subjects. The stimuli manipulations applied by previous studies (e.g., scrambling the narrative or presenting narratives that lacked causality) created differences between experimental conditions similar to those in our study. However, while we shared with these studies their interest in the subjects' engagement and other factors related to the reception of the narrative, we were also interested in investigating the possibility of classifying and discriminating structural characteristics of narrative artifacts. Therefore, we defined the narrativity levels (HL and LL) based on narrative structural elements and properties such as the presence of defined characters in an identifiable context, clear causal links between events, and closure resulting from the intertwining of these events. Our conceptualization of narrativity levels was inspired by Ryan [30].

From this perspective, our results align with those of previous studies, indicating that information processing is highly consistent across subjects when participants are exposed to HL narrativity. In other words, the fact that the ISC of HL video ads is significantly higher than that of LL video ads in the PC condition (Fig 2-a) is a promising indication that EEG-ISC could potentially be used to explore levels of narrativity (and correlative engagement) in different kinds of media artifacts. Furthermore, we evaluated the activity of the first two components separately (Fig 2-b). Results showed that the second component (in the PC condition) was most reliably modulated by the narrativity level. However, further investigation is necessary to

elucidate why narrativity level didn't appear to affect ISC in the same way in the VR condition. Chang et al. reported that ISC might change due to fatigue effects for participants in an fMRI scanner [32]. Thus, one possible explanation for the non-significant effect of narrativity level on ISC in the VR condition is fatigue. Wearing the VR headset and the EEG cap for almost 25 minutes might have caused fatigue and discomfort (e.g., related to posture, weight, or itching), and such an effect might have masked the effects of narrativity level on ISC. Another plausible explanation is also related to the VR feature. As VR is an increasingly popular immersive technology, participants' expectations about the VR modality may have affected their level of attention. Participants may have been disappointed to find a 2D stimulus that perhaps failed to meet their expectations and therefore diminished their active attention.

In the narratology field, there is a substantial, ongoing debate about whether narrative should be considered an either-or property or a scalar property (i.e., a matter of more or less) [30,33]. Therefore, in the last two decades, some narratologists have introduced the notion that different artifacts may have different levels of narrativity [30,34]. It is widely accepted that, even though events in a story do not have to be chronologically ordered, the sequence of the events must nonetheless follow a narrative logic if closure is to be achieved (this is something that is commonly encapsulated in the distinction between "story" and "discourse"). Therefore, a "scrambled" narrative ceases to be a narrative; instead, it is a series of unrelated events, lacking an internal temporal logic. If there is no discernable, recoverable chronological order of connected events, the sequence could hardly be considered a narrative. Although the previous studies presented in our literature review provided the methodological bases for our study, they were not centered on narrative qualities. In our study, we focus on investigating the plausibility of testing a particular expressive artifact for levels of narrativity. Therefore,

our stimuli presented two different degrees of narrativity (LL and HL). We showed that small differences in narrativity levels (i.e., between HL and LL) have effects on ISC and engagement similar to the effects of more evident differences between such levels (e.g., scrambled vs. non-scrambled). These findings are a step toward narrative comprehension, especially when considering narrative as a scalar property [35].

To test whether viewing order modulates engagement and ISC, we separated the data into two groups—first and second viewing—and conducted statistical analyses considering narrativity level and viewing order to be independent variables. Domchowski et al. found that attentional engagement decreases when participants watch a stimulus for the second time compared to the first time [2]. Moreover, they reported significantly lower ISC for the second viewing. Ki et al. confirmed and extended their results by declaring that neural responses become less reliable upon second viewing of a stimulus, showing significantly lower ISC [11]. In an fMRI study, Song et al. investigated the effect of viewing order on neural activity, watching scrambled videos twice [27]. Their results replicated the findings of the aforementioned study and other studies [19,28] by showing that neural states across participants were less synchronized when watching the videos, including the scrambled videos, for the second time. However, Chang et al. conducted a combined EEG-magnetoencephalography (MEG) study and reported increased ISC during the second viewing of the stimuli [32]. They declared that participants' prediction of the story structure increased during the second viewing, resulting in a more similar EEG-MEG activity across participants, and that this contradiction with previous studies might be due to the different time window selected for ISC calculation. Our results showed that engagement level dropped during the second viewing, supporting previous studies [2]. For the HL videos ads, engagement scores for the first viewing were significantly

higher than those for the second viewing, although this was not the case for the LL ads (Fig 3-a). The decrease in engagement scores for HL ads but not LL ads might be due to differences intrinsic to the classification of our stimuli. While videos in the HL category included all or almost all narrativity elements [30], those in the LL category included only a few. Thus, we could say that the HL video ads were more storytelling-based than the LL video ads. Therefore, it could be that factors such as suspense, expectation, and suspension of disbelief, which naturally occur when watching stories, are attenuated when the story is viewed for the second time, lowering engagement levels in the case of HL video ads. These same factors are not present or are present to a lesser extent when watching LL video ads, and therefore, engagement levels were not harmed. However, there was no significant difference in ISC between the first and second viewing in either HL or LL video ads, suggesting that viewing order does not modulate ISC. Though the current dataset cannot provide definite answers, these differences in findings might be explained by our study design. First, our participants were exposed to 12 short video ads, a greater number of stimuli than previous studies employed when investigating viewing order [2,11,27,28,32]. Watching 12 videos in sequence might have reduced participants' memories of details of the stories. Hence, when watching the videos for the second time (although they were perceived as less engaging) there remained a substantial amount of information to be processed, which could have been reflected in the ISC of the second viewing. Another possible factor that might have hindered participants' short-term memories of the videos and affected ISC levels is the time span and the tasks performed between the two viewings. In our study, the second exposure to the video ads was separated from the first by about 20 min. During those 20 min, participants performed two different tasks for a separate study. Therefore, the current

dataset was not able to capture the neural underpinnings of viewing order. Considering the limitations of this study, future studies could be conducted to capture this effect.

Materials and methods

This study was approved by the local ethics committee (Technical Faculty of IT and Design, Aalborg University) and performed in accordance with the Danish Code of Conduct for Research and the European Code of Conduct for Research Integrity. All participants signed an informed consent form at the beginning of the session, were debriefed at the end of the experiment, and were given a symbolic payment as a token of gratitude for their time and effort. Data collection took place throughout November 2020. This study was part of a larger study; here, we report only the information relevant to the present study.

