Moving toward reality: Electrocortical reactivity to naturalistic multimodal emotional videos

Dean Sabatinelli[1], Andrew Farkas[1], and Matthew Gehr[1]

[1]University of Georgia

August 4, 2023

Abstract

While previous research has investigated the effects of emotional videos on peripheral physiological measures and conscious experience, this study extends the research to include cortical measures, specifically the steady-state visual evoked potential (ssVEP). A carefully curated set of 45 videos, designed to represent a wide range of emotional and neutral content, were presented with a flickering border. The videos featured a continuous single-shot perspective, natural soundtrack, and excluded elements associated with professional films, to enhance realism. The results demonstrate a consistent reduction in ssVEP amplitude during emotional videos which strongly correlates with the rated emotional intensity of the clips. This suggests that narrative audiovisual stimuli have the potential to track dynamic emotional processing in the cortex, providing new avenues for research in affective neuroscience. The findings highlight the potential of using realistic video stimuli to investigate how the human brain processes emotional events in a paradigm that increases ecological validity. Future studies can further develop this paradigm by expanding the video set, targeting specific cortical networks, and manipulating narrative predictability. Overall, this study establishes a foundation for investigating emotional perception using realistic video stimuli and has the potential to expand our understanding of real-world emotional processing in the human brain.

Moving toward reality: Electrocortical reactivity to naturalistic multimodal emotional videos

Dean Sabatinelli[1,2], Andrew H. Farkas[1], & Matthew C. Gehr[1]

[1]Department of Psychology

[2]Department of Neuroscience

University of Georgia, Athens, GA

Corresponding author:

Dean Sabatinelli, sabat@uga.edu

Keywords: Emotion, EEG, video, ssVEP, ecological validity, realism

Abstract

While previous research has investigated the effects of emotional videos on peripheral physiological measures and conscious experience, this study extends the research to include cortical measures, specifically the steady-state visual evoked potential (ssVEP). A carefully curated set of 45 videos, designed to represent a wide range of emotional and neutral content, were presented with a flickering border. The videos featured a continuous single-shot perspective, natural soundtrack, and excluded elements associated with professional films, to enhance realism. The results demonstrate a consistent reduction in ssVEP amplitude during emotional videos which strongly correlates with the rated emotional intensity of the clips. This suggests that narrative audiovisual stimuli have the potential to track dynamic emotional processing in the cortex, providing new

avenues for research in affective neuroscience. The findings highlight the potential of using realistic video stimuli to investigate how the human brain processes emotional events in a paradigm that increases ecological validity. Future studies can further develop this paradigm by expanding the video set, targeting specific cortical networks, and manipulating narrative predictability. Overall, this study establishes a foundation for investigating emotional perception using realistic video stimuli and has the potential to expand our understanding of real-world emotional processing in the human brain.

The substantial impact of smartphone videos on society is unquestionable. Watching a recording of a spontaneous event carries the viewer into a situation with realism and context that is missing in photographs. Seeing and hearing the natural sounds as a narrative unfolds captures our attention and can evoke powerful, long-lasting emotional states (Burns et al., 2008; Holman et al., 2014). Some studies have added emotional audio tracks (Brown & Cavanaugh, 2016; Gerdes et al., 2013) or music (Spreckelmeyer et al., 2006) to static emotional scene presentations while collecting ERPs to scene onset, and report modulatory effects primarily on early latency components. However, the complexity of multimodal video makes it a challenging stimulus to control experimentally, particularly for electrocortical measures.

Several peripheral psychophysiological studies have presented participants with studio film excerpts to evoke emotion. Emotional as compared to neutral videos lead to elevated skin conductance, differential heart rate modulation, and expressive facial muscle activity similar to that evoked by emotional and neutral scenes (Bos et al., 2013; Codispoti et al., 2008; Christie & Friedman, 2004; Kolodyazhniy et al., 2011; Koruth et al., 2015; Kriebig et al., 2007; Palomba et al., 2000). Affective startle modulation during video perception also shows effects consistent with emotional scene studies (Bradley, 2007; Kaviani et al, 1999; Koukounas & McCabe, 2001). In a series of studies using 27 content-matched emotional videos and scenes, videos were shown to enhance ratings of emotional intensity (Detenber et al., 1998; Simons et al. 1999; Simons et al., 2000; but see also Detenber & Reeves, 1996) and elevate skin conductance (Detenber et al., 1998; Simons et al., 1999; but see also Simons et al, 2000). These studies of peripheral and reflex physiology demonstrate the utility of emotional videos and suggest that the video may offer a more potent medium to induce states of emotion in a research setting.

To our knowledge, only 2 studies have used video stimuli to induce emotional states while recording EEG, both of which assessed changes in alpha-band power. One reported no differential effects of fearful, sad, or neutral video content on frontal alpha power (Dennis & Solomon, 2010), and one employed both videos and scenes (Simons et al., 2003), and reported reliable reductions in parietal alpha power during arousing, compared to neutral videos. A possible reason for the scarcity of EEG studies of emotional video perception is simply methodological, in that there is no single stimulus event from which an event-related potential (ERP) can be averaged, the most common means of analyzing electrocortical activity.

This practical difficulty could be circumvented through the use of steady-state visual evoked potentials (ssVEP) in combination with video presentation. Just as instructed spatial attention toward flickering stimuli enhances ssVEP amplitude at the attended flicker frequency (Hillyard et al., 1997), motivated attention enhances ssVEP amplitude during emotional relative to neutral flickering scene perception (Bekhtereva et al., 2015, 2021; Keil et al., 2003; 2005). Conversely, if nonflickering scenes are presented in combination with a competing flickering visual stimulus, ssVEP amplitude is reduced during emotional relative to neutral scene perception (Deweese et al., 2014; Müller et al., 2008; Weiser et al., 2016). Here we explore whether multimodal videos presented with a flickering border might show a similar modulation, such that emotional videos would be associated with reduced ssVEP amplitude compared to neutral videos.

Several emotional video sets have been assembled and used to investigate emotional reactivity in self report and peripheral physiology. However, these emotional video sets are not optimal for EEG recording, for several reasons. Nearly all video sets use long and variable clip duration (1-15 minutes; Gilman et al., 2017; Gross & Levenson, 1995; Hewig et al., 2005; Kreibig et al., 2007; Philippot, 1993; Schaefer et al., 2011). Many emotion film sets also lack audio, or include narration and music (Cowen & Keltner, 2017; Gilman et al., 2017; Philippot, 1993; Samson et al., 2016). In addition, all available video sets contain primarily content drawn from studio productions. These professional films often switch between multiple camera viewpoints, such that the viewer experiences the scene from more than one perspective, creating a 'suspension of disbelief' and limiting ecological validity (Holland, 2003; Lee, 2004; Prentice & Gerrig, 1999). Studio films also often feature well-known actors and special effects that lead the viewer to recognize that the action is artificial, which suppresses emotional reactivity (Gross & Thompson, 2007; Hacjak, MacNamara & Olvet, 2010). Ideally, a video intended to evoke emotion might present a situation as a true event, as it actually happened in the real world. Because of the high prevalence smartphone video ownership, and the public posting of these videos, there is now sufficient 'raw material' available to assemble a realistic video set for experimental use.

