Visual Recognition of the Memento10k dataset: A P300 component ERP study

Rachelle Hamelink[1]

[1] *Radboud University, Netherlands*

Abstract

In modern and more specifically social media, most advertisements nowadays consist of very short, few second videos. This leads to an increase of interest in recognizability and rememberability of visual media. The present study event-related potentials (ERP's) across 1000 trials of the Memento10k dataset to explore neurophysiological differences between videos that are later remembered or not when shown again to the same subject. The posterior brain region was analyzed across three channels and the right temporal cortex was analyzed in one channel. A significant difference, measured as $p < .05$, in amplitude was found in the $340 - 408$ ms window after onset between remembered and not remembered videos as well as around the 476 ms timepoint across all trials, channels and participants in the visual cortex. In the right temporal cortex, a significant difference, measured as $p < .05$, in amplitude was observed in the $306 - 816$ ms window. These results suggest, in line with previous literature, that a stronger P300 component can be found for remembered videos than not remembered videos in the right temporal lobe. In the visual cortex, an opposite effect was found, as a higher positivity was observed for not remembered videos.

1 INTRODUCTION

In an increasingly more fast-paced and social media oriented lifestyle, the need for companies to investigate how easily media can become recognizable is rising. More often than not nowadays, advertisements can only consist of several second videos to grab the attention of the viewer. As smart phones and apps containing these short advertisements have become more and more accessible to a wider audience, companies find themselves competing over attention of the viewers at an increasing rate. This creates a need to understand why certain short videos or still phrames have increased recognizability over others. Neurophysiological research can provide a foundation to investigate which underlying neural processes cause some multimedia items to be memorized over others, even after just one short viewing. Memorability is defined in the context of this paper as recognition upon second viewing of the same stimulus. Memorability of visual content can either consciously of subliminally influence people's decision making in cases such as purchasing a certain brand over another unknown option. Therefore, an interest in research investigating the properties of visual data that can increase recognizability is growing [1]. This present study investigates whether a difference can be observed in event-related potentials (ERP's) within the 300 millisecond – 1000 millisecond time window after the onset of watching a video between videos that will later be remembered and not remembered via binary recognition. This paper is written as a part of the MediaEval 2022 Predicting Video Memorability task [2] , using the Memento10k consisting of three-second videos [3]. All analyzed data in this study was provided by the MediaEval 2022 Predicting Video Memorability Task [2].

2 RELATED WORK

An example of feature research has shown that increased creativity used in creating advertisement can be an important factor in later recall and recognition of shown advertisements [4]. Before important feature information causing these peaks in recognizability can be extracted, however, it is important to explore the difference between remembered and not remembered videos or the extent to which there is one. This specific type of visual explicit memory was explained to be declarative memory by Osipova et al. [5]. Both formation [5] and retrieval [6] of declarative memory have been studied using brain activity during encoding of items and during recognition of these items after a short time period.

Because of this research, in the neuroscience field it has become well known that neural oscillations occur in specific brain areas when a declarative memory formation is successfully formed [4]. Based on evidence from functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), studies showed that the anterior prefrontal cortex, parietal cortex and the medial-frontal areas were activated during an old/new paradigm, which shows the participants a set of images and then after a short time window shows these images again mixed with new images and if a recognition memory was successfully formed, a P300 component would be seen in these mentioned brain areas [7] [8] [9].

In addition, Osipova et al. [5] showed using magnetoencephalogram data that during encoding of images in an old/new paradigm, stronger gamma band activity was shown over the posterior brain regions for later remembered items than forgotten items and a systematic difference of theta band activity was found over right temporal regions, all in the 0.3ms – 1000 ms window, or the so-called P300 component[5]. A P300 component indicates the observation of a larger positivity for the manipulated variable in the 300+ ms time window from onset of the stimulus. Relating to this study, the would indicate a larger positivity from the 300ms mark after onset for remembered videos than not-remembered videos, in both the posterior brain region and the right temporal lobe.

The posterior brain regions are correlated with the visual cortex and are in close proximity to the hippocampus, respectively important for the processing of visual imput and for initial consolidation into memory systems. The right temporal lobe of the brain is involved in the processes of learning and remembering non-verbal informatin, such as music and videos. Polich [10] explains the P300 component by amplitude and latency.