Participants and stimuli

We recruited 32 (13 women) right-handed participants of 16 nationalities between the ages of 20 and 37 (M = 26.84, SD = 4.33). Regarding occupation, 69% were students, 16% were workers, and 15% were both. Regarding educational level, 12% had completed or were completing a bachelor's degree, and 88% had completed or were completing a master's degree. We requested that participants not drink caffeine products at least two hours prior the experimental session. A statistical power analysis of the ISC effect, with alpha equal to 0.05 and power equal to 80%, was used to determine the sample size. The largest sample size required by this analysis was 32 participants.

To select the stimuli, we initially pre-selected 22 video ads (eight brands; Fig 5-a) varying in the number of elements that constitute different narrative levels according to

Ryan [30]. The narrativity levels of these video ads were then preliminarily rated from 0 to 100 by an expert in the narrative field. Based on these scores, we selected six brands that had the largest difference in score between two of their ads (to maximize the difference in narrativity level between two ads of the same brand), coinciding with an absolute score below 50 for one ad and above 50 for the other ad (Fig 5-b). The final selection consisted of six video ads with HL narrativity (M = 83.17, SD = 10.70) and another six with LL narrativity (M = 33.33, SD = 10,70), with each of the six brands contributing one video to each of the two groups (Fig 5-c). To validate the categorization of the ads into these two categories, another independent expert in the narrative field rated the narrativity level of the 12 video ads from 0 to 100, and their ratings were consistent with those of first expert (HL: M = 81.00, SD = 13.70; LL: M = 37.33, SD = 15.37; Fig 5-d). To assess whether the general public would also perceive a difference in narrativity level between the two categories of videos, we conducted an online validation test using the Clickworker platform (https://www.clickworker.com) with an independent panel of participants, which confirmed that our classification was valid (see the results section). For the online test, a total of 156 participants were assigned to one of four groups. Each group was assigned to watch three video ads, each of which was from a distinct brand. To capture the degree of perceived narrativity in each video ad, participants were asked to answer five questions used by Kim et al. [36], such as “the commercial tells a story” and “the commercial shows the main actors or characters in a story.” All questions were rated from 0 (strongly disagree) to 100 (strongly agree). Average responses to these five questions showed whether the videos ads were indeed perceived to be in the HL (higher scores) or LL (lower scores) category. The responses confirmed that our initial categorization of the videos was valid (see results and Fig 5-e).

The stimuli therefore comprised 12 2D video ads from six brands (Barilla, Coke, Disney, Kellogg's, Nike, and Oculus), with two video ads from each brand, one for each narrativity level. We selected different brands to mitigate the influence of brand and product category. The ads were real commercials retrieved from YouTube. We removed the audio and edited some of the videos slightly to adjust the length (which varied from 57 to 63 seconds). We selected non-verbal narratives in video format because motion picture narratives are less susceptible to interindividual differences and generate more homogeneous experiences across individuals than verbal (oral) narratives [37]. This may be because the visual images are directly related to the narrative content, which reduces personal interpretations of the story [9].

Fig 5 about here

Fig 5: Video ads selection procedure, division into HL and LL categories, results of the ratings, and the resulting design. **(a)** The eight preselected brands from which we chose 22 video ads. **(b)** Process of selecting videos based on expert's rating. Suppose that, from brand X, we preselected three videos that were rated by the expert. The videos that had the maximum difference (df1) in score, for which one score was > 50 and the other was < 50, were selected as the final videos from brand X. **(c)** The six selected brands with average scores. For each brand, the video with a higher score was assigned to the HL category (six videos) and the video with a lower score was assigned to the LL category (six videos). **(d)** Average of the second expert's ratings across the six selected videos. **(e)** Average scores (across participants and stimuli) for narrativity level, rated by 124 participants. An independent sample t test showed that the narrativity scores of the

HL videos were significantly higher than those of the LL videos ($p < .001$). **(f)** The 2 × 2 experiment design consisted of the following conditions: HL–PC, HL–VR, LL–PC, and LL–VR.

Design, data collection, and task

We conducted a 2 × 2 full factorial within-subjects study with two levels of narrativity (high level [HL] vs. low level [LL]) and two media (computer screen [PC] vs. virtual reality [VR]). The study included four conditions: six HL video ads presented on a PC (HL–PC), six HL video ads presented in VR (HL–VR), six LL video ads presented on a PC (LL–PC), and six LL video ads presented in VR (LL–VR). See Fig 5-f.

Initially, the EEG device (32-channel, 10-20 system) was placed on each participant's scalp, and the impedance of the active electrodes was set to less than 25 kΩ using a conductive gel, according to the manufacturer's guidelines. The signals were recorded at a sampling rate of 500 Hz using Brain Products software. The HTC Vive Pro VR headset was placed on top of the EEG electrodes, and the impedance of the electrodes was checked again. The task comprised two sessions of approximately 25 min each, separated by a 20 min interval. During one session, participants watched ads on a PC; during the other session, they watched ads in VR. During each session, participants watched all of the video ads and answered a questionnaire after each. A two-second fixation cross was shown before each new video. The videos were displayed first on a PC for half of the participants. The video presentation order was counterbalanced across participants but the same across media. The ten-item questionnaire included two questions of interest for this study: (i) "this commercial really held my attention" and (ii) "this ad draws me in"; both were scored from 0 (strongly disagree) to 100 (strongly agree). We retrieved these questions from the "being hooked" scale[38], and they were

meant to capture consumers' sustained attention to the advertisement [38]. Our engagement metric was calculated based on the combined average scores of these two questions.

Preprocessing

We performed all preprocessing and processing steps using Matlab R2020b (The Math Works, Inc) with in-house codes and tools from the FieldTrip 20210128 (http://fieldtriptoolbox.org) and EEGLAB 2021.0 (https://eeglab.org/) toolboxes. To remove high and low frequency noise, we applied a third-order Butterworth filter with 1–40 Hz cut-off frequencies to the raw data. Next, we detected bad channels using an automated rejection process with voltage threshold $\pm$500 μV and, after confirmation by an expert, rejected them from the channel list. We then interpolated the removed channels using the spherical spline method based on the activity of six surrounding channels in the FieldTrip toolbox. The average number of rejected channels per participant was 1.92 $\pm$ 1.48. One participant was excluded from the analyses due to having more than five bad channels, and two participants were excluded due to missing trials. Subsequently, we segmented the filtered EEG data corresponding to the 12 video ads and concatenated the data. We then excluded the EEG data corresponding to the time when participants were answering the questionnaires. Next, we conducted independent component analysis on the concatenated data to remove remaining noise. We estimated source activity using the second-order blind identification method. We then identified eye-related artifacts and other noisy components, which were confirmed by an expert and removed from the component list. Applying the inverse independent component analysis coefficients to the remaining components, we obtained the denoised data. Finally, we re-referenced the denoised data to the average activity of all electrodes.