Compared to static scenes, multimodal video clips present several problems with regard to experimental control when recording electroencephalography (EEG). In order to interpret the potential differences in electrocortical reactivity across video contents, the video set should not differ systematically in basic sensory and perceptual features, while retaining factors that promote realism (Allison, Wilcox, & Kazimi, 2013; Lin & Peng, 2015). These include a continuous single shot, a landscape perspective that places the viewer in a ground-level position, ambient sound track without music or narration, and videographic quality that does not suggest professional creation, to enhance the impression that the clip represents a genuine event. A relatively brief duration of 10 seconds was chosen to provide short enough periods to enable multiple exemplars from different emotional content categories, but long enough to engage and sustain an emotional state during which a cortical response could be recorded.

Based on past work with emotional scenes, we hypothesize that emotional video perception will result in reduced ssVEP amplitude toward the competing flickering border. As with past ssVEP and ERP studies of emotional perception, the impact of pleasant and unpleasant videos is expected to be equivalent, and strongly related to arousal ratings (Bekhtereva et al., 2015; Cuthbert, et al., 2000; Frank & Sabatinelli, 2019; Keil et al., 2008). Alternatively, the dynamic nature of the video stimuli and the addition of an ambient audio track may engage continuous perceptual processing across all videos to the extent that any differential impact of emotional videos on ssVEP amplitude fails to rise above the ssVEP variability.

Methods

Participants . Forty-four participants were recruited from the University of Georgia student body and were compensated with course credit. All participants gave informed consent after reading a description of the study approved by the University of Georgia Human Subjects Institutional Review Board. Data from 2 participants were excluded due to excessive EEG artifacts (described below) leaving 42 participants in all subsequent analyses. The mean age for final sample was 18.9 years ($SD = 1$) with a range from 18 to 22. Thirty-four of the participants identified as female, 6 of the participants identified as Asian, 2 as Black, 1 as multiracial, and 33 as White.

Video Stimuli Set . The goal in assembling the video set was to represent a wide variety of emotional and neutral content, while holding basic perceptual features of the videos reasonably constant. The videos were clipped to 10 seconds in duration with the intention to allow the viewer to recognize the nature of the situation quickly, with the narrative maintained across the interval. The videos featured a single lens perspective that placed the viewer roughly at eye level, and included a natural soundtrack. The videos excluded recognizable actors or high production value elements associated with professional film (e.g. expert lighting, composition). The intention was to minimize artificial elements that may break the viewer's belief that the clips depict true events. Videos were selected to depict roughly equivalent degrees of loudness and motion across emotional and neutral videos to avoid a confound of emotion and action.

Forty-five clips meeting these criteria were collected from the internet that depicted a range of pleasant, neutral, and unpleasant situations. These videos were divided into 5 groups of 9 videos each. Four of these video categories depicted emotional situations, judged by the experimenters to be highly arousing (roller coasters, passionate couples, graphic surgery, direct threats) or modestly arousing content (puppies, cute babies; indirect threats). One group of 9 videos depicted active, but common life experiences (walking down a busy street, a kitchen staff hard at work). The video clips were quantified on a number of basic perceptual qualities. This included sound intensity, measured with a BAFX 3370 Digital Sound Level Meter placed at ear level at the participant's chair. Audio was presented with a Dell A525 3-speaker system with subwoofer, placed directed under the video monitor, and loudness was assessed as decibel (A weighted) values recorded twice per second, and averaged across each video. Brightness was defined by converting the color videos to grayscale and averaging the 0-255 values for each video frame. Movement depicted in the videos was quantified as the average difference in grayscale pixel values between successive frame of each video, using the Magick R package version 2.7.3 (Ooms, 2021). For example, during periods with high levels of movement in a video, many pixels will show large changes in brightness from frame to frame. These changes were averaged to yield a score representing the total movement represented in each video. Lastly, Shannon's entropy was used as an index of perceptual complexity, by quantifying the entropy value within each frame to provide a mean and standard deviation across each 10 s video. These quantified video features will be correlated with the electrocortical data along with emotional ratings of pleasantness and arousal collected from our participant sample.

Experimental Design and Procedure. After providing informed consent, each participant was given instructions and seated in a chamber shielded for sound and electromagnetic noise. A 64-channel EEG net (described below) was placed and adjusted over the course of 10-15 minutes. The research assistants then reminded the participants to remain still and maintain fixation on a red cross at the center of the video screen throughout the series. Participants were also asked to avoid blinking during each video clip presentation, to the best of their ability.

The video series started with an acclimation trial, which presented a 10 s fixed checkerboard surrounded by a flickering border. A delay of 12 s was then followed by the 45 experimental video clips, which were arranged in a pseudo-randomized order with an average inter-trial interval (ITI) of 12 s (range 10 to 14 s). Videos were ordered such that no more than two videos from the same category were shown in succession, and that video contents were equivalently distributed across the series.

PsychoPy open source software (Peirce et al., 2019) was used to present the video clips and send triggers to the EEG acquisition computer. The PsychoPy control files and video stimuli are available on Open Science Framework (final link TBD). A Dell Optiplex 380 computer presented videos to a 60 Hz Westinghouse 32-in LCD monitor, which was placed 1.6 m from the participants eyes, at a 960 by 648 pixel resolution, with the video clip shown in the central 720 by 405 pixels (20º x 15º visual angle). The remaining monitor space displayed a gray border (RGB value of [148, 148, 148]) around the video which flickered to black (RGB value of [0, 0, 0]) at a 7.5 Hz frequency to evoke a steady-state visual potential. To ensure a precise flash rate, the black border was drawn every eighth screen refresh on the 60 Hz monitor. The presentation code logged the exact time of each flash of the border, which was consistent at 7.5 Hz for all videos and all participants.

