DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-126

Physiological capture of augmented viewing states: objective measures of high-dynamic-range and wide-color-gamut viewing experiences

Dan Darcy, Evan Gitterman, Alex Brandmeyer, Scott Daly and Poppy Crum; Dolby Laboratories, Inc.; San Francisco, CA

Abstract

The capability for improved image and video reproduction quality is notably growing. For example, augmentation of resolution and dynamic range allows for a markedly transformed viewer experience. These can influence a viewer's experience of content in a manner not captured by standardized methods of subjective quality assessment. Changes in viewer experiences may include elevated arousal, enhanced emotional impact, and increased engagement. Here we describe objective methods and data metrics allowing for the assessment of individual responses to high-dynamic-range (HDR) and wide-color-gamut (WCG) motion imagery using electroencephalography (EEG) measurements.

All test content was mapped from a HDR/WCG source, with the HDR/WCG content mapped to the parameters of the stimulus display (0.005-1000 cd/m^2 dynamic range, DCI P3 color space). Comparison baseline content was mapped to the parameters of a standard consumer TV (0.05-100 cd/m^2 dynamic range, Rec. 709 color space). The difference between viewing the HDR/WCG content and baseline content was captured using EEG to probe modulation of visual cortical drive, elicited with a fixed-frequency reversing checkerboard stimulus. These metrics are combined with a broader set of measures to collectively quantify content- and dynamic range-dependent impacts of engagement and attentional processing.

Introduction

Motivations for physiological testing of motion imagery

There is a long history of studies that assess changes in image quality[1] resulting from display design parameters and image processing algorithms [1]. The majority of this work has considered the assessment of still images and made use of behavioral forced-choice testing methodologies that include side-by-side and sequential comparisons. Subjective metrics of image quality (such as threshold detectability of changes in perceptual parameters, suprathreshold differences in appearance, and viewer preferences) can inform algorithmic design development or targeted business strategies. The revival of 3D displays and popularity of streaming video have expanded the subject of image quality into the broader topic known as quality of experience (QoE) [2, 3]. Here we describe a methodology (and supporting results) for QoE assessment of high-dynamic-range (HDR) and wide-color-gamut (WCG) motion imagery that aims to mitigate many of the problems typically associated with QoE assessment and the experience of dynamic content.

Limitations of side-by-side comparisons

While many studies of still image quality exist, QoE for motion imagery is not well understood. In addition to difficulties relating to content selection and acquisition, there are many experimental challenges that must be solved. For example, popular still-image testing approaches using side-by-side comparisons have been shown to produce significantly lower response variability than those using sequential comparisons. This has been shown for reporting of image features such as color and tone assessment [4]. However, side-by-side empirical methodologies are not optimal for assessment of motion imagery. Limitations in the viewer's time-dependent foveation ability paired with inherent frame dependency of image distortions introduce notable problems in producing directly comparative data. By the time the viewer notices a distortion or improvement in one image and then foveates to the paired side-by-side image, one or more frame cycles have elapsed, and the viewer is unable to make a direct comparison of corresponding frames. To address this, studies of motion imagery have employed tools for side-by-side methodological design (e.g. 'butterfly' or 'mirrored' comparisons). Nonetheless, these methods often introduce additional problems such as nausea and discomfort, resulting from general changes in the overall looming and receding optical flow. In comparison, single-screen real-time objective QoE assessment methodologies as described in this paper avoid these well-known limitations in data acquisition.

Contextual exposure can influence QoE

Another notable issue in motion imagery testing is the impact of content narrative and expertise on the viewer response. Typical motion imagery content includes conveyance of an experiential, time-dependent, contextually relevant story. Experts and professionals working in the production of movies and video are attuned to small differences in image quality that non-expert viewers may not be trained to focus on. To them, the various subtleties that they work to achieve as artistic craft may not be describable or explicitly noticed by non-experts, but will have an intended effect on the viewer's emotion, engagement level, immersion, or 'suspension of disbelief'. According to this line of thinking, improvements that viewers are unable to describe may still have an effect on the overall impression of the video's narrative, by assuming non-actionable or subconscious processes and biases.

[1] In this article we will use 'image quality' as a superset of still-image quality and motion imagery quality.