Inter-subject correlation

To evaluate whether the evoked responses to the stimuli were shared among participants, we calculated the ISC of neural responses to the video ads by calculating the correlation of EEG activity among participants. Researchers have established that, for an evoked response across participants (or trials) to be reproduceable, it is necessary that each participant (trial) provides a reliable response [11]. In this sense, ISC is similar to traditional methods that capture the reliability of a response by measuring the increase in magnitude of neural activity; the important difference is that, by measuring reliability across participants, we avoid presenting stimuli multiple times to a single participant. In addition, this method is compatible with continuous naturalistic stimuli such as video ads. Therefore, the goal of ISC analysis is to identify correlated EEG components that are maximally shared across participants. By "EEG component," we mean a linear combination of electrodes, which can be considered "virtual sources" [11]. This correlated component analysis method is similar to principal component analysis, for which, instead of the maximum variance within a dataset, the maximum correlation among datasets is considered. Like other component extraction methods, this method identifies components by solving a eigenvalue problem [39,40]. Below, we explain our ISC calculation procedure for multiple stimuli (provided in [11,22]) .

To construct the input data, we combined data from participants in all four conditions (six videos for each condition; see Fig 6-a for one stimulus) and obtained 24 three-dimensional EEG matrices (channel × data samples × participants). For each of these 24 matrices, we separately calculated the between-subject cross-covariance using Eq1:

$$R_b = \frac{1}{N(N-1)} \sum_{k=1}^{N} \sum_{l=1, l \neq k}^{N} R_{kl} \qquad \text{Eq1}$$

where R_{kl} is as follows:

$$R_{kl} = \sum_{t} (x_k(t) - \bar{x}_k)(x_l(t) - \bar{x}_l)^T$$

R_{kl} indicates cross-covariance among all electrodes of subject *k* with all electrodes of subject l. The matrix $x_k(t)$ contains 32 electrode activities (preprocessed data) of subject *k* measured in time, and $\bar{x}_k$ is the average of $x_k(t)$ over time. Additionally, we separately calculated the within-subject cross-covariance for each of the abovementioned matrices using Eq2:

$$R_w = \frac{1}{N} \sum_{k=1}^{N} R_{kk} \qquad \text{Eq2}$$

R_{kk} is calculated in an identical manner to R_{kl}, except it considers only the electrode activity of subject *k*. We then summed the calculated R_b and R_w of all 24 matrices to obtain the pooled within- and between- subject cross-covariances, representing data on all stimuli.

Then, to improve the robustness of our analysis against outliers, we used the shrinkage regularization method to regularize the within subject-correlation matrix[41] (Eq3):

$$R_w \leftarrow (1 - \gamma) R_w + \gamma \bar{\lambda} I \qquad \text{Eq3}$$

where $\bar{\lambda}$ is the average of eigenvectors of R_w, with γ equal to 0.5.

Next, by solving the eigenvalue problem for $R_w^{-1} R_b$ (Eq4), we obtained eigenvectors v_i. Such eigenvectors are projections indicating the maximum correlation among participants, from strongest to weakest, provided by eigenvalues λ_i.

$$(R_w^{-1} R_b) v_i = \lambda_i v_i \qquad \text{Eq4}$$

After this step, we obtained the projections that are maximally correlated among all participants considering all stimuli (Fig 6-b). Next, to measure the reliability of individual participants' EEG responses, we calculated the correlation of projected data for each subject with projected data for the group separately for each stimulus (Fig 6-c). This metric shows how similar the brain activity of a single subject was to that of all other participants. Next, we calculated the ISC using the correlation of such projections for each stimulus (video ads) and component, averaged across all possible combinations of participants. Thus, for each stimulus and each subject, we obtained a component matrix. By summing the first two strongest components, we obtained the ISC for each subject, as follows (Eq5):

$$ISC_k = \sum_{i=1}^{2} r_{ki} \qquad \text{Eq5}$$

where

$$r_{ki} = \frac{1}{N(N-1)} \sum_{l=1, l\neq k}^{N} \frac{\sum_t y_{ik}(t) y_{il}(t)}{\sqrt{\sum_t y_{ik}^2(t)} \sqrt{\sum_t y_{il}^2(t)}}$$

r_{ki} is the Pearson correlation coefficient averaged across all pairs of participants applied to component projection y_{ik}, which is defined as follows (Eq6):

$$y_{ik}(t) = v_i^T (x_{ik}(t) - x_k) \qquad \text{Eq6}$$

Next, we summed the first two strongest components to represent the ISC of each subject for each stimulus (i =1:2) and ignored the weaker components because they were below chance level in one condition. Finally, we averaged the calculated ISC over the six stimuli of each condition to obtain the ISC of all participants in each of the four conditions. To illustrate the scalp activity of each component, we used the

corresponding forward model following previous studies [42,43] (Fig 6-d). In Fig 2-a, the bars illustrate the sum of the two first components.

To determine chance-level ISC, we first built up phase-randomized EEG data using a method that randomizes EEG signal phases in the frequency domain [11,44]. Using this method, we obtained new time series where the temporal alterations were not necessarily aligned with the original signals and therefore not correlated across participants. We then implemented the aforementioned steps on the randomized data, identically: for each condition, we computed within- and between-subject cross-covariances, projected the data on eigenvectors, calculated ISC, and averaged them over stimuli. We generated 100 sets of such randomized data and continued the process as described above. Finally, we obtained the chance-level ISC by averaging the values computed in these 100 iterations. The results attributed to chance level are shown by the gray bars in Fig 2-b.

Fig 6 about here

Fig 6: The ISC calculation and component activity estimation procedure. We repeated this procedure separately for each stimulus. **(a)** We first concatenated the EEG activity of all participants together. **(b: upper line)** For each subject, we calculated the $\boldsymbol{r_b}$ and $\boldsymbol{r_w}$ (using Eq1 and Eq2, respectively) based on the cross-covariance matrices. By solving the eigenvalue problem (Eq3), we computed eigenvectors of each subject. The red arrows represent each subject's eigenvector ($\boldsymbol{v'}$). Therefore, we obtained 29 $\boldsymbol{v'}$, corresponding to the number of participants. Note that each $\boldsymbol{v'_i}$ is a 32-dimensional matrix corresponding to the number of electrodes (for simplicity, the vectors are shown

in two dimensions). **(b: lower line)** We summed all the $\boldsymbol{r_b}$ across all participants to calculate $\boldsymbol{R_b}$ and did the same for $\boldsymbol{r_w}$ to calculate $\boldsymbol{R_w}$. Then, by solving the eigenvalue problem, we obtained $\boldsymbol{v_i}$, which represents maximal correlation across participants. Note that $\boldsymbol{v}$ is a 32-dimensional matrix corresponding to the number of electrodes (for simplicity, the vectors are shown in two dimensions). **(c)** We calculated the similarity between the representative vector of all participants ($\boldsymbol{v}$) and each of the vectors corresponding to each subject ($\boldsymbol{v'_1}$ ... $\boldsymbol{v'_{29}}$) to obtain ISC. **(d)** Using the calculated eigenvectors and forward model, we calculated the scalp activity of each component, which are linear combinations of electrode activity (32 components were calculated with the order from strongest to weakest). In sum, the activity of first two strongest components is displayed here.