Following the presentation of the video series, there was a brief (~2 m) break while research assistants entered the chamber and checked in with the participant and the status of the sensors. The video series was then repeated in a new pseudo-random order. After the second video block, the recording equipment was removed, and participants left the chamber and were given towels and water to clean the electrolyte gel out of their hair. Participants then sat at a computer and viewed the 45 videos again and provided ratings of each video using a computer-based Qualtrics survey. Videos were viewed to completion before participants could give their ratings and advance to the next video. Ratings of experienced valence and arousal were recorded using a computer-based version of the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994). Participants input their ratings of pleasantness and arousal on a scale of 1-9 in units of .1 by clicking and dragging a cursor. After ratings were completed, participants filled out a brief post-experimental questionnaire, and were debriefed on the rationale for the experiment.

EEG Data Collection. The EEG data were recorded using a BioSemi ActiveTwo 64-channel system (BioSemi Amsterdam, Netherlands). The electrode cap was positioned according to the 10-20 system. Data were recorded continuously in reference to common mode electrodes (CMS and DRL). The electrode offsets were kept between 50 and -50 millivolts before data collection. The data were sampled at 512 Hz with no online low- or high-pass filters. Triggers corresponding to video onsets were sent from the PsychoPy presentation computer to the EEG system via a BioSemi USB Trigger Interface cable.

EEG preprocessing. The EEG data were preprocessed using the MATLAB-based Electro Magnetic Encephalography Software (EMEGS; emegs.org; Peyk, De Cesarei, & Junghöfer, 2011). The software implements an artifact correction procedure designed for use with dense array EEG (Junghöfer, Elbert, Tucker, & Rockstroh, 2000). Offline, the EEG data were filtered using a low-pass Butterworth filter with a passband of 30 Hz and stopband of 40 Hz. Because we were primarily interested in our steady-state driven frequency of 7.5 Hz, we used a high-pass filter with a passband of 3 Hz and stopband of 1 Hz. From the continuously recorded data, each trial was segmented from 100 ms prior to 10 s after video onset.

To identify trials and sensors that were contaminated by artifacts, sensor by trial distributions were made by a composite measure from the maximum amplitudes, standard deviation, and the maximum first derivative by EMEGS (Junghöfer et al., 2000). These distributions were jointly used to locate noisy channels as compared to the distribution medians. Trials that contained excess artifact are removed by EMEGS, and sensors that contain excess artifact are replaced by spherical spline interpolation with a weighted average of all remaining sensors. The largest weights for this procedure go to the closest sensors that have the smallest standard deviations. The data were then re-referenced to an average reference and the artifact detection procedure was repeated.

To estimate the elicited ssVEP amplitude, waveforms for each video were created using a custom-built function in MATLAB. The function used a Hilbert transform to isolate the 7.5 Hz phase and derive the instantaneous amplitude at this frequency. This allowed us to transform our data to represent the border-driven 7.5 Hz amplitude over each video presentation. Due to onset / offset transition effects, the first and last second of video duration was excluded from analysis, leaving 8 seconds of ssVEP amplitude.

Following preprocessing, trials were averaged together by category for each of the 42 participants. To be included in the final group, each participant were required to retain at least 50% of the trials from each video category, otherwise the participant was excluded from further analysis, which resulted in the exclusion of 2 participants. Of the remaining sample of 42, 78.1% of trials were retained (standard deviation 9%) across participants. Averaged ssVEP amplitude across the sample was used for the by-video correlation analyses with emotional ratings and video features.

ssVEP data analyses. The amplitude of the ssVEP signal was sampled from 14 occipital sensors in which the ssVEP signal is strongest, based on prior studies using similar designs (Müller et al., 2008; Deweese et al., 2014; Bekhtereva et al., 2017), and confirmed with topographical representation of 7.5 Hz power in the current dataset. These sensors included P1, P3, PO3, PO7, O1, Iz, Oz, POz, Pz, P2, P4, PO4, PO8, & O2. To estimate ssVEP amplitude, the middle 8 seconds of video presentation (from 1 to 9 seconds after onset) was averaged and used for all analyses to avoid onset and offset artifacts. There is considerable inter-subject variability in overall ssVEP amplitude (Moratti et al., 2004; Weiser et al., 2016), thus within-subject data were z-scored across video contents.

Self-reported emotion ratings and ssVEP amplitudes were analyzed across the 5 content categories using repeated measure ANOVAs corrected for violations of sphericity as needed, with effects broken down by paired t-tests, with Tukey correction. Effect sizes were quantified by the generalized eta squared ($\eta_{\Gamma}{}^{2}$) measure. Generalized eta can be interpreted such that .01 is a small effect, .06 a medium effect, and .14 a large effect (Olejnik & Algina, 2003). Pearson's correlations were used to assess the potential relationships between ssVEP amplitude and video estimates of self-reported valence, arousal, luminance, sound intensity, entropy, entropy SD, and pixel motion.

Results

Emotion Ratings. The video set was assembled to evoke both pleasant and unpleasant emotion states across a range of intensities. Consequently, a repeated measures ANOVA found a strong difference between SAM valence ratings between the 5 video categories (F (4,164) = 135.60, $p<$ .001; $\eta_{\Gamma}{}^{2}$ = .690). As shown in Figure 1, the pleasant videos elicited the largest valence ratings, followed by highly arousing pleasant, neutral, unpleasant, and highly arousing unpleasant videos, with all differences significant. This pattern of valence ratings is quite consistent with ratings of comparable categories of scenes (Farkas & Sabatinelli, 2023;

Frank & Sabatinelli, 2019). Emotional arousal was also strongly modulated by video content (F (4,164) = 46.64, $p < .001$;$\eta_{\Gamma}{}^{2} = .285$). Arousal ratings were equivalently enhanced by highly arousing pleasant, highly arousing unpleasant, and unpleasant videos. These were followed by arousal ratings for pleasant videos, while neutral videos were rated as least arousing. Table 1 lists the emotion rating means for each video, in addition to luminance, sound intensity, entropy, entropy standard deviation, and pixel motion.

Steady-State visual evoked potential. Steady state visual evoked potential amplitude did not differ between the first and second presentation of the videos (see Table 2) and were averaged together. Shown in Figure 2, raw steady-state evoked amplitude was maximal over occipital regions and significantly modulated by video content from 1-9 seconds after video onset (F (4,164) = 14.21, $p< .001$; $\eta_{\Gamma}{}^{2} = .013$), with decreases in 7.5 Hz power evident during emotional, compared to neutral videos. Standardizing ssVEP amplitude across participants yielded the same pattern of modulation, with considerably greater effect size (F (4,164) = 16.17, $p < .001$;$\eta_{\Gamma}{}^{2} = .283$). Shown in Figure 3, video categories sharing the same letter in the legend (a, b, or c) do not significantly differ. Thus, highly arousing unpleasant, unpleasant, and highly arousing pleasant videos reduced ssVEP amplitude equivalently, followed by ssVEP amplitude during pleasant videos, which differed only from ssVEP amplitude during high unpleasant and unpleasant videos. Lastly, neutral videos evoked the smallest reduction in ssVEP amplitude, different from ssVEP amplitude during all other video categories but pleasant, which was marginal (t = 2.72, p = .056). Also shown in Figure 3 is a histogram of individual participants' ssVEP amplitude across the 5 categories, representing the consistency of ssVEP modulation across video content across the sample.

