

Overview of the EEG Pilot Subtask at MediaEval 2021: Predicting Media Memorability

Lorin Sweeney¹, Ana Matran-Fernandez², Sebastian Halder²,
Alba G. Seco de Herrera², Alan Smeaton¹, Graham Healy¹

¹School of Computing, Dublin City University

²School of Computer Science and Electronic Engineering, University of Essex

lorin.sweeney8@mail.dcu.ie,{amatra,s.halder,alba.garcia}@essex.ac.uk,{alan.smeaton,graham.healy}@dcu.ie

ABSTRACT

The aim of the Memorability-EEG pilot subtask at MediaEval'2021 is to promote interest in the use of neural signals—either alone or in combination with other data sources—in the context of predicting video memorability by highlighting the utility of EEG data. The dataset created consists of pre-extracted features from EEG recordings of subjects while watching a subset of videos from Predicting Media Memorability subtask 1. This demonstration pilot gives interested researchers a sense of how neural signals can be used without any prior domain knowledge, and enables them to do so in a future memorability task. The dataset can be used to support the exploration of novel machine learning and processing strategies for predicting video memorability, while potentially increasing interdisciplinary interest in the subject of memorability, and opening the door to new combined EEG-computer vision approaches.

1 INTRODUCTION AND RELATED WORK

Even though the nature and constitution of people's memories remains elusive, and our understanding of what makes one thing more/less memorable than another is still nascent, combining computational (e.g., machine learning) and neurophysiological (e.g., electroencephalography; EEG) tools to investigate the mechanisms (formation and recall) of memory may offer insights that would be otherwise unobtainable. While EEG is not a tool that can directly explain what makes a video more/less memorable, it can help us trim the umbral undergrowth surrounding the subject, shedding light and offering a potential leap forward in our understanding of the interplay between the mechanisms of memory and memorability.

The purpose of this pilot study at MediaEval'2021 [8] was to collect enough EEG data for proof of concept and demonstration purposes, showcasing what could be done in subsequent work on predicting media memorability. The study involved the collection, filtering, and interpretation of neurophysiological data, and the use and evaluation of machine learning methods to enable the assessment of EEG data as a predictor of video memorability. The study has culminated in a demonstration of the utility of EEG in the context of video memorability, along with the public release of processed EEG features for others to explore¹. This study has the potential to not only broaden the research horizons of computing

¹Dataset and examples of use, as well as the code to replicate the results in this paper, are available at <https://osf.io/zt6n9/>

researchers, allowing them to explore and leverage EEG features without any of the requisite domain knowledge, but also increase the interdisciplinary interest in the subject of memorability more broadly.

Applying EEG to the question of whether an experience will be subsequently remembered or forgotten is a well researched area [7, 9, 12, 15]. Memorability, however, has been shown to be distinct from subsequent memory effects [1, 14], and received little interdisciplinary attention. Additionally, even though the application of machine learning to EEG is an active area of interest—allowing for the automation or augmentation of neurological diagnostics [3–5, 10], and the classification of emotional states [16], mental tasks [11], and sleep stages [2]—the use of EEG to predict visual memorability has yet to be firmly established, and was previously limited to static content [6]. To the best of our knowledge, this paper outlines the first application of EEG to video memorability.

2 EXPERIMENT DESIGN AND STRUCTURE

The stimuli used in the study are a subset of the subtask 1 data (i.e., the short-term video memorability prediction task) in MediaEval'2021 [8], and consists of 450 videos, 96 of which were designated as targets and selected to reflect the bottom and top 50 memorable videos from the TRECVID dataset, 200 were selected to reflect the next top and bottom 100, and 100 were selected to reflect the middle 100 memorable videos (95 selected + 5 duplicates) from the set of subtask videos. EEG data was collected from 11 subjects while they completed a short-term memory experiment, which was used to annotate the videos for memorability. EEG data acquisition² was carried out in two separate locations using a shared experimental procedure, and each location annotated the same set of videos. Rather than being split into separate encoding and recognition phases, the experiment was continuous in nature.