ISC-HL

We also tested whether we could assign a single participant to the HL or LL group based on participant's neural brain activity. In other words, we tested whether we could predict the level of narrativity the participant was exposed based on EEG activity when participants attended to HL and LL videos. To do so, we calculated the projections v_i using only the data from the HL group. We then calculated the ISC of each participant (from both the HL and LL groups) based on the projections calculated only from the HL group and called the result "ISC-HL." Therefore, in each group, each participant had an ISC-HL value showing how similar their neural activity was to the activity of all participants in the HL group. The difference between ISC-HL and ISC is that, while calculating ISC, we explored how similar the activity of a participant was to their own group (whether HL or LL). When calculating ISC-HL, however, we calculated how similar the activity of participants in both groups (HL and LL) was to the activity of

only the participants in the HL group. We hypothesized that the neural activity of a single subject from the HL group would be more similar to the overall activity of all participants in HL group (i.e., higher ISC-HL) than to the neural activity of a single subject from the LL group. To avoid bias in this procedure, we excluded the test subject from the HL group while computing the projections v_i and then calculated the ISC-HL for the corresponding subject. As described in the previous section, we summed the two first stronger components of ISC-HL and also averaged them across six stimuli for each condition. Additionally, we assessed classification performance for both groups (HL vs. LL) using the AUC characteristics of a fitted support vector machine model with 10-fold cross-validation on the data. To determine the chance-level AUC, we randomly shuffled the labels 1,000 times. In each iteration, AUC was calculated through an identical process, starting from extracting the correlated components of HL-labeled group and following all subsequent steps described above. We then determined the significance level by comparing the actual AUC value computed from the correct labels to the average of 1,000 randomized AUC values. We repeated all the aforementioned steps separately for the PC and VR conditions.

Statistical analysis

To test statistical differences across narrativity levels, we conducted a two-way ANOVA. Based on the study design, our independent variables were narrativity level (HL vs. LL) and medium (PC vs. VR). The dependent variables were the scores from the two questionnaires (i.e., attention and immersion) and calculated neural reliability (as expressed by ISC). The output of such a comparison includes the main effects of "narrativity level" and "medium" and the interaction effect of "narrativity level × medium." The same procedure and dependent variables, but with narrativity level (HL vs. LL) and viewing order (first vs. second) as independent variables, were used to test

statistical differences across viewing order. In the present study, we focused on the effect of narrativity level and viewing order on brain activity patterns by comparing the calculated features in HL vs. LL conditions, and first vs. second viewing. Therefore, we do not report results for the main effect of medium. Nevertheless, we included medium type as a factor in the statistical analyses to control for its effect. Additionally, we performed post-hoc analysis when we identified interaction effects but also when we judged it appropriate to report simple effects. We corrected the p-values of the post-hoc tests using the Bonferroni method. Finally, to avoid multiple comparison error of multiple statistical tests (such as comparing HL and LL within two extracted components; see results), we corrected the p-values based on the number of statistical test repetition errors (FDR correction) using the Benjamini-Hochberg method [45].

Acknowledgements: We are grateful to Tirdad Seifi Ala for helping with technical aspects of the data analysis and for revising an early version of the manuscript, and to Thomas Anthony Pedersen for providing technical support and help during the early stages of the study.

References

1. Chun MM, Golomb JD, Turk-Browne NB. A Taxonomy of External and Internal Attention. Annu Rev Psychol [Internet]. 2011 Jan 10;62(1):73–101. Available from: https://www.annualreviews.org/doi/10.1146/annurev.psych.093008.100427

2. Dmochowski JP, Sajda P, Dias J, Parra LC. Correlated Components of Ongoing EEG Point to Emotionally Laden Attention – A Possible Marker of Engagement?

Front Hum Neurosci [Internet]. 2012;6(MAY 2012):1–9. Available from: http://journal.frontiersin.org/article/10.3389/fnhum.2012.00112/abstract

3. Busselle R, Bilandzic H. Measuring Narrative Engagement. Media Psychol [Internet]. 2009 Nov 23;12(4):321–47. Available from: http://www.tandfonline.com/doi/abs/10.1080/15213260903287259

4. Vuilleumier P. How brains beware: neural mechanisms of emotional attention. Trends Cogn Sci [Internet]. 2005 Dec;9(12):585–94. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1364661305003025

5. Hogan PC. Affective Narratology: The Emotional Structure of Stories. Frontiers of Narrative Series. University of Nebraska Press; 2011. 304 p.

6. Houghton DM. Story elements, narrative transportation, and schema incongruity: a framework for enhancing brand storytelling effectiveness. J Strateg Mark [Internet]. 2021 Aug 19;1–16. Available from: https://www.tandfonline.com/doi/full/10.1080/0965254X.2021.1916570

7. Green MC, Brock TC. The role of transportation in the persuasiveness of public narratives. J Pers Soc Psychol. 2000 Nov;79(5):701–21.