Pearson's correlations were used to assess the relationship between raw ssVEP amplitude and emotional and perceptual video features. Thus the averaged ssVEP amplitude for the subset of participants with data for each of the 45 videos (average 34.2, SD = 4.6) was compared with the emotional ratings and quantitative measures from all 42 participants. This analysis could reveal relationships between ssVEP amplitude and video features that may be confounded with 5 video categories. For example, if videos with high luminance, sound intensity, or motion tended to reduce ssVEP amplitude, and occurred more often in emotional videos, correlations of these values may emerge if these features were not sufficiently distributed across the 5 categories. Shown in Table 3, ssVEP amplitude correlated only with arousal ratings (R^2 = .308, $p < .001$), while valence ratings and all perceptual features showed no significant association. A trend suggests a relationship between entropy SD (variability in perceptual complexity over time) across the video and ssVEP amplitude ($R^2 = .074$, $p = .07$). Shown in Figure 4, when averaged by video category across the sample of 42, the correlation of arousal ratings and ssVEP amplitude was very high ($R^2 = .97$, $p < .01$).

Discussion

In this study we demonstrated the feasibility of differentiating electrocortical reactivity evoked during multimodal emotional and neutral video perception. A considerable literature has employed video stimuli to evoke states of emotion while assessing peripheral physiological measures and conscious experience, and here we expand this area of study to cortical measures. This effort involved the curation of 45 brief video clips, and the addition of a competing visual flicker to evoke a steady-state response that serves as a continuous index of uninstructed emotional engagement toward each video clip. This novel paradigm resulted in reliably reduced ssVEP amplitude that was clearly correlated with the rated emotional intensity of emotional videos, both pleasant and unpleasant. Other perceptual video features that might be confounded with emotional intensity and modulate ssVEP amplitude independent from emotion did not show any significant relationships. These data demonstrate that narrative audiovisual stimuli can be employed to track dynamic emotional processing in the cortex, potentially enabling new research questions to be addressed in affective neuroscience. Realistic audiovisual clips have the potential to recruit multimodal brain networks involved in the dynamic perception of emotional situations and thus evoke emotional states that more closely represent real life experience. While technically challenging, this move toward increased ecological validity could prompt qualitatively different brain states and expand our opportunities to understand how the human brain processes realistic emotional events.

Multiple brain networks are involved in the processing of any dynamic, multimodal stimulus which continuously draws perceptual and integrative resources across visual and auditory modalities. The steep drop in ssVEP amplitude in the first second of video presentation (Figure 2) likely reflects the rapid shift in cortical entrainment from the initially dominant border flicker to the processing of the content of each video. The additional reduction in ssVEP amplitude during emotional, compared to neutral videos appears to represent the enhanced activation of widespread sensory and association areas engaged by frontal cortical and (indirectly) subcortical structures involved in emotional discrimination and response processing (Bo et al., 2021; Frank et al., 2019; Liu et al., 2012; Pessoa, 2017; Sabatinelli & Frank, 2019). Prior ssVEP studies of scene processing suggest an enhanced contribution of superior parietal, middle temporal and inferior frontal cortex during emotional compared to neutral scenes (Keil et al., 2009; 2012; Moratti et al., 2004) that may also support perceptual processing of motivationally relevant videos. The current study was intended to demonstrate the feasibility of the paradigm, and thus we focused on modulation of ssVEP amplitude over occipital sensors where the visual flicker entrainment is strongest. Given the relatively sparse coverage of the 64-channel EEG net and limited number of videos per category, a more detailed analysis was underpowered. In future work with a larger video set and greater EEG sensor density, source localization analyses might be applied to differentiate visual and auditory cortical contributions, as well as other cortical networks that are dynamically engaged during emotional video processing.

A recent and relevant study by Stegmann & colleagues (2022) employed classical conditioning of shock with 32 s neutral videos that placed the viewer's perspective as walking through empty office hallways, and used a flickering overlay (a black frame appeared in the video every 50 ms) to induce a ssVEP and thus track cortical engagement. Their data showed a decrease in ssVEP amplitude over occipital areas during CS+ acquisition, consistent with the effects of the current study, though our videos were inherently emotion-evoking. The use of a flickering border may therefore be equivalent to interspersed black frames, with the emotional or conditioned content resulting in reduced ssVEP amplitude.

Compared to scenes, video stimuli are difficult to standardize, as each exemplar is a series of 240 images, and their unfolding narrative is unpredictable to the viewer. The 45 videos assembled for this study were selected to convey a reasonably consistent narrative, without radically unexpected changes. For example, a video from the perspective of a person walking on a city street did not transform into a mugging, or a surprise reunion with an old friend. While narrative variability is a potential confound if uncontrolled, future studies might manipulate this predictability by including videos that shift from neutral to pleasant or unpleasant, to potentially reveal how emotional networks characterize changing events. The audio track could also be manipulated independent of the video clip to investigate the impact of consistent relative to conflicting information, such as ambiguous language and voice inflection. Subtle changes may have large effects in the interpretation of dynamic narratives which might be assessed with time-locked shifts of ssVEP amplitude to video events.

Limitations . While encouraging, the use of video stimuli to evoke emotional states involves several limitations. Compared to scenes, considerable labor is involved in video collection, editing, and balancing. The use of a flickering border to elicit the ssVEP works against the intention to move the evocative stimulus farther toward realism, and depends on an indirect reduction is flicker entrainment as the index of emotional engagement. The lack of a single stimulus onset precludes the averaging of event-related potentials, and the assessment of evoked cortical oscillations by video clips with scalp-derived EEG has not yet been demonstrated, perhaps due to the difficulty of capturing consistent emotional network activity across participants (Shen & Yi, 2019).

A common means of assessing emotional perception in the laboratory is with the late positive potential (LPP; Ferrari et al., 2017; Hajcak et al., 2010; Schindler & Bublatsky, 2020; Schupp & Kirmse, 2021). Thus it would be helpful to compare the modulation of the LPP with modulation of the ssVEP during video perception within a single sample. The individual pattern of modulation (e.g., a bias toward pleasant or unpleasant stimuli) and overall effect size of emotion may or may not be consistent, depending on the relationship between the conscious recognition and elaborative processing reflected in the LPP to a static scene and the ongoing deciphering and evaluation of videos.