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-126

Undesired effects of comparison and rating tasks

Additionally, time-dependent task interference can confound image quality testing. The temporally sequential act of comparing image differences followed by a motor response can influence the data in an undesirable manner. For example, one artifact introduced by time-dependent exposure is accommodative hysteresis [5] and load when the displayed imagery and the response interface are presented at different distances. This can be mitigated by placing the visual response interface within the same display, as done in the SAMVIQ method [6], or by placing it at the same distance when it is on a separate display [7]. Nonetheless, even if this problem is avoided, the task of image comparison for motion imagery deviates notably from natural viewing. For motion imagery, iconic memory is a key limitation in side-by-side comparisons, while visual short-term memory is a limitation in magnitude estimation for both side-by-side and sequential viewing. Long-term memory becomes an issue with attempts to compare overall quality using categorical ranking approaches. Rating (Likert) scales with descriptive responses (e.g., 'excellent', 'good', 'fair', 'poor', and 'bad') are frequently used and suffer from these unsolicited influences. Additional deviations from natural viewing are also expected to occur in comparison or rating tasks: eye movements are altered, attention is taken away from the narrative, and the viewer is necessarily placed in an analytical state, all of which serve to lessen the emotional and visceral impact of the content and to influence state-dependent responses.

In summary, there is an immediate need for alternative testing methodologies in QoE assessment of motion imagery content. New developments in professional and consumer technology enable significant changes in both the dynamic range and color gamut of motion imagery. The impact of these technological developments on the user experience of content has to date had limited study, despite the numerous venues for content consumption. Here, we discuss single-screen real-time methods for collecting objective physiological metrics pertaining to QoE. Such methodologies circumvent the many limitations associated with dual-screen viewing environments and allow for more direct study of the content and technologies underlying viewer experience. For example, what type of content is most impacted by HDR and WCG? Is there an interaction with these technologies and the content that influences the emotional impact and arousal of the viewer, depending on the length of the temporal viewing window?

Physiological approaches to image quality testing

One issue with traditional assessment tools (such as the rating scales previously mentioned) is that results can be influenced by subjective interpretations of the quantity or quality of the reported response. As an alternative, physiological measurements recorded from an individual are constrained by biological properties that are not subjective. One such measure is the electroencephalogram (EEG), which measures small electrical currents on the scalp generated by populations of neurons in the central nervous system. EEG has been used extensively to study the basic neural mechanisms underlying both perceptual and cognitive processes. We set out to explore EEG responses under viewing conditions that we expect will elicit notable changes in physiological state, in order to understand how the physiological metrics may augment traditional approaches to assessment.

There is some history of physiological monitoring for motion imagery in the 20th century [8], but the majority of studies were done for marketing purposes without corresponding technical publications. There has been a recent resurgence due to the increasing ubiquity of non-invasive physiological sensors, as well as advances in neuroscience. EEG has been used in countless studies in which visual stimuli are used to evoke event-related potentials (ERPs). For example, Bentin [9] used the N170 component of the ERP to study face-recognition during still-image viewing. Still images have also been used to manipulate the viewer's emotion, which can then be classified into affective states using facial thermal imaging [10] among other techniques.

Recently, EEG has been increasingly used to measure various aspects of image quality. Palmateer proposed using the steady-state visual evoked potential (SSVEP) and QEEG (quantitative EEG) techniques in assessing display quality and tested the concept for display rate flicker [11]. Several EEG studies have used evoked potentials to characterize still image compression [12,13] and video compression [14,15]. Depth-related aspects of perception have been studied for stereoscopic (3D) displays: in comparison to 2D [16], as well as characterizing crosstalk in 3D displays [17], both using EEG to identify differences in ERP topography. Video frame rate differences have been studied with EEG frequency power spectra using Canberra distances [18]. More basic research has used ERPs and eye movements to study visual target search [19]. Most of these studies, however, were focused on perceptual visibility and did not pursue higher-level cognitive processes.

Recently, studies aimed at entertainment media have used physiological monitoring to assess higher-level aspects of perception, rather than visibility. These include the use of functional magnetic resonance imaging (fMRI) to understand audience preferences for television content [20] and to characterize cortical activity while reading literature with different levels of engagement [21], as well as the use of electrocorticography (ECoG) to understand spatial attention [22].

Our goal is to develop methodologies and data analyses for EEG in an image quality application that assesses viewer engagement. The present study is aimed at narrative video content, and we seek to determine if new improvements in imaging can cause higher levels of viewer engagement, regardless of whether the improvement is explicitly noticed by the viewer. A secondary goal is to emulate natural viewing circumstances and avoid excessive repetition of stimuli while maintaining a good signal-to-noise ratio (SNR).