8. Tchernev JM, Collier J, Wang Z. There and Back Again? Exploring the Real-Time Cognitive Journey of Narrative Transportation. Communic Res [Internet]. 2021 May 27;009365022110185. Available from: http://journals.sagepub.com/doi/10.1177/00936502211018577

9. Richardson DC, Griffin NK, Zaki L, Stephenson A, Yan J, Curry T, et al. Engagement in video and audio narratives: contrasting self-report and physiological measures. Sci Rep [Internet]. 2020 Dec 9;10(1):11298. Available

from: https://doi.org/10.1038/s41598-020-68253-2

10. Song H, Finn ES, Rosenberg MD. Neural signatures of attentional engagement during narratives and its consequences for event memory. Proc Natl Acad Sci [Internet]. 2021 Aug 17;118(33):e2021905118. Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.2021905118

11. Ki JJ, Kelly SP, Parra LC. Attention Strongly Modulates Reliability of Neural Responses to Naturalistic Narrative Stimuli. J Neurosci [Internet]. 2016 Mar 9;36(10):3092–101. Available from: http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.2942-15.2016

12. Sonkusare S, Breakspear M, Guo C. Naturalistic Stimuli in Neuroscience: Critically Acclaimed. Trends Cogn Sci [Internet]. 2019 Aug;23(8):699–714. Available from: https://doi.org/10.1016/j.tics.2019.05.004

13. Dini H, Simonetti A, Bigne E, Bruni LE. EEG theta and N400 responses to congruent versus incongruent brand logos. Sci Rep [Internet]. 2022 Dec 16;12(1):4490. Available from: https://doi.org/10.1038/s41598-022-08363-1

14. Escalas JE. ADVERTISING NARRATIVES: What are they and how do they work? In: Stern B, editor. Representing Consumers [Internet]. Routledge; 1998. p. 267–89. Available from: https://www.taylorfrancis.com/books/9781134669875/chapters/10.4324/9780203380260-19

15. Coker KK, Flight RL, Baima DM. Video storytelling ads vs argumentative ads: how hooking viewers enhances consumer engagement. J Res Interact Mark [Internet]. 2021 Oct 15;15(4):607–22. Available from: https://www.emerald.com/insight/content/doi/10.1108/JRIM-05-2020-

0115/full/html

16. Wang RWY, Chang YC, Chuang SW. EEG Spectral Dynamics of Video Commercials: Impact of the Narrative on the Branding Product Preference. Sci Rep [Internet]. 2016 Dec 7;6(1):36487. Available from: http://www.nature.com/articles/srep36487

17. Nastase SA, Gazzola V, Hasson U, Keysers C. Measuring shared responses across subjects using intersubject correlation. Soc Cogn Affect Neurosci [Internet]. 2019 May 16;14(6):669–87. Available from: https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsz037/5489905

18. Redcay E, Moraczewski D. Social cognition in context: A naturalistic imaging approach. Neuroimage [Internet]. 2020 Aug;216(November 2019):116392. Available from: https://doi.org/10.1016/j.neuroimage.2019.116392

19. Imhof MA, Schmälzle R, Renner B, Schupp HT. Strong health messages increase audience brain coupling. Neuroimage [Internet]. 2020 Aug;216(August 2019):116527. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1053811920300148

20. Petroni A, Cohen SS, Ai L, Langer N, Henin S, Vanderwal T, et al. The Variability of Neural Responses to Naturalistic Videos Change with Age and Sex. eneuro [Internet]. 2018 Jan;5(1):ENEURO.0244-17.2017. Available from: https://www.eneuro.org/lookup/doi/10.1523/ENEURO.0244-17.2017

21. Hasson U, Furman O, Clark D, Dudai Y, Davachi L. Enhanced Intersubject Correlations during Movie Viewing Correlate with Successful Episodic Encoding. Neuron [Internet]. 2008 Feb;57(3):452–62. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0896627307010082

22. Cohen SS, Parra LC. Memorable Audiovisual Narratives Synchronize Sensory and Supramodal Neural Responses. eneuro [Internet]. 2016 Nov 10;3(6):ENEURO.0203-16.2016. Available from: https://pubmed.ncbi.nlm.nih.gov/27844062

23. Simony E, Honey CJ, Chen J, Lositsky O, Yeshurun Y, Wiesel A, et al. Dynamic reconfiguration of the default mode network during narrative comprehension. Nat Commun [Internet]. 2016 Nov 18;7(1):12141. Available from: http://www.nature.com/articles/ncomms12141

24. Nummenmaa L, Glerean E, Viinikainen M, Jaaskelainen IP, Hari R, Sams M. Emotions promote social interaction by synchronizing brain activity across individuals. Proc Natl Acad Sci [Internet]. 2012 Jun 12;109(24):9599–604. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1206095109

25. Dmochowski JP, Bezdek MA, Abelson BP, Johnson JS, Schumacher EH, Parra LC. Audience preferences are predicted by temporal reliability of neural processing. Nat Commun [Internet]. 2014 Dec 29;5(1):4567. Available from: http://www.nature.com/articles/ncomms5567

26. Regev M, Simony E, Lee K, Tan KM, Chen J, Hasson U. Propagation of Information Along the Cortical Hierarchy as a Function of Attention While Reading and Listening to Stories. Cereb Cortex [Internet]. 2019 Sep 13;29(10):4017–34. Available from: https://academic.oup.com/cercor/article/29/10/4017/5158241

27. Song H, Park B yong, Park H, Shim WM. Cognitive and Neural State Dynamics of Narrative Comprehension. J Neurosci [Internet]. 2021 Oct 27;41(43):8972–90. Available from:

https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.0037-21.2021

28. Poulsen AT, Kamronn S, Dmochowski J, Parra LC, Hansen LK. EEG in the classroom: Synchronised neural recordings during video presentation. Sci Rep [Internet]. 2017 Apr 7;7(1):43916. Available from: http://www.nature.com/articles/srep43916

29. Cohen SS, Henin S, Parra LC. Engaging narratives evoke similar neural activity and lead to similar time perception. Sci Rep [Internet]. 2017 Dec 4;7(1):4578. Available from: http://www.nature.com/articles/s41598-017-04402-4

30. Ryan ML. Toward a definition of narrative. In: Herman D, editor. The Cambridge Companion to Narrative [Internet]. Cambridge University Press; 2007. p. 22–36. Available from: https://www.cambridge.org/core/product/identifier/CBO9781139001533A006/type/book_part

31. Lim S, Yeo M, Yoon G. Comparison between Concentration and Immersion Based on EEG Analysis. Sensors [Internet]. 2019 Apr 8;19(7):1669. Available from: https://www.mdpi.com/1424-8220/19/7/1669

32. Chang WT, Jääskeläinen IP, Belliveau JW, Huang S, Hung AY, Rossi S, et al. Combined MEG and EEG show reliable patterns of electromagnetic brain activity during natural viewing. Neuroimage [Internet]. 2015 Jul;114:49–56. Available from: http://dx.doi.org/10.1016/j.neuroimage.2015.03.066

33. Abbott HP. The Cambridge Introduction to Narrative [Internet]. 2nd ed. Cambridge: Cambridge University Press; 2008. Available from: http://ebooks.cambridge.org/ref/id/CBO9780511816932