Though arousal ratings clearly coincide with the degree of ssVEP reduction, the valence of the videos do not appear to differentially affect ssVEP amplitude, similar to the LPP (Frank & Sabatinelli, 2019; Codispoti et al., 2021). Future studies may exploit the 10 second duration to contrast pleasant and unpleasant videos to potentially reveal valence (or reward) related effects that may originate in anterior medial regions (Costa et al., 2010; Junghöfer et al., 2017; Sabatinelli et al., 2015) that may become evident during the comparatively long duration video stimuli.

In the current design, the video and its flickering border were presented simultaneously, thus intermixing the brain's response, and delaying the evidence of emotional modulation until entrainment had stabilized and the content of the video could be understood by the participant. Future research could separate these 2 events by initiating the flicker prior to video onset, ideally using an audiovisual stimulus that shared the basic sensory features of the upcoming video. This refinement to the paradigm may allow a more temporally resolved assessment of emotional reactivity (Bekhtereva et al., 2018).

Conclusions . In summary, this study found a significant modulation of ssVEP amplitude during naturalistic multimodal videos, which correlated strongly with rated arousal. This finding suggest that narrative audiovisual stimuli can be used to track emotional processing in the cortex, potentially enabling new research questions to be investigated which could facilitate a better understanding of how the human brain processes realistic emotional events. Future development and improvement can expand the utility of this approach to studying emotional processing in the laboratory.

Figure Legends

Figure 1 . Self-reported emotional valence and arousal ratings (1-9 scale) from the sample for the 45 video clips, with standard error bars.

Figure 2 . At left, a topographical distribution of 7.5 Hz ssVEP power, demonstrating a concentration over occipital sensors. At right, the time course of the sample's ssVEP power averaged over 14 occipital sensors for each of the 5 video contents.

Figure 3 . At top, the standardized ssVEP power average for each video content category, averaged from 1-9 seconds after onset of each 10 second video, with standard errors. Categories sharing a letter in the legend do not differ reliably. At bottom, a histogram of individual participants' average ssVEP power by video category, demonstrating the general consistency of the grand mean across the sample.

Figure 4 . The inverse relationship between rated arousal of the videos and ssVEP amplitude during video perception.

References

Allison, R. S., Wilcox, L. M., & Kazimi, A. (2013). Perceptual artefacts, suspension of disbelief and realism in stereoscopic 3D film.*Public* , 24(47), 149-160.

Bekhtereva, V., Craddock, M., & Muller, M. M. (2021). Emotional content overrides spatial attention. Psychophysiology, 58(8), 1–14. https://doi.org/10.1111/psyp.13847.

Bekhtereva, V., & Müller, M. M. (2015). Affective facilitation of early visual cortex during rapid picture presentation at 6 and 15 Hz. Social Cognitive and Affective Neuroscience, 10(12), 1623–1633. https://doi.org/10.1093/scan/nsv058

Bekhtereva, V., Pritschmann, R., Keil, A., & Muller, M. M. (2018). The neural signature of extracting emotional content from rapid visual streams at multiple presentation rates: A cross-laboratory study.*Psychophysiology* , 55(12), e13222.

Bradley, M. M., Keil, A., & Lang, P. J. (2012). Orienting and emotional perception: facilitation, attenuation, and interference. *Frontiers in Psychology* , *3* , 493.

Bradley MM, Lang PJ. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *J Behav Ther Exp Psychiatry* . 25:49–59.

Bradley, M. M., Sambuco, N., & Lang, P. J. (2022). Affective perception: The power is in the picture. In B. Ionescu, W. A. Bainbridge & N. Murray (Eds.) *Human perception of visual information* (pp. 59–83). Springer.

Bradley, S.D. (2007) Examining the Eyeblink Startle Reflex as a Measure of Emotion and Motivation to Television Programming, *Communication Methods and Measures* , 1:1, 7-30, doi: 10.1080/19312450709336658

Brown, D. R., & Cavanagh, J. F. (2017). The sound and the fury: Late positive potential is sensitive to sound affect.*Psychophysiology* , 54(12), 1812–1825. https://doi.org/10.1111/psyp.12959.

Bo, K., Yin, S., Liu, Y., Hu, Z., Meyyappan, S., Kim, S., Keil, A., & Ding, M. (2021). Decoding neural representations of affective scenes in retinotopic visual cortex. *Cerebral Cortex* , 31(6), 3047-3063.

Bos, M. G. N., Jentgens, P., Beckers, T., & Kindt, M. (2013). Psycho- physiological response patterns to affective film stimuli. *PLoS One* , 8(4):62661. https://doi.org/10.1371/journal.pone.0062661

Burns, C. M., Morley, J., Bradshaw, R., & Domene, J. (2008). The emotional impact on and coping strategies employed by police teams investigating internet child exploitation. *Traumatology* , 14(2), 20–31.

Christie, I.C.,& Friedman, B.H. (2004). Autonomic specificity of discrete emotion and dimensions of affective space: A multivariate approach. *International Journal of Psychophysiology* , 51, 143-153.

Codispoti, M., Surcinelli, P., & Baldaro, B. (2008). Watching emotional movies: Affective reactions and gender differences. *International Journal of Psychophysiology* , 69(2), 90–95. https://doi.org/10.1016/j.ijpsycho.2008.03.004

Codispoti, M., Micucci, A., & De Cesarei, A. (2021). Time will tell: Object categorization and emotional engagement during pro- cessing of degraded natural scenes. *Psychophysiology* , 58(1), 1– 16. https://doi.org/10.1111/psyp.13704

Costa, V.D., Lang, P.J., Sabatinelli, D., Versace, F., & Bradley, M.M. (2010). Emotional imagery: assessing pleasure and arousal in the brain's reward circuitry. *Human Brain Mapping* , 31, 1446-1457.

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences* , 114(38), E7900-E7909.

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biological psychology* , *52* (2), 95-111.

Dennis, T. A., & Solomon, B. (2010). Frontal EEG and emotion regulation: Electrocortical activity in response to emotional film clips is associated with reduced mood induction and attention interference effects. *Biological Psychology* , 85(3), 456–464. https://doi. org/10.1016/j.biopsycho.2010.09.008

Deweese, M. M., Bradley, M. M., Lang, P. J., Andersen, S. K., Müller, M. M., & Keil, A. (2014). Snake fearfulness is associated with sustained competitive biases to visual snake features: hypervigilance without avoidance. *Psychiatry Research* , 219(2), 329-335.