Several properties of the EEG signal can be used to study attention, a key component of engagement. A large body of scientific literature studying the basic neurophysiological mechanisms underlying the allocation and dynamics of attention in and across different sensory modalities has established metrics that can be derived from EEG data. The present study will focus on three of these measures: EEG power in the 8-12 Hz range (the 'alpha' band) measured using scalp electrodes in the posterior region of the head, amplitude of SSVEPs elicited under different stimulus conditions, and a time-locked component of the ERP known as the 'P300 response'.

Alpha-band activity was first described in the pioneering EEG research done by Hans Berger in 1924, and is one of the strongest,

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-126

most readily observed components of the EEG signal. Current theories propose that alpha activity reflects the functional inhibition of sensory processing, such that increased alpha power reflects reduced sensory processing [23]. In the visual domain, increased alpha power has been shown to predict reduced performance in signal detection tasks [24]. Thus, it can be predicted that there will be a negative correlation between EEG signal power in the alpha band measured over visual cortex and the level of visual engagement.

SSVEPs are a type of evoked response elicited by visual stimuli which repeat at a characteristic frequency. At each visual stimulus onset, a visual evoked response is observed, such that the individual responses overlap with one another to produce a time-locked oscillatory pattern in the data. This oscillatory response can be observed in the frequency-domain representation of the EEG data as a spectral peak at the stimulus' characteristic frequency [25]. Importantly, the amplitude of the SSVEP is modulated by visual attention [25], and can potentially index differences in visual engagement with the content immediately preceding the eliciting stimulus.

The P300 component of the ERP is another well-established marker of attention that is elicited by 'pop-out' sensory stimuli, either due to their salience or context-dependent value [26]. It is typically observed as a relatively high amplitude (> 5 μV) positive peak in the ERP, reaching maximum amplitude between 300-600 ms after the appearance of the eliciting stimulus at central electrode locations near the crown of the head [27]. Crucially, its amplitude is known to vary depending on the allocation of attentional resources to secondary tasks unrelated to the eliciting stimulus [28, 29]. As such, a salient 'pop-out' stimulus can be used to probe the depth of attentional allocation in a secondary task.

High Dynamic Range and Wide Color Gamut

The particular image quality improvement being studied here is the result of a combination of high dynamic range (HDR) and wide color gamut (WCG), as well as the necessary bit-depth increase. HDR and WCG involve major improvements in the image capture, signal formatting, and display capabilities. By combining the luminance range description (HDR) with the traditional 2D color gamut description, a color solid results, and the size of the color solid is described by the overall color volume. Color volume can be used to describe the overall color and luminance capability. This is illustrated in Figure 1, using a vertical luminance axis. The particular parameters of improvement that we will study are shown in Figure 2, where the 'SDR' (standard dynamic range and gamut) volume is determined by Rec. 709 primaries, a maximum luminance of 100 cd/m^2, and a minimum luminance (black level) of 0.05 cd/m^2, for a total dynamic range of 2,000:1. The improved color volume (referred to in this paper simply as 'HDR', even though the color gamut has also improved) has the digital cinema primaries (referred to as P3), a maximum luminance of 1000 cd/m^2, and a minimum luminance of 0.005 cd/m^2, for a total dynamic range of 200,000:1.

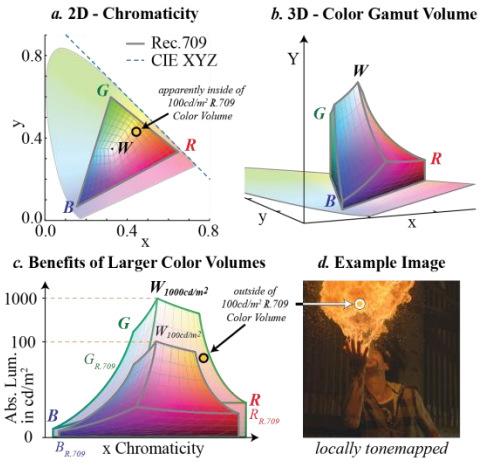

Figure 1. *a) Traditional 2D color gamut. b) A 3D color volume. c) New capability achieved by HDR. d) Example HDR image.*

The visible advantages of HDR and WCG are numerous, and are described in several books [30, 31, 32]. Some of the more interesting advantages include better renditions of specular reflections and emissive colors, and the ability to portray both indoor and outdoor scenes in a single image (looking outside from within a room). In addition, there have been descriptions of the differences in terms of visceral reactions by the consumer electronics and professional imaging press [33]. Recent demos have used maximum luminance levels of 4000 to 8000 cd/m^2 and achieved easily visible dramatic effects. However, the level of HDR and color volume tested here is on the lower end of HDR improvement [4]. For these luminance levels, the differences are obvious in side-by-side comparisons, but in sequential comparisons, it can be difficult for non-experts to notice the difference.