34. Oatley K. Such stuff as dreams: The psychology of fiction. John Wiley & Sons; 2011.

35. Bruni LE, Dini H, Simonetti A. Narrative Cognition in Mixed Reality Systems: Towards an Empirical Framework. In: Chen JYC, Fragomeni G, editors. Virtual, Augmented and Mixed Reality [Internet]. Cham: Springer International Publishing; 2021. p. 3–17. Available from: https://link.springer.com/10.1007/978-3-030-77599-5_1

36. Kim E (Anna), Ratneshwar S, Thorson E. Why Narrative Ads Work: An Integrated Process Explanation. J Advert [Internet]. 2017 Apr 3;46(2):283–96. Available from: https://www.tandfonline.com/doi/full/10.1080/00913367.2016.1268984

37. Jajdelska E, Anderson M, Butler C, Fabb N, Finnigan E, Garwood I, et al. Picture This: A Review of Research Relating to Narrative Processing by Moving Image Versus Language. Front Psychol [Internet]. 2019 Jun 26;10(JUN). Available from: https://www.frontiersin.org/article/10.3389/fpsyg.2019.01161/full

38. Escalas JE, Moore MC, Britton JE. Fishing for Feelings? Hooking Viewers Helps! J Consum Psychol. 2004;14(1–2):105–14.

39. de Cheveigné A, Parra LC. Joint decorrelation, a versatile tool for multichannel data analysis. Neuroimage. 2014;98:487–505.

40. Parra L, Sajda P. Blind source separation via generalized eigenvalue decomposition. J Mach Learn Res. 2003;4:1261–9.

41. Blankertz B, Lemm S, Treder M, Haufe S, Müller KR. Single-trial analysis and classification of ERP components—a tutorial. Neuroimage. 2011;56(2):814–25.

42. Haufe S, Meinecke F, Görgen K, Dähne S, Haynes JD, Blankertz B, et al. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage. 2014;87:96–110.

43. Parra LC, Spence CD, Gerson AD, Sajda P. Recipes for the linear analysis of EEG. Neuroimage. 2005;28(2):326–41.

44. Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD. Testing for nonlinearity in time series: the method of surrogate data. Phys D Nonlinear Phenom. 1992;58(1–4):77–94.

45. Gerstung M, Papaemmanuil E, Campbell PJ. Subclonal variant calling with multiple samples and prior knowledge. Bioinformatics. 2014;30(9):1198–204.

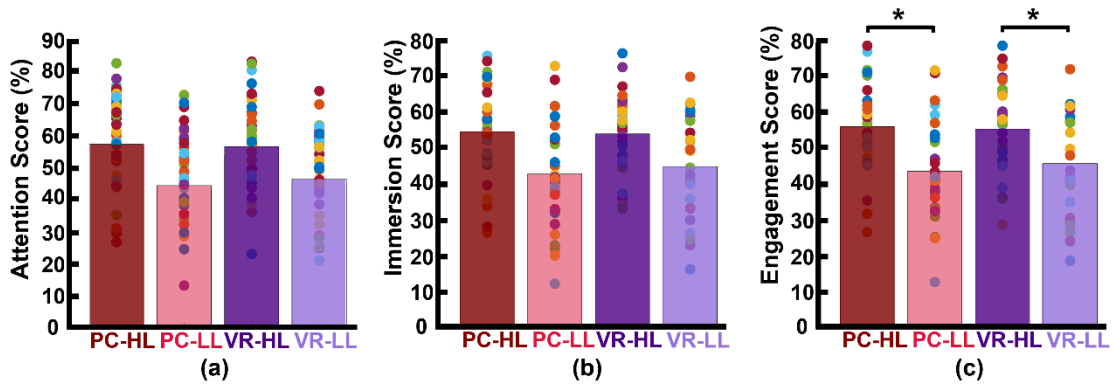

Figure 1: self-reported scores rated by participants. Each bar is the average of the scores over participants and over corresponding video ads. The colored dots represent the score for each participant averaged across corresponding videos. **(a)** Self-reported scores for attention **(b)** Self-reported scores for Immersion. **(c)** The scores for our defined engagement metric which is average of attention and immersion scores. It show that the HL scores are significantly higher than LL in PC condition ($p < .001$) and VR condition ($p < .001$)

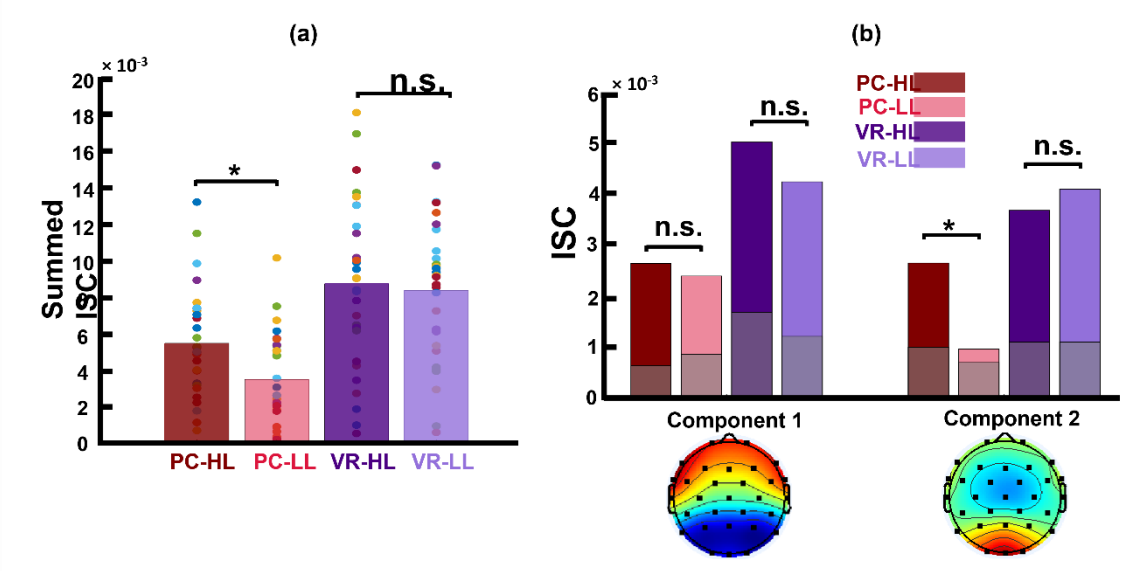

Figure 2: (a) Summed-ISC over first two strongest components in four conditions. It shows that in PC condition, the ISC of HL video ads is significantly higher that LL video ads ($p = .029$), and in the VR condition the ISC of HL is still higher than LL but not significantly ($p = .638$). **(b-top)** Calculated ISC for each component separately, and within each condition. It shows that in component 1, the difference between HL vs LL video ads is not significant neither for PC nor VR condition. However, in component 2, the ISC of HL video ads is significantly higher that LL video ads ($p = .002$) in PC condition, but not VR condition ($p = .313$). The gray bars shows the average of calculated chance level for each condition **(b-bottom)** Scalp activity of first two strongest components. The activity in component 1 is over frontal occipital area while in component 2, the activity is concentrated on occipital area. In both figures, the colored dots represent the calculated ISC values of each participant.