Detenber, B. H., & Reeves, B. (1996). A bio-informational theory of emo- tion: Motion and image size effects on viewers. *Journal of Communication* , 46, 66–84.

Detenber, B. H., Simons, R. F., & Bennett, G. G. (1998). Roll 'em: The effects of picture motion on emotional responses. *Journal of Broadcasting and Electronic Media* , 42, 112–126.

Ferrari, V., Codispoti, M., & Bradley, M. M. (2017). Repetition and ERPs during emotional scene processing: A selective review. International Journal of *Psychophysiology* , https://doi.org/10.1016/j.ijpsycho.2016.07.496

Frank, D. W., Costa, V. D., Averbeck, B. B. & Sabatinelli, D. (2019) Directional interconnectivity of the human amygdala, fusiform gyrus, and orbitofrontal cortex in emotional scene perception. *J. Neurophysiol. 122* , 1530–1537.

Frank, D. W., & Sabatinelli, D. (2019). Hemodynamic and electrocortical reactivity to specific scene contents in emotional perception. *Psychophysiology* , *56* (6), e13340.

Gerdes, A. B. M., Wieser, M. J., Bublatzky, F., Kusay, A., Plichta, M. M., and Alpers, G. W. (2013). Emotional sounds modulate early neural processing of emotional pictures. *Front. Psychol* . 4:741. doi: 10.3389/fpsyg.2013.00741

Gilman, T. L., Shaheen, R., Nylocks, K. M., Halachoff, D., Chapman, J., Flynn, J. J., . . . & Coifman, K. G. (2017). A film set for the elicitation of emotion in research: A comprehensive catalog derived from four decades of investigation. *Behavior research methods* , 49(6), 2061-2082.

Gross, J. J., and Levenson, R. W. (1995). Emotion elicitation using films. *Cogn. Emot* . 9, 87–108. doi: 10.1080/02699939508408966

Gross, J. J., & Thompson, R. A. (2007). Emotion Regulation: Conceptual foundations, *Handbook of emotion regulation* . New York: Guilford.

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: An integrative review.*Developmental Neuropsychology* , 35(2), 129–155. https://doi.org/10.1080/87565640903526504

Hewig, J., Hagemann, D., Seifert, J., Gollwitzer, M., Naumann, E., and Bartussek, D. (2005). A revised film set for the induction of basic emotions. *Cogn. Emot* . 19, 1095–1109. doi: 10.1080/02699930541000084

Hillyard SA, Hinrichs H, Tempelmann C, Morgan ST, Hansen JC, Scheich H, Heinze HJ (1997) Combining steady-state visual evoked potentials and fMRI to localize brain activity during selective attention. *Hum Brain Mapping* 5:287–292.

Holland, N. (2003). The willing suspension of disbelief: A neuro-psychoanalytic view. *PsyART* , 7.

Holman EA, Garfin DR and Silver RC (2014) Media's role in broadcasting acute stress following the Boston Marathon bombings. *Proceedings of the National Academy of Sciences of the United States of America*111, 93–98.

Kaviani, H., Gray, J. A., Checkley, S. A., Kumari, V., and Wilson, G. D. (1999). Modulation of the acoustic startle reflex by emotionally-toned film-clips. *Int. J. Psychophysiol* . 32, 47–54. doi: 10.1016/S0167-8760(98)00063-4

Keil, A., Gruber, T., Müller, M. M., Moratti, S., Stolarova, M., Bradley, M. M., & Lang, P. J. (2003). Early modulation of visual perception by emotional arousal: evidence from steady-state visual evoked brain potentials. *Cognitive, Affective, & Behavioral Neuroscience* , 3(3), 195-206.

Keil, A., Moratti, S., Sabatinelli, D., Bradley, M. M., & Lang, P. J. (2005). Additive effects of emotional content and spatial selective attention on electrocortical facilitation. *Cerebral Cortex* , 15(8), 1187–1197. https://doi.org/10.1093/cercor/bhi001

Junghöfer, M., Elbert, T., Tucker, D. M., & Rockstroh, B. (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology* , *37* (4), 523-532.

Junghöfer, M., Winker, C., Rehbein, M.A., & Sabatinelli, D. (2017). Noninvasive stimulation of the ventromedial prefrontal cortex enhances pleasant scene processing. *Cerebral Cortex* . doi: 10.1093/cercor/bhx073

Keil, A., Smith, J. C., Wangelin, B. C., Sabatinelli, D., Bradley, M. M., & Lang, P. J. (2008). Electrocortical and electrodermal responses covary as a function of emotional arousal: A single-trial analysis.*Psychophysiology* , 45(4), 516-523.

Keil, A., Sabatinelli, D., Ding, M., Lang, P. J., Ihssen, N., & Heim, S. (2009). Re-entrant projections modulate visual cortex in affective perception: Evidence from Granger causality analysis. *Human Brain Mapping* , 30(2), 532–540. https://doi.org/10.1002/hbm.20521

Keil, A., Costa, V. D., Smith, J. C., Sabatinelli, D., McGinnis, E. M., Bradley, M. M., & Lang, P. J. (2012). Tagging cortical networks in emotion: A topographical analysis. *Human Brain Mapping* , 33, 2920-2931.

Kolodyazhniy, V., Kreibig, S. D., Gross, J. J., Roth, W. T., & Wilhelm, F. H. (2011). An affective computing approach to physiological emotion specificity: Toward subject-independent and stimulus-independent classification of film-induced emotions. *Psychophysiology* , 48(7), 908-922. doi: 10.1111/j.1469-8986.2010.01170.x

Koruth, K. J., Lang, A., Potter, R. F., & Bailey, R. L. (2015). A comparative analysis of dynamic and static indicators of parasympathetic and sympathetic nervous system activation during TV viewing. *Communication Methods and Measures* , 9 (1-2), 78–100.

Koukounas, E., & McCabe, M. P. (2001). Emotional responses to filmed violence and the eye blink startle response: A preliminary investigation. *Journal of Interpersonal Violence* , 16, 476–488.

Kreibig, S. D., Wilhelm, F. H., Roth, W. T., & Gross, J. J. (2007). Cardiovascular, electrodermal, and respiratory response patterns to fear and sadness-inducing films. *Psychophysiology* , 44, 787–806.

Lee, K. (2004). Presence, explicated. *Communication Theory* , 14(1), 27–50.

Lin, J. H., & Peng, W. (2015). The contributions of perceived graphic and enactive realism to enjoyment and engagement in active video games. *International Journal of Technology and Human Interaction* , 11(3), 1-16.