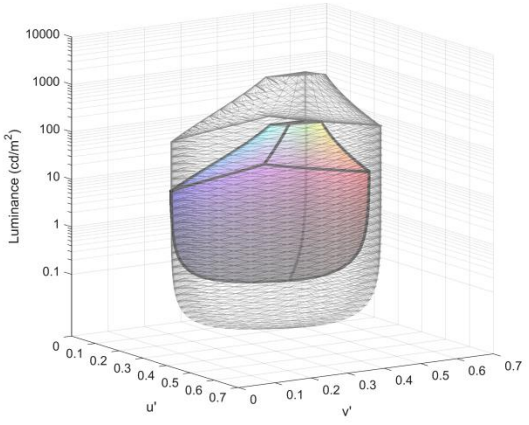

Figure 2. *Color volumes used in this experiment. Inner solid is max. lum. = 100, min. lum. = 0.05, and color gamut = Rec. 709 (referred to as simply 'SDR' in this paper). Outer mesh is max. lum. = 1000, min. lum. = 0.005, and color gamut = P3 (referred to as simply 'HDR' in this paper). All luminance units in cd/m^2.*

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-126

Methods

Subjects and Data Collection

Nine subjects participated in the study (7 right-handed, 4 females and 5 males, mean age: 33 years, range: 21-47). All subjects reported normal hearing and normal or corrected to normal vision, with the exception of one subject with partial color-blindness. Subjects were recruited from a pool of external listeners employed part-time by Dolby and were paid hourly for their participation in the experiment. All subjects had no history of photosensitive epilepsy and were shown a preview of the flashing checkerboard stimulus to gauge visual discomfort before deciding to proceed with the experiment.

EEG was recorded from 32 active dry-sensor g.Sahara electrodes with a g.Nautilus wireless headset/amplifier (g.tec medical engineering GmbH, Graz, Austria). Electrodes were positioned according to the modified 10-20 system at the following locations: Oz, PO7, PO3, PO4, PO8, P7, P3, Pz, P4, P8, CP5, CP1, CP2, CP6, T7, C3, Cz, C4, T8, FC5, FC1, FC2, FC6, F7, F3, Fz, F4, F8, AF3, AF4, FP1, FP2. Reference and ground electrodes were placed on the right and left earlobes, respectively. Data were recorded with a 250Hz sampling rate and the sensitivity of the active system set to +/- 750mV.

Stimuli and Procedure

Testing was done in the Physiology Lab at Dolby's 1275 Market headquarters, with room lights turned off. Visual stimuli were presented using Adobe Speedgrade (Adobe Systems, San Jose, CA) on a 42-inch Dolby Professional Reference Monitor PRM-4200 that was modified to reach a maximum luminance of 1000 cd/m^2 with a minimum luminance of 0.005 cd/m^2 and a P3 color gamut. Viewing distance was approximately three picture heights from the 42-inch-diagonal display image, set by chair position without a chin rest in order to keep viewing as natural as possible. Due to the limitations of Speedgrade, each movie clip's audio was presented as a stereo downmix, on ATC SCM11 speakers (ATC Loudspeakers, Gloucestershire, UK).

Stimuli consisted of two 22-minute segments, each containing three movie clips. The clips were excerpts from action movies, each approximately 6 minutes in duration. These clips were selected due to their visually engaging action sequences, which featured bright explosions, dark detail, and imagery that took advantage of HDR and WCG. An additional criterion was that the clips were engaging without the context of the entire movie, since most subjects had not seen the movies before. The clips had been originally graded for presentation on a 4000-nit high-end reference display, and an internal display-mapping algorithm was used to adapt the clips for the present experimental parameters. The mapping algorithm has been used in other psychophysical quality studies involving HDR [4,34]. Rather than simply compress the luminance linearly in a luminance or gamma domain, it tends to preserve the mean level while more severely compressing the highlights. Each clip was shown twice, once in an 'HDR' condition, which used the full dynamic range of the modified PRM-4200 display, and once in an 'SDR' condition, which emulated a typical consumer display. The display matching parameters for each condition are shown below, and illustrated in Figure 2.