Before the experiment was carried out, participants were given a verbal description of the experiment procedure, presented with a set of written instructions, and taken through a practice run of 3 videos to familiarise them with the experiment. The experiment used a total of 450 videos, 192 of which were the target videos (96 targets, shown twice), and the remaining 258 videos were the fillers. The experiment was broken into 9 blocks of 50 videos, where a

²Data collection for participants 1–5 was carried out at Dublin City University (DCU) with approval from the university's Research Ethics Committee (DCUREC / 2021 / 171), and for participants 6–11 at the University of Essex (UoE) with approval from the Ethics Committee (ETH2122-0001). Data at DCU was collected using a 32-channel ANT Neuro eego system with a sampling rate of 1000 Hz. Data at UoE was collected using a 64-channel BioSemi ActiveTwo system at a sampling rate of 2048 Hz.

fixation cross was displayed for 3–4.5s, followed by the video presentation for its ~6 second duration, followed by a “get ready to answer” prompt of 1–3 seconds, followed by a 3s period for recognition response (repeated video or not). The time per block was approximately 700 seconds (~12 minutes) without accounting for 30-second closed/open eye baselines and breaks, which occurred between blocks. In order to account for recency effects, the first 50 videos presented did not include targets, but had 5 filler repeats, and the presentation positions of targets between each of the participants was pseudo-randomised, with the distances between target and repeat videos roughly fitting a uniform distribution, and the position of each block aside from block 1 being rotated by 1 for each participant.

3 ANALYSIS AND RESULTS

EEG data from both locations were processed in the same way for the 30 channels that were common across the two setups: data were first referenced using a common average and band-pass filtered between 0.1–30 Hz using a symmetric linear-phase FIR filter. Independent Component Analysis (ICA) was used to remove artifacts, and trial rejection using subject-specific thresholds was applied.

To establish a baseline using features extracted from the time domain, the EEG was low-pass filtered with a cutoff frequency of 15 Hz and downsampled to 30 Hz. We applied baseline correction to the average of the 250-ms pre-stimulus interval and extracted the data corresponding to the first second of each repeated clip, from each of the 30 channels, and concatenated it to form a feature vector. We term these the Event-Related Potential (ERP) features.

A second set of features were extracted from the EEG, this time from the time-frequency domain, which we refer to as ERSP (Event-Related Spectral Perturbation) features. For this, we extracted 4-second long epochs and computed a trial-by-trial time-frequency representation using Morlet wavelets for frequencies between 2–30 Hz. For this set of features, we used data from only 4 channels, namely Fz, Cz, Pz, and O1.

Since there were very few forgotten clips, in this task we differentiate between the first and the second viewing of clips that were successfully remembered based only on EEG data. To establish a baseline, we standardised the data to have mean zero and unit standard deviation, and used *scikit-learn*'s Bayesian Ridge regressor with default parameters. Results were obtained through 20-fold cross-validation with a 20% train-test split, separately for ERP and ERSP features. The individual classification results for each participant are shown in Table 1, measured using Area Under the Receiver Operating Characteristic Curve [13].

4 DISCUSSION AND OUTLOOK

This was an exploratory pilot task to guide the development of a future experimental protocol for capturing EEG signatures relating to successful memory encoding and retrieval to be used in predicting video memorability. While our experimental protocol resulted in too little data to examine differences between successful and unsuccessful encoding, we show EEG-related differences exist between the encoding and recognition phases of previously seen videos. These results indicate that EEG signatures relating to memory processes for video are present, and thus suitable to be

Table 1: Mean AUC values obtained for each participant across all folds, separately for ERP and ERSP features.

Participant	ERP-based classification	ERSP-based classification
1	0.564 ± 0.09	0.522 ± 0.09
2	0.585 ± 0.11	0.558 ± 0.07
3	0.520 ± 0.07	0.532 ± 0.07
4	0.666 ± 0.07	0.626 ± 0.09
5	0.714 ± 0.06	0.649 ± 0.08
6	0.555 ± 0.11	0.522 ± 0.10
7	0.601 ± 0.10	0.525 ± 0.08
8	0.590 ± 0.08	0.674 ± 0.08
9	0.609 ± 0.09	0.489 ± 0.06
10	0.628 ± 0.06	0.618 ± 0.09
11	0.477 ± 0.08	0.611 ± 0.12
Mean	0.591 ± 0.06	0.575 ± 0.06

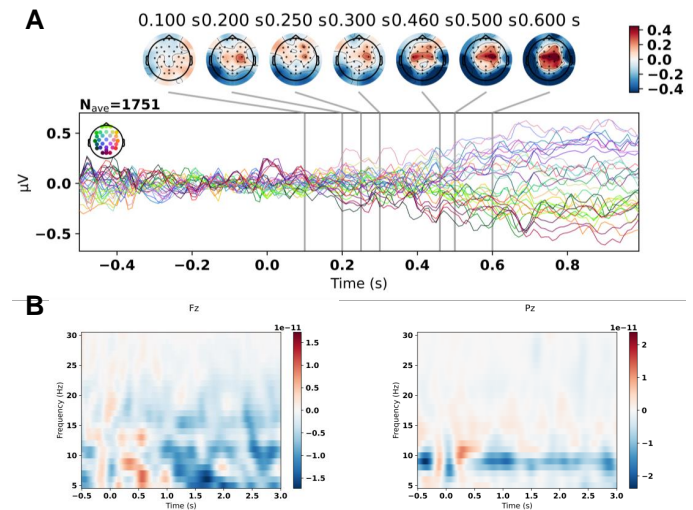


Figure 1: Grand-averaged butterfly plot showing differences in EEG activity for the second minus first presentation for videos for the first second (top-A). Averaged time-frequency differences in power for the second presentation minus that for the first presentation of videos for the first 3 seconds for channels Fz and Pz (bottom-B left and right, resp.).