Liu, Y., Huang, H., McGinnis-Deweese, M., Keil, A., & Ding, M. (2012). Neural substrate of the late positive potential in emotional processing. *Journal of Neuroscience* , 32(42), 14563–14572. https://doi.org/10.1523/JNEUROSCI.3109-12.2012

Moratti S, Keil A, Stolarova M (2004). Motivated attention in emotional picture processing is reflected by activity modulation in cortical attention networks. *Neuroimage* 21:954-964.

Muller, M. M., Andersen, S. K., & Keil, A. (2008). Time course of competition for visual processing resources between emotional pictures and foreground task. *Cerebral Cortex* , 18(8), 1892-1899.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods* , 8(4), 434–447. https://doi. org/10.1037/1082-989X.8.4.434

Ooms J (2021). magick: Advanced Graphics and Image-Processing in R_. R package version 2.7.3, <https://CRAN.R-project.org/package=magick>.

Palomba D, Sarlo M, Angrilli A, Mini A, Stegagno L (2000) Cardiac responses associated with affective processing of unpleasant film stimuli. *Int J Psychophysiol* 36: 47–57.

Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Hochenberger, R., Sogo, H., Kastman, E., Lindelov, J.K., 2019. PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods 51* (1), 195–203.

Pessoa, L. (2017). A network model of the emotional brain. *Trends in cognitive sciences* , 21(5), 357-371.

Philippot, P. (1993) Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition and Emotion* , 7:2, 171-193.

Prentice, Deborah A. and Richard J. Gerrig. (1999). Exploring the boundary between fiction and reality. *Dual-Process Theories in Social Psychology* . Ed. Shelly Chaiken and Yaacov Trope. New York: Guilford Press: 529–46.

Sabatinelli, D. McTeague, L.M., Dhamala, M., Frank, D.W., Wanger, T.J., & Adhikari, B. (2015). Reduced medial prefrontal-subcortical modulation in dysphoria: Granger causality analyses of rapid functional MRI. *Brain Connectivity* , 5, 1-9. doi:10.1089/brain.2013.0186

Sabatinelli, D. & Frank, D.W. (2019). Assessing the primacy of human amygdala-inferotemporal emotional scene discrimination with rapid whole-brain fMRI. *Neuroscience* . doi: 10.1016/j.neuroscience.2019.03.001.

Samson, A. C., Kreibig, S. D., Soderstrom, B., Wade, A. A., & Gross, J. J. (2016). Eliciting positive, negative and mixed emotional states: A film library for affective scientists. *Cognition and Emotion* , 30(5), 827-856.

Schindler, S., & Bublatzky, F. (2020). Attention and emotion: An integrative review of emotional face processing as a function of attention. *Cortex* , 130, 362–386. https://doi.org/10.1016/j.cortex.2020.06.010

Schupp, H. T., & Kirmse, U. M. (2021). Case-by-case: Emotional stimulus significance and the modulation of the EPN and LPP. *Psychophysiology* , 58(4), 1–13. https://doi.org/10.1111/psyp.13766

Simons, R. F., Detenber, B. H., Roedema, T. M., & Reiss, J. E. (1999). Emotion processing in three systems: The medium and the message. *Psychophysiology* , 36, 619–627.

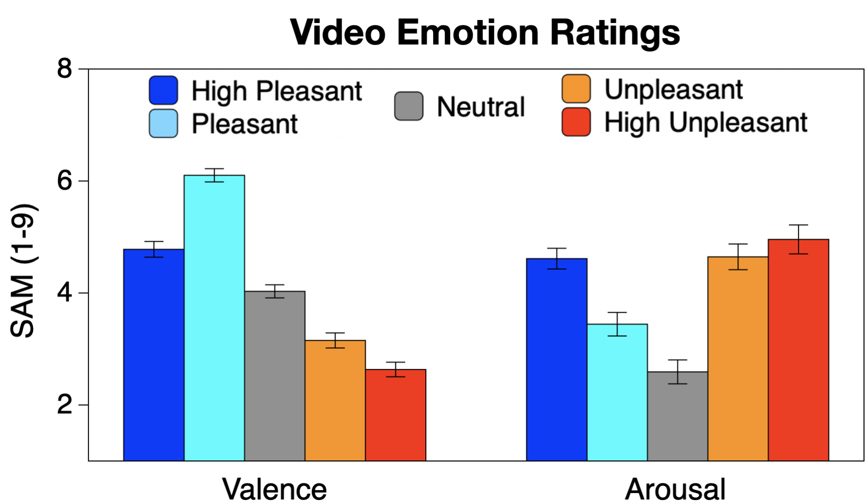
Simons, R. F., Detenber, B. H., Reiss, J. E., & Shults, C. W. (2000). Image motion and context: A between- and within-subject comparison. *Psychophysiology* , 37, 706–710.