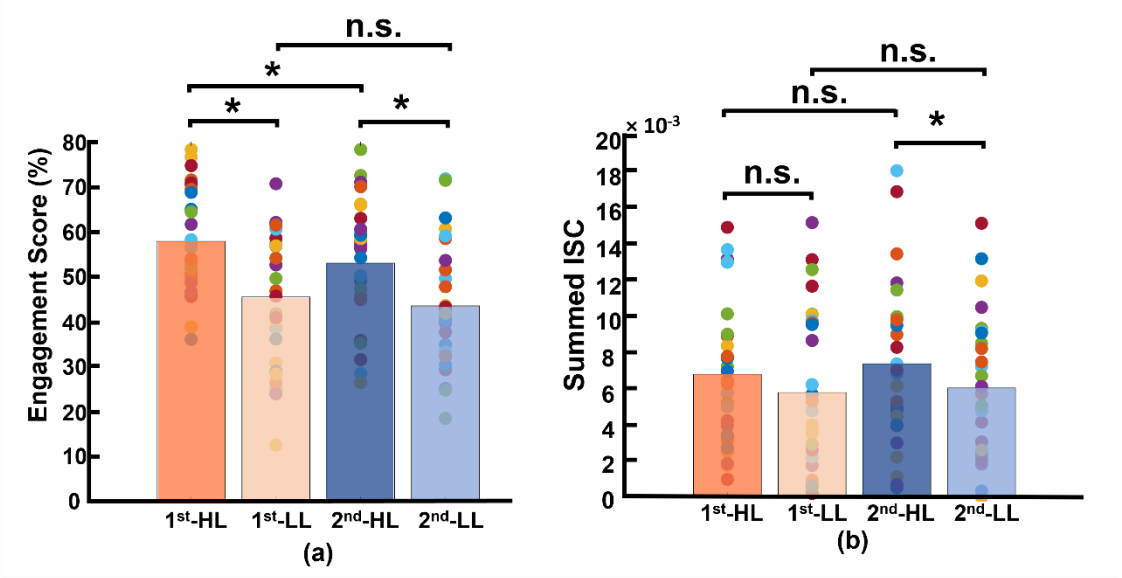

Figure 3: Results of viewing order (1st and 2nd) considering narrativity level (HL and LL). **(a)** Results of engagement scores, where in the HL condition, engagement scores of 1st viewing is significantly higher than 2nd viewing. Moreover, HL has significantly higher engagement scores than LL in both 1st and 2nd viewing. (**b)** Results for calculated ISC of viewing order, where the only significant difference occurs in comparison between HL vs LL of 2nd viewing.

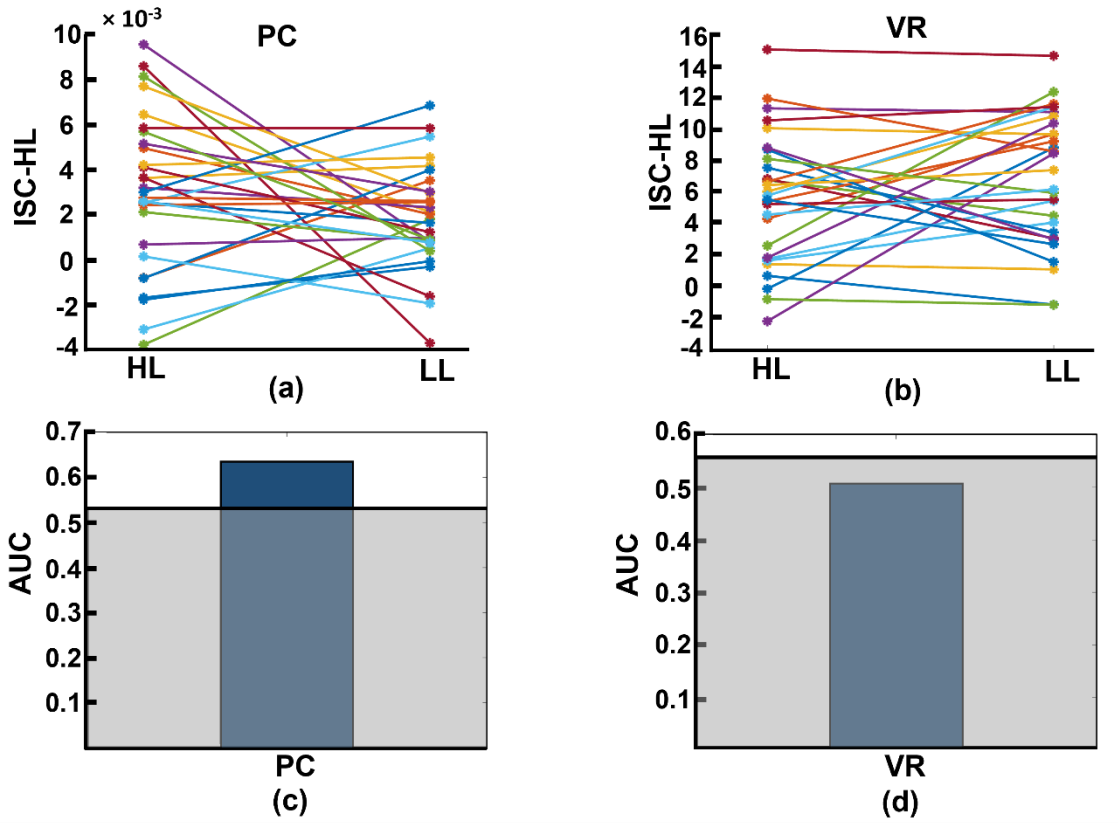

Figure 4: (a) and (b) Results for calculated ISC-HL for each participant in PC and VR, respectively. The colored line and markers shows the ISC-HL of one participant from HL and LL condition. **(a)** It shows that 76.6% of participants had higher ISC-HL when they attended to HL video ads than attending to LL video ads. This shows that most of the participants who were exposed to HL video ads showed higher similarity to the trained model (which is based on HL) than the ones who are exposed to LL video ads. (**b)** It relates to the same evaluation as (a), but in VR condition, where 50% of participants had higher ISC-HL. **(c) and (d)** AUC of classification performance for predicting the exposure to HL and LL video ads in PC and VR conditions, respectively. The gray area shows the chance-level performance. **(c)** It shows that the AUC of classification based on participant's neural activity in PC condition is higher than chance level, while **(d)** shows the AUC of classification based on participant's neural activity in VR condition is below chance.