The amplitude is defined by the difference between the mean baseline voltage and the largest positive peak of the ERP wavefrom within a time window, in this study 300-1000ms after stimulus onset based on [5]. Latency is defined as the time in ms from stimulus onset to the point of maximum positive amplitude within a time window, 300-1000ms in this study based on [5]. Analyzing ERP amplitudes for memory is an interesting method for declarative memory, as this method has shown that at formation differences can already be observed that can later predict memorability [5]. Classic declarative memory studies would have to rely on a consolidation period of time and possible sleep factors in order to measure successful memorization of the stimulus input.

This study aims to investigate whether this P300 effect can be observed in videos of the Memento10k dataset. The aim is to research whether a significant P300 difference effect can be found between remembered and not remembered videos in the encoding phase of the old/new paradigm in this experiment. This data is then used to analyze whether a P300 effect in the encoding phase of an old/new paradigm can predict successful video memorability. ERP's rely on the onset of a stimulus to observe a difference in brain function. EEG recordings were collected for the first second of the videos [2]. Therefore, as it can be argued that a one second video does not differ severely from an still image, the main hypothesis in this paper is that a between the 300-1000 ms time window in the data, a significant P300 component can be observed.

3 APPROACH

3.2 Participants, materials and procedure

The participants, materials and procedure of the Predicting Video Memorability task data collection can be found in the overview paper [2]. For this analysis, the posterior brain region or visual cortex was analyzed by means of the EEG cap channels Oz, O1 and O2. In addition, the right temporal lobe was analyzed by means of the EEG cap channel P8. Both these region were analyzed based on previous results using MEG by Osipova et al. [5]. The posterior brain regions correspond with the visual cortex, which is mostly activated when processing multimedia input and has close proximity to the hippocampus, which is a key are for memorization. The right temporal lobe is considered involved in learning and remembering of non-verbal stimuli, such as music and videos. Both these areas were analyzed separately to observe a difference in remembered versus non-remembered videos. Participants had to indicate on second viewing whether remembered the presented video or not. Only trials that included a memorability score for that video were included in final analysis.