Display Specifications for Test Clip Conditions

	HDR Condition	SDR Condition
Maximum brightness	1000 cd/m^2	100 cd/m^2
Minimum brightness	0.005 cd/m^2	0.05 cd/m^2
Color space	DCI P3	Rec. 709

Both HDR and SDR conditions were presented on the same PRM-4200 display using the Dolby PQ electro-optical transfer function (EOTF), with a resolution of 1920 by 1080 pixels, 12 RGB bits/pixel, and at 24 frames per second. Clip order was fixed within each of the two sessions, while condition order alternated between HDR and SDR throughout both sessions to avoid habituation effects, with the starting condition randomized across subjects.

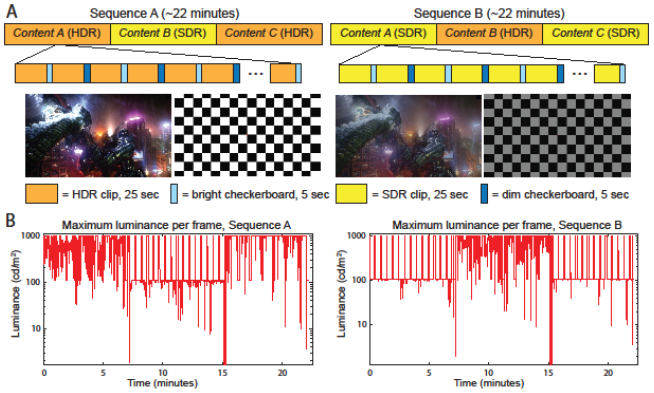

Figure 3. *Experimental design. A) Two video sequences were presented to each subject in counterbalanced order. Each contained interleaved HDR and SDR versions of 3 cinematic clips. 'Bright' (light blue, mapped to HDR range) and 'dim' (dark blue, mapped to SDR range) checkerboard stimuli were alternately presented at 25-second intervals throughout both sequences. B) Peak luminance over time for each sequence.*

All clips had rapidly flashing checkerboard segments interspersed throughout, designed to serve as a 'probe' stimulus in which the effects of the HDR and SDR contexts could be measured indirectly by analyzing the evoked potentials. Flashing checkerboard stimuli are known to elicit SSVEP responses, which are desirable for their high SNR (providing many stimulus onsets in a short period of time). These segments consisted of a full-screen black-and-white 16-by-9 checkerboard, which inverted its colors at 24 frames per second. A constant red dot was overlaid on the center of the checkerboard, and subjects were instructed to fixate on the dot during checkerboard segments. The checkerboard design was chosen in order to elicit maximal SSVEP amplitudes, and the 24 Hz frequency was chosen in order to match the movie clips and ensure reliable presentation. Checkerboard segments were always 5 seconds in duration, and were interspersed

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-126

following each 25-second period of the movie clips, after which the movie clip would resume. No audio was presented during the checkerboard segments. The checkerboard segments were presented in two conditions, 'bright' and 'dim', with maximum and minimum luminance values identical to the HDR and SDR movie clip conditions, respectively. The bright and dim checkerboard segments were presented in alternation throughout the experiment to control for the contrast difference when presented in the context of the two clip conditions.

EEG Data Preprocessing and Analysis

All EEG data processing and analysis was conducted using the open source FieldTrip toolbox [35] in MATLAB in conjunction with custom scripts. Raw EEG signals were high-pass filtered with a finite impulse response filter at 0.5 Hz, and low-pass filtered using a Butterworth filter at 40 Hz. Optical trigger signals in the unprocessed EEG recordings were then used to segment the data into 1-second epochs, which were grouped into sets representing the different visual conditions: SDR viewing, HDR viewing, and 'bright' and 'dim' checkerboard stimuli.

An independent component analysis (ICA) was then performed using the 'runica' method [36, 37], and the data were visually inspected to identify components corresponding to muscle movement and eye-blink artifacts. These artifacts typically contain transient activity orders of magnitude larger than the neural signals of interest, and the corresponding components can be readily identified based on both their scalp topography and signal statistics. Removal of such artifacts through ICA has the benefit of preserving data epochs that would otherwise be rejected in subsequent artifact removal steps based on the amplitude of the measured signals. Following the identification and removal of artifactual components, the analysis proceeded in the EEG channel space.