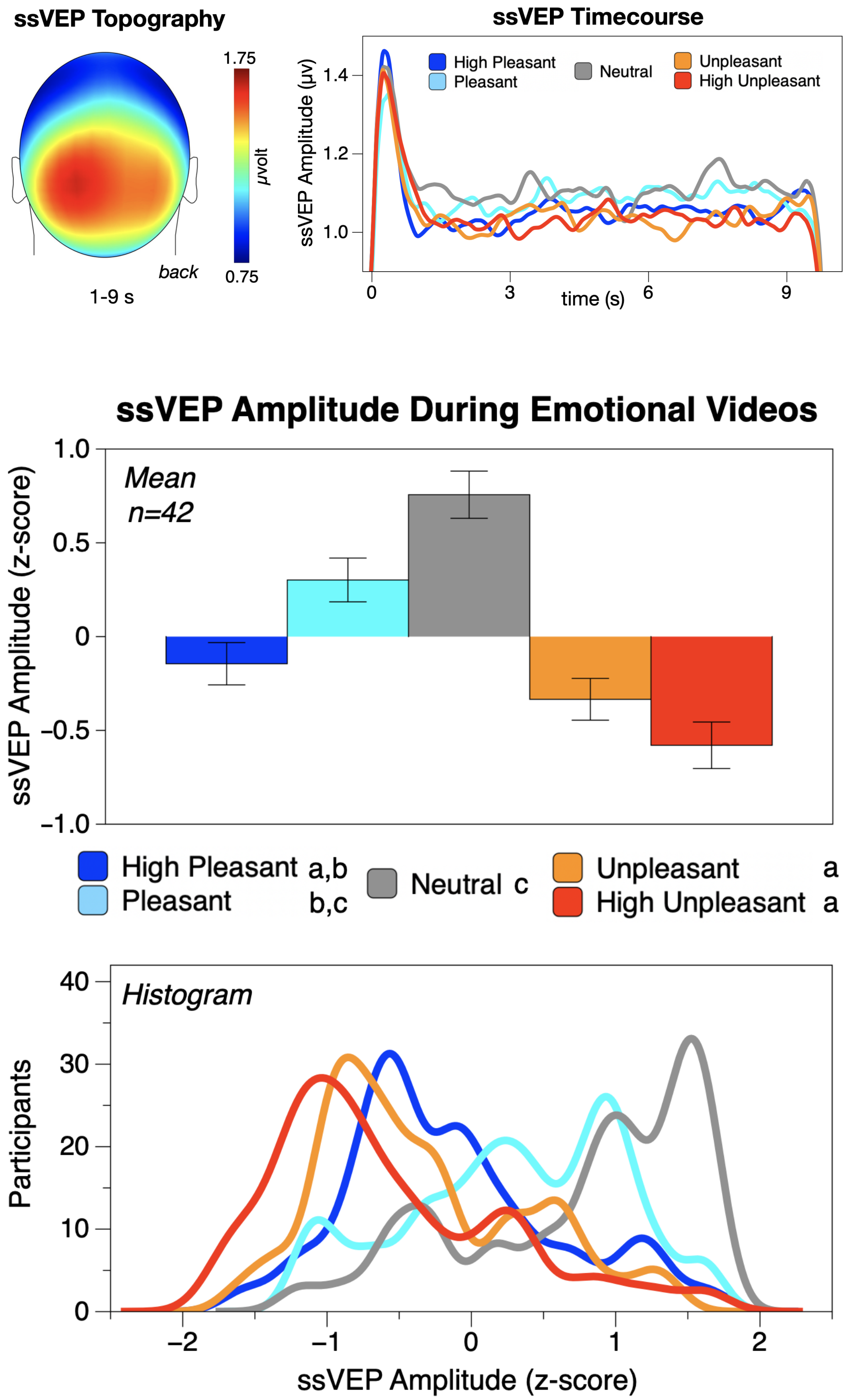
Simons, R. F., Detenber, B. H., Cuthbert, B. N., Schwartz, D. D., & Reiss, J. E. (2003). Attention to television: Alpha power and its relationship to image motion and emotional content. *Media Psychology* , 5, 283–301.

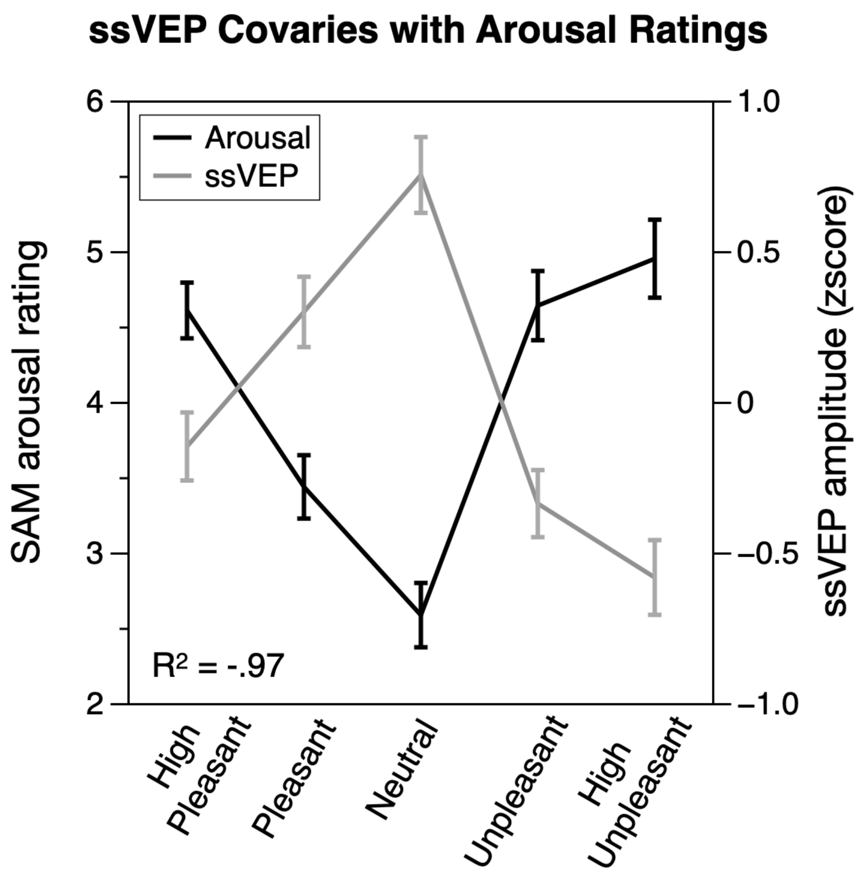
Spreckelmeyer K. N., Kutas M., Urbach T. P., Altenmuller E., Munte T. F. (2006). Combined perception of emotion in pictures and musical sounds. *Brain Research* , 1070, 160-170.

Stegmann, Y., Andreatta, M., Pauli, P., Keil, A., & Weiser, M. J. (2023). Investigating sustained attention in contextual threat using steady-state VEPs evoked by flickering video stimuli. *Psychophysiology* , 60:e14229, doi.org/10.1111/psyp.14229

Wieser, M. J., Miskovic, V., & Keil, A. (2016). Steady-state visual evoked potentials as a research tool in social affective neuroscience. *Psychophysiology* , 53(12), 1763–1775. https://doi.org/10.1111/ psyp.12768




ssVEP Topography
1.75
µvolt
0.75
back
1-9 s
ssVEP Timecourse
High Pleasant
Pleasant
Neutral
Unpleasant
High Unpleasant
ssVEP Amplitude (µv)
1.4
1.2
1.0
0
3
6
9
time (s)
ssVEP Amplitude During Emotional Videos
Mean
n=42
ssVEP Amplitude (z-score)
1.0
0.5
0
−0.5
−1.0
High Pleasant a,b
Pleasant b,c
Neutral c
Unpleasant a
High Unpleasant a
Histogram
Participants
40
30
20
10
0
−2
−1
0
1
2
ssVEP Amplitude (z-score)



Hosted file

Table1_Videos.docx available at https://authorea.com/users/646514/articles/658393-moving-toward-reality-electrocortical-reactivity-to-naturalistic-multimodal-emotional-videos

Hosted file

Table2_Block.docx available at https://authorea.com/users/646514/articles/658393-moving-toward-reality-electrocortical-reactivity-to-naturalistic-multimodal-emotional-videos

Hosted file

Table3_Corr.docx available at https://authorea.com/users/646514/articles/658393-moving-toward-reality-electrocortical-reactivity-to-naturalistic-multimodal-emotional-videos