3.3 Data cleaning and statistical analysis

Preprocessing steps were undertaken to create the final dataset [2]. In addition to the preprocessing steps described in the overview paper of the task, outlier amplitudes across all participants, the four included channels and all trials were removed based on 1.5 standard devation below the first quartile of the data and above the third quartile of the amplitude data. A linear mixed effect model analysis with memorability score, channel and timepoint as fixed effects and subject as random effect was used to analyze the dataset in the visual cortex. An additional linear effect model analysis with memorability score and timepoint as fixed effects and subject as random effect was performed to analyze the dataset in the right temporal cortex. All data was processed and plotting and analyzing of the data was performed using RStudio [12].
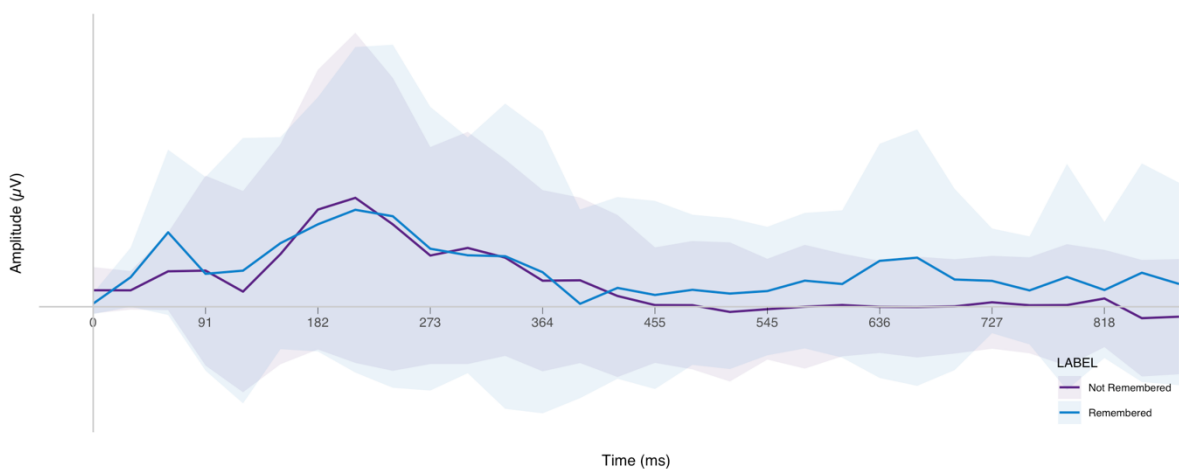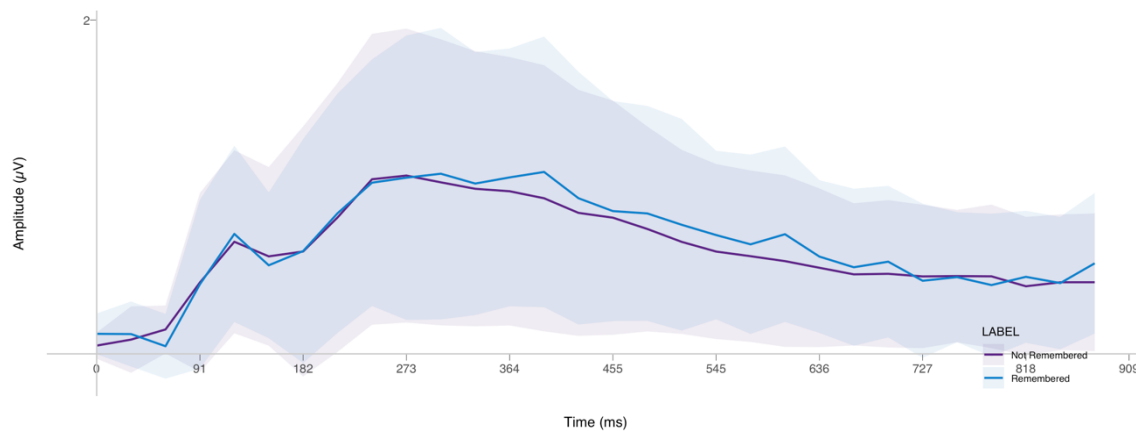
4 RESULTS AND ANALYSIS

Figure 1: **Visual representation of the ERP data of remembered versus not remembered videos in the Memento10k dataset in the visual cortex. The ribbons displayed in the graph represent the confidence intervals of the ERP amplitudes, measures at 95% upper and lower bound.**

A linear mixed effects model with fixed effects Time, Channel and Memorability Score was performed on the Amplitude ERP data, with an added random factor of Subject. Table 1 represents the significant Time and Label interactions discovered in the data, indicating that the time window of ~ 340 ms - ~ 408 ms and the timepoint at ~ 476 ms showed a significant difference in ERPs between remembered and not remembered videos across all trials and participants. These differences in ERP amplitudes can be observed in Figure 1. The largest differences can be observed at timepoint 10 (374 ms) and timepoint 13 (475 ms) in Figure 1. All other timepoints showed no significant difference between remembered and not remembered videos (all p's > 0.05).

Table 1 : **Table containing significant Memorability Score (Remembered vs Not remembered videos) x Timepoints (in ms) interactions in the visual cortex (Channels Oz, O1 and O2)**

	Beta (Fixed effect estimate)	Standard Error	t	p
Memorability Score * Timepoint 9 (~ 340 ms)	-0.0835	0.0425	-1.963	0.050
Memorability Score * Timepoint 10 (~ 374 ms)	-0.0937	0.0426	-2.198	0.023
Memorability Score * Timepoint 11 (~ 408 ms)	-0.0895	0.0428	-2.093	0.036
Memorability Score x Timepoint 13 (~ 476 ms)	-0.1144	0.0428	-2.670	0.008

Figure 2: **Visual representation of the ERP data of remembered versus not remembered videos in the Memento10k dataset in the right temporal lobe. The ribbons displayed in the graph represent the confidence intervals of the ERP amplitudes, measures at 95% upper and lower bound.**

A linear mixed effects model with fixed effects Time and Memorability Score was performed on the Amplitude ERP data, with an added random factor of Subject. Table 2 represents the significant Time and Label interactions discovered in the data, indicating that the time window of ~ 306 ms - ~ 816 ms showed a significant difference in ERPs between remembered and not remembered videos across all trials and participants. These differences in ERP amplitudes can be observed in Figure 2. The largest differences can be observed at 13 (475 ms) in Figure 2. All other timepoints showed no significant difference between remembered and not remembered videos (all p's > 0.05).