collected with a revised experimental protocol and more participants to support a future fully-fledged task for predicting video memorability. The preprocessed EEG data captured is released to the research community.

ACKNOWLEDGMENTS

This work was part-funded by NIST Award No. 60NANB19D155 and by Science Foundation Ireland under grant number SFI/12/RC/2289_P2.

REFERENCES

- [1] Wilma A Bainbridge, Daniel D Dilks, and Aude Oliva. 2017. Memorability: A stimulus-driven perceptual neural signature distinctive from memory. *NeuroImage* 149 (2017), 141–152.
- [2] Farideh Ebrahimi, Mohammad Mikaeili, Edson Estrada, and Homer Nazeran. 2008. Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 1151–1154.
- [3] Denis A Engemann, Federico Raimondo, Jean-Rémi King, Benjamin Rohaut, Gilles Louppe, Frédéric Faueras, Jitka Annen, Helena Cassol, Olivia Gosseries, Diego Fernandez-Slezak, and others. 2018. Robust EEG-based cross-site and cross-protocol classification of states of consciousness. *Brain* 141, 11 (2018), 3179–3192.
- [4] Behshad Hosseinifard, Mohammad Hassan Moradi, and Reza Rostami. 2013. Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Computer methods and programs in biomedicine* 109, 3 (2013), 339–345.
- [5] Cosimo Ieracitano, Nadia Mammone, Amir Hussain, and Francesco C Morabito. 2020. A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia. *Neural Networks* 123 (2020), 176–190.
- [6] Sang-Yeong Jo and Jin-Woo Jeong. 2020. Prediction of visual memorability with EEG signals: A comparative study. *Sensors* 20, 9 (2020), 2694.
- [7] Demetrios Karis, Monica Fabiani, and Emanuel Donchin. 1984. “P300” and memory: Individual differences in the von Restorff effect. *Cognitive Psychology* 16, 2 (1984), 177–216.
- [8] Rukiye Savran Kiziltepe, Mihai Gabriel Constantin, Claire-Hélène Demarty, Graham Healy, Camilo Fosco, Alba García Seco de Herrera, Sebastian Halder, Bogdan Ionescu, Ana Matran-Fernandez, Alan F. Smeaton, and Lorin Sweeney. 2021. Overview of The MediaEval 2021 Predicting Media Memorability Task. In *Working Notes Proceedings of the MediaEval 2021 Workshop*.
- [9] Wolfgang Klimesch. 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews* 29, 2-3 (1999), 169–195.
- [10] Christoph Lehmann, Thomas Koenig, Vesna Jelic, Leslie Prichep, Roy E John, Lars-Olof Wahlund, Yadolah Dodge, and Thomas Dierks. 2007. Application and comparison of classification algorithms for recognition of Alzheimer’s disease in electrical brain activity (EEG). *Journal of neuroscience methods* 161, 2 (2007), 342–350.
- [11] Nan-Ying Liang, Paramasivan Saratchandran, Guang-Bin Huang, and Narasimhan Sundararajan. 2006. Classification of mental tasks from EEG signals using extreme learning machine. *International journal of neural systems* 16, 01 (2006), 29–38.
- [12] Eunho Noh, Grit Herzmann, Tim Curran, and Virginia R de Sa. 2014. Using single-trial EEG to predict and analyze subsequent memory. *NeuroImage* 84 (2014), 712–723.
- [13] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, and others. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* 12 (2011), 2825–2830.
- [14] Michael D Rugg and Tim Curran. 2007. Event-related potentials and recognition memory. *Trends in cognitive sciences* 11, 6 (2007), 251–257.
- [15] Thomas F Sanquist, John W Rohrbaugh, Karl Syndulko, and Donald B Lindsley. 1980. Electro cortical signs of levels of processing: Perceptual analysis and recognition memory. *Psychophysiology* 17, 6 (1980), 568–576.
- [16] Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu. 2014. Emotional state classification from EEG data using machine learning approach. *Neurocomputing* 129 (2014), 94–106.