Two additional steps were taken to clean the data. First, an iterative procedure was used to identify individual EEG channels showing poor connectivity. This is typically characterized by an overall signal power that is orders of magnitude larger than the ongoing EEG signal. EEG recordings made using dry-electrode systems will typically contain multiple channels with poor connectivity, as there is no additional conductive medium applied to the scalp as part of the cap fitting procedure. In each iteration, the mean signal power across all data epochs was calculated for each EEG channel. Then, based on the median value, any channels showing overall signal power more than 2.5 standard deviations above the median were marked as bad, and were repaired using the mean of all immediate neighbors not also marked as bad. This process was repeated until no channels were marked as outliers, or until 10 iterations had been completed. One participant's data was excluded from further analysis due to excessive mean signal power ($> 10^5$ μV^2) following bad channel repair. For the remaining participants, an average of 11.1 channels were repaired.

Second, signal amplitudes at all 32 channels in individual data epochs were checked against a threshold of ± 150 μV, such that any epochs exceeding the threshold were marked as bad and excluded from any subsequent analyses. On average, 79/2648 data epochs, or approximately 3% of the collected data, were marked as containing artifacts.

Additional 6-second long data epochs corresponding to a checkerboard stimulus and the data in the second immediately preceding it were obtained from the raw data and used to calculate the ERP for the onset of the checkerboard stimulus, as well as its time-frequency representation. From a total of 90 epochs per subject, an average of 9.1 epochs, or 10.1% of the available data, were excluded from subsequent analyses that made use of these longer data epochs.

The data analysis proceeded along two principal lines: 1) A frequency-domain analysis of the EEG power spectra and topographies across the different experimental conditions, and 2) time-domain analysis of the ERPs elicited by the onset of the checkerboard stimuli. The frequency domain analysis was performed using a multi-taper method on individual one-second data epochs in the 1-65 Hz range with a resolution of 1 Hz and a spectral smoothing parameter of ±1 Hz. The ERP analyses made use of the six-second data epochs, which were baseline corrected using the mean signal value at each electrode in the 200 ms prior to the onset of the checkerboard, prior to averaging. An additional time-frequency analysis was performed using the six-second data epochs as part of the analysis of the neural response to the checkerboard stimulus. This was done using a wavelet-based method in the range between 18-30 Hz in .33 Hz steps. Individual wavelets representing 7 cycles at each of the target frequencies were convolved with the time-domain data to obtain an estimate of the time-varying signal power at that frequency. Following the wavelet analysis, individual data were baseline corrected with respect to the time interval from 600 ms to 200 ms prior to the onset of the checkerboard stimuli, such that signal power at each time-frequency point was represented as the percent signal power change relative to the baseline value for the corresponding frequency bin.

Results

Alpha Power Modulation during HDR and SDR Viewing Conditions

The first analyses of the EEG focused on the dynamics of alpha-band power during HDR and SDR content viewing. For this analysis, the continuous EEG data representing the time course of the HDR and SDR content viewing was reassembled from the individually preprocessed epochs. Epochs corresponding to the checkerboard stimuli were not included, while epochs containing artifacts were replaced with randomly selected data from the same visual condition in order to preserve the overall time course of the individual sessions. This time course was band-pass filtered between 9-12 Hz using a Butterworth filter, and then squared to obtain a power estimate. Finally, the time course was smoothed using a low-pass filter at 0.2 Hz.

Next, data from both the HDR and SDR portions of the experiment were averaged at the individual and group levels in order to estimate the overall grand-average scalp topography of peak alpha activity (Figure 4a, left panel). A clear peak over visual and parietal cortices was observed, corresponding to the typical EEG scalp distributions of alpha power [38]. This topography was used as a spatial filter on individual subjects' continuous data to derive a 'virtual electrode' representing the summary time course of alpha activity in HDR and SDR viewing conditions. From these

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-126

time courses, the mean alpha power in each condition was estimated (Figure 4a, right panel). Relatively higher levels of alpha power, indicating reduced visual engagement, were observed in the SDR condition. However, this difference failed to reach statistical significance.