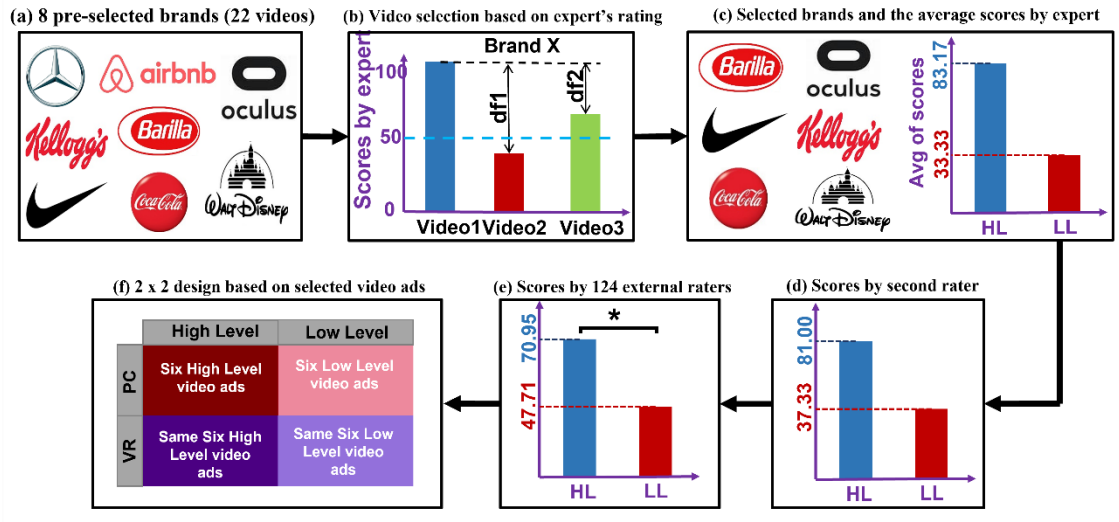

Figure 5: Procedure of video ads selection, dividing them into HL and LL categories, results of the ratings, and the resulted design. **(a)** The eight pre-selected brands from which we chose 22 video ads. **(b)** Process of selecting videos based on expert's rating. Suppose that from brand X, we have pre-selected three videos (e.g., Video 1 to Video 3) which are rated by the expert. The videos which had the maximum difference (df1), and one above 50 and the other below 50 score is selected as final videos from such a brand. **(c)** The six selected brands with average scores. The video with higher score is assigned to HL (six videos) and the other with lower score is assigned to LL (six videos) category. **(d)** Average of the scores rated by the second expert across six selected videos. **(e)** Average (across subjects and stimuli) scores for narrativity level by 124 subjects. An independent sample t-test showed the scores of HL are significantly higher than LL ($p < .001$). **(f)** 2x2 experiment design consisted of HL-PC, HL-VR, LL-PC, and LL-VR.

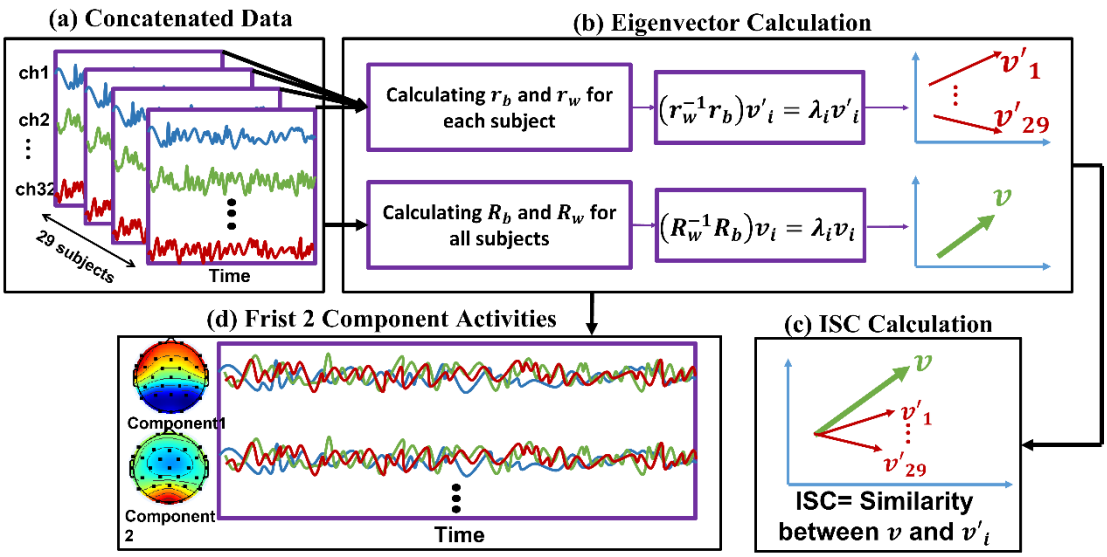

Figure 6: The procedure of ISC calculation and component activity estimation. This procedure has been repeated for each stimuli separately. **(a)** We first concatenated the EEG activity of all subjects together. **(b-upper line)** For each subject, we calculated the r_b and r_w (Eq1 and Eq2, respectively) based on the cross-covariance matrices. By solving the eigenvalue problem (Eq3), we computed eigenvectors of each subject. The red arrows exemplify each subject's eigenvector (v'). Therefore, we obtained 29 v' corresponding to the number of subjects. Note that each v'_i is a 32-dimensional matrix corresponding to the number of electrodes (for simplicity we showed them in 2 dimensions). **(b-bottom line)** We summed up all the r_b across all subjects to reach R_b and the same for r_w to reach R_w. Then by solving the eigenvalue problem, we obtained v_i, which is a representative of maximal correlation across subjects. Note that v is a 32-dimensional matrix corresponding to the number of electrodes (for simplicity we showed them in 2 dimensions). **(c)** By calculating the similarity between the representative vector of all subjects (v) and each of the vectors corresponding to each subject ($v'_1 \ldots v'_{29}$) to obtain ISC. **(d)**Using the calculated eigen vectors and forward model, we calculated the activity of each component, which are linear combinations of electrode's activity (32 components ordered from strongest to weakest were calculated). For summarization, the activity of fist two strong components is visualized here.