Table 2 : **Table containing significant Memorability Score (Remembered vs Not remembered videos) x Timepoints (in ms) interactions in the right temporal lobe (Channel P8)**

	Beta (Fixed effect estimate)	Standard Error	t	p
Memorability Score * Timepoint 8 (~ 306 ms)	0.3778	0.0993	3.804	<.001
Memorability Score * Timepoint 9 (~ 340 ms)	0.4539	0.0994	4.565	<.001
Memorability Score * Timepoint 10 (~ 374 ms)	0.5698	0.0995	5.728	<.001
Memorability Score * Timepoint 11 (~ 408 ms)	0.5614	0.0997	5.629	<.001
Memorability Score * Timepoint 12 (~ 442 ms)	0.6010	0.0996	6.036	<.001
Memorability Score x Timepoint 13 (~ 476 ms)	0.6615	0.0999	6.624	<.001
Memorability Score x Timepoint 14 (~ 510 ms)	0.5458	0.0998	5.470	<.001
Memorability Score x Timepoint 15 (~ 544 ms)	0.4495	0.0997	4.510	<.001
Memorability Score x Timepoint 16 (~ 578 ms)	0.4439	0.1000	4.438	<.001
Memorability Score x Timepoint 17 (~ 612ms)	0.4098	0.0998	4.105	<.001
Memorability Score x Timepoint 18 (~ 646 ms)	0.3556	0.1002	3.548	<.001
Memorability Score x Timepoint 19 (~ 680 ms)	0.2963	0.1001	2.961	0.003
Memorability Score x Timepoint 20 (~ 714 ms)	0.3181	0.0100	3.183	0.001
Memorability Score x Timepoint 21 (~ 748 ms)	0.2381	0.1001	2.378	0.017
Memorability Score x Timepoint 22 (~ 782 ms)	0.2353	0.1005	2.341	0.019
Memorability Score x Timepoint 23 (~ 816 ms)	0.2354	0.1007	2.339	0.119

5   CONCLUSIONS

The aim of this study was to investigate whether a difference in amplitude of ERP's could be observed within the first second after onset of thousand three-second videos in the Memento10k dataset between videos that participants would later indicate as remembered versus not remembered, as part of the MediaEval 2022 Predicting Video Memorability task [2]. A significant difference in amplitude was found at four timepoints, 340ms, 374ms, 408ms and 476ms in the posterior brain region. This difference was found in the opposite direction than was hypothesized, as the neural oscillations were of less positivity for remembered videos than for not-remembered videos. Contrary to the hypothesis based on [5], that showed a greater positivity for remembered videos than for non-remembered videos. In addition, a significant difference in amplitude in the right temporal lobe was observed, in the 306 – 816 ms time window. These results are in line with earlier studies [5][7][8][9] that have observed a greater positive difference in ERP amplitude around the 300ms mark after onset of stimulus, up until one second after the onset of the stimulus. The results from this project add evidence that suggests successful remembrance of a video later on can be predicted by looking for a P300 component after the onset of the video in the right temporal lobe. In the visual cortex, a positive peak was observed for the not-remembered videos. Future studies could explore the interaction effect between this positivity and negativity around the 300 ms time window after onset of a video.

This study adds to the body of literature in a way that it provides evidence for the importance of a P300 component for memorability beyond the scope of image research. In addition, the paradigm in which the videos were presented to the participants, can be argued as mimicking how a person would normally perceive visual imput via, e.g. social media. Most media platforms nowadays consist of an endless stream of very short videos and only certain grab and hold the attention of the viewer beyond the point of retrieval. This is in line with the increase in interest described in the introduction with regard to predicting video memorability in modern media.