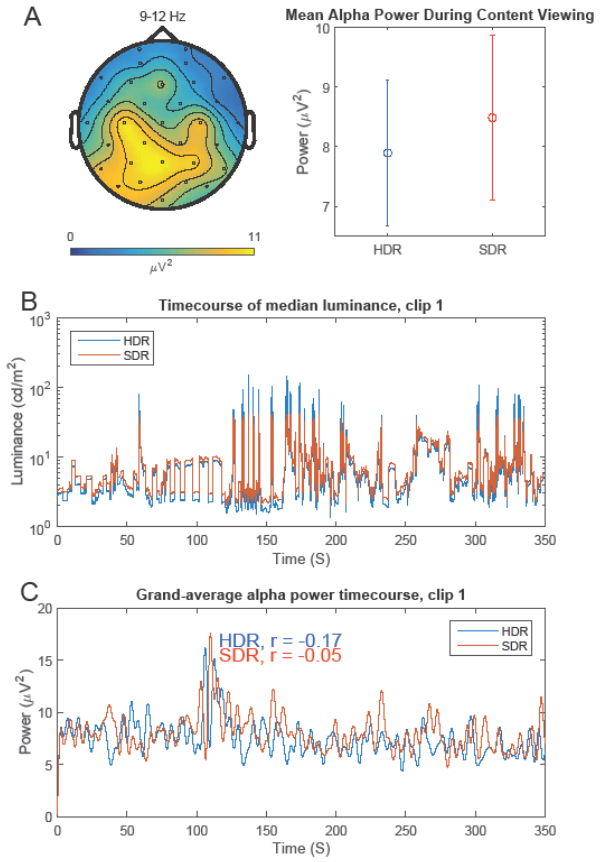

Figure 4. *Alpha-band power dynamics during matched HDR and SDR content viewing. A) Grand-average scalp topography and mean power in HDR and SDR conditions across subjects. B) Median luminance levels in HDR and SDR versions of the first cinematic clip. C) Grand-average time course of alpha power for the first cinematic clip. Correlations of the alpha power time courses with the luminance time course in panel B are indicated.*

Subsequent analysis of the alpha power dynamics focused on their relationship with the luminance changes in the movie content. The median luminance levels for both of the visual sequences were obtained on a frame-by-frame basis. The data for the first of the three cinematic content excerpts are shown in Figure 4b, and are illustrative of the increased dynamic range in the HDR condition. This plot shows that HDR images are not simply brighter overall – they can be darker as well. The benefit of HDR is that the overall range of luminance is increased, both spatially within an image, and temporally across scenes or frames (as shown in Figure 4b). A correlational approach was then taken to relate changes in the image dynamics over time to changes in visual processing, as indexed by the alpha power time course. The grand-average time courses across subjects for both the HDR and SDR versions of the first cinematic content excerpt were resampled to 24 Hz to match the frame rate of the image data. The time courses for both the HDR and SDR conditions are presented in Figure 4c.

The present hypothesis regarding the relationship between image dynamics and visual engagement would predict a negative relationship between median luminance levels and alpha power, as alpha power will be reduced as the level of visual processing increases. This is indeed what is observed, with a significant negative correlation ($p < .001$) of median luminance levels and alpha power across time for both the HDR and SDR conditions. The observed correlation coefficient was substantially more negative for the HDR condition ($r = -0.17$) than for the SDR condition ($r = -0.05$), indicating that visual processing is modulated by the image dynamics of the cinematic content to a greater extent when presented in HDR.

Evoked Responses to Checkerboard Stimuli

Grand-averaged ERPs at the onset of the checkerboard stimuli across visual conditions are presented in Figure 5. Figure 5a presents the average scalp topography of the response at between 90-95 ms following the onset of the checkerboard stimulus in the left panel, together with the average signal trace from electrode Oz for the periods immediately preceding and following stimulus onset. The Oz electrode is centered over the foveal region representation within primary visual regions (striate cortex). A stereotypical visual evoked response was observed, with clear negative and positive peaks at approximately 70 and 100 ms, respectively, corresponding to the N70 (also known as C1) and P100 components generated in striate and extrastriate cortices [39]. Additionally, a clear oscillatory response can be observed, corresponding to the steady-state visual evoked potential (SSVEP) elicited by inversions of the checkerboard stimulus (with an interstimulus interval of ~42 ms, corresponding to 24 frames per second).

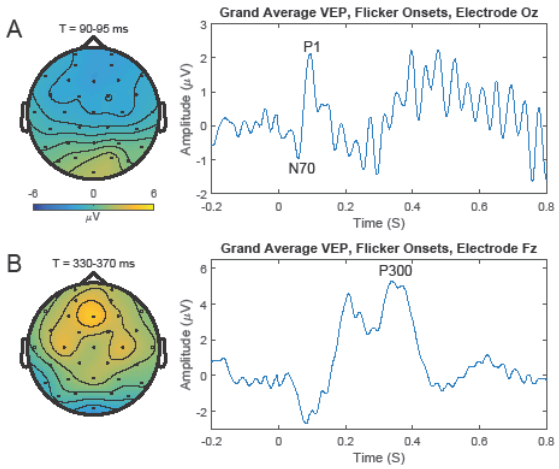

Figure 5. *Event-related potentials during checkerboard stimulus. A) Scalp topography of P100 response and grand-average waveform at midline occipital electrode (Oz). B) Scalp topography of P300 response and grand-averaged waveform at midline frontal electrode (Fz).*

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-126

The left panel of Figure 5b presents the average scalp topography between 330-370 ms following stimulus onset. The corresponding signal trace from electrode Fz is shown in the right panel, where a clear positive peak corresponding to the P300 response can be observed (Fz is located immediately anterior to the crown of the head). This response provides an index of attentional capture of the checkerboard stimulus, which itself is a function of the preceding levels of attention directed towards the cinematic content. Variability of these responses across experimental conditions is further explored in the following sections.

Frequency-domain analyses of the responses elicited by the checkerboard stimulus are presented in Figure 6. Figure 6a presents the scalp topography of the power spectrum at 24 Hz, with a clear peak observed over visual cortices at electrode Oz. Figure 6b presents the power spectra obtained across the different experimental conditions at electrode Oz. An average peak with a magnitude of approximately 2 μV^2 was observed in all conditions. Statistical tests revealed no differences in power levels for bright vs. dim checkerboard stimuli, nor were any differences observed in responses following HDR vs. SDR cinematic content. Figure 6c presents the average time-frequency representation of the response across the entire 5-second duration of the checkerboard stimuli. A clear increase in power centered at 24 Hz can be observed relative to the baseline period prior to the onset of the stimulus. Thus, it can be concluded that checkerboard stimulus reliably elicited an SSVEP, but no overall differences in the magnitude of this response were observed across the experimental conditions.

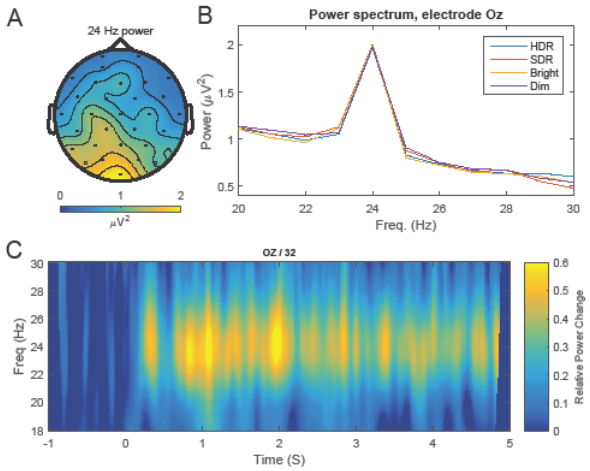

Figure 6. *Steady-state visual evoked potential (SSVEP) at 24 Hz during checkerboard stimulus presentation. a) Scalp topography of 24 Hz response during middle portion of checkerboard stimulus (1-4 seconds). b) Power spectra of SSVEP during different experimental conditions. c) Grand-average time-frequency representation across conditions.*

Relationship between visual engagement and P300 response amplitude

Individual subjects' P300 responses in the different experimental conditions were captured by first spatially filtering the data using the mean individual response topographies in the 330-370 ms range. The amplitude of the largest positive peak in the filtered signal between 300-500 ms in each visual condition was submitted to further analysis. Results of these analyses are presented in Figure 7.

The first analysis focused on differences in the P300 response amplitudes elicited by checkerboard stimuli following either HDR or SDR cinematic content excerpts. In Figure 7a, significantly larger response amplitudes were observed following SDR content relative to HDR content ($p < 0.05$, non-parametric permutation test). This suggests that the attentional capture effect of the checkerboard stimulus was enhanced following SDR content viewing.

One possibility is that the relative difference in brightness between the two checkerboard stimuli enhanced the P300 response following the SDR condition due to lower overall luminance levels during content viewing. To assess this, two additional analyses were carried out, and are presented in Figures 7b-c. The first compared overall P300 response amplitudes for bright and dim checkerboard stimuli. Mean amplitudes between the two conditions were not significantly different, indicating that attentional capture was not principally modulated by stimulus luminance.

The second analysis looked specifically at the P300 response amplitudes to dim checkerboard stimuli following SDR content relative to bright checkerboard stimuli following HDR content. This served as an additional control for the results presented in Figure 7a, but focusing only on responses elicited