However, further studies would be needed to investigate further how video relates to image literature beyond the scope of this paper. As only neurophysiological data of the first second of the videos was recorded in this paper, a limitation to this study is to draw any bridging conclusions between the literature of image memorability to video memorability. Even though the dataset consists of moving pictures, it can be argued that one second of viewing is not extensive enough to compare video and image media. Full recordings of the entire video could provide more insight as to where the spark of activation occurs over time and potentially over the biological structure of the brain. Future studies would have to indicate whether other brain regions such described in [7][8][9] (e.g., anterior prefrontal cortex, parietal cortex and the medial-frontal areas) show importance during a video viewing task and thus extending this body of knowledge beyond the scope of image research. It could be interesting to combine the behavioral data and EEG data in this study with diffusion magnetic resonance imaging, to deepen the understanding of the interplay between those brain regions that might be a factor in the later remembrance of a short video.

In addition, future research is necessary to investigate the features that play a role in memorability. This study only included the first second of the video in final analysis and it was assumed that since the videos were only three-seconds long, the main element of the subject of the video would already present itself in the first second of the video. Feature extraction could, however, analyze this statement in more depth to investigate whether there are common subject features that can be extracted from the Memento10k dataset that ensure memorability. This has the potential to answer the questions from the introduction section of this paper, to investigate what, if any, features of a short video can predict memorability. If a common benchmark of easily remembered features in videos can be established, this has a wide range of practical implementations for advertisers, influencers and other stakeholders that might have an interest of getting short videos remembered.

To conclude, this present paper provides evidence for a P300 effect differences in the posterior brain regions and right temporal lobe at retrieval for the predictability of memorability of videos from the Memento10k dataset [3]. More research is needed to understand how this data relates to image studies in the past and feature extraction could potentially identify common features that correlate with an increase in memorability.

REFERENCES

[1] Jo, S. Y., & Jeong, J. W. (2020). Prediction of visual memorability with EEG signals: A comparative study. *Sensors*, *20*(9), 2694.

[2] Sweeney, L., Constantin, M., Demarty, C.H., Fosco, C., Herrera, A., Halder, S., Healy, G., Ionescu, B., Matran-Fernandez, A., Smeaton, A., & Sultana, M. (2023). Overview of The MediaEval 2022 Predicting Video Memorability Task. In *MediaEval Multimedia Benchmark Workshop Working Notes*.

[3] Maria Authorsen and Jacques de Coauthor. Cool Task: Challenges, Dataset and Evaluation. Proc. of the MediaEval 2022 Workshop, Bergen, Norway and Online, 12-13 January 2023.

[4] Shen, W., Liu, Z., Ball, L. J., Huang, T., Yuan, Y., Bai, H., & Hua, M. (2020). Easy to remember, easy to forget? The memorability of creative advertisements. *Creativity Research Journal*, *32*(3), 313-322.

[5] Osipova, D., Takashima, A., Oostenveld, R., Fernández, G., Maris, E., & Jensen, O. (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. *Journal of neuroscience*, *26*(28), 7523-7531.

[6] Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high-and low-frequency words. *Memory & cognition*, *18*(4), 367-379.

[7] Rugg, M. D., & Coles, M. G. (1995). Electrophysiology of mind: Event-related brain potentials and cognition. Oxford University Press.

[8] Konishi, S., Wheeler, M. E., Donaldson, D. I., & Buckner, R. L. (2000). Neural correlates of episodic retrieval success. *Neuroimage*, *12*(3), 276-286.

[9] Donaldson, D. I., Petersen, S. E., & Buckner, R. L. (2001). Dissociating memory retrieval processes using fMRI: evidence that priming does not support recognition memory. *Neuron*, *31*(6), 1047-1059.

[10] Henson, R. N., Rugg, M. D., Shallice, T., Josephs, O., & Dolan, R. J. (1999). Recollection and familiarity in recognition memory: an event-related functional magnetic resonance imaging study. *Journal of neuroscience*, *19*(10), 3962-3972.

[11] Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128-2148.

[12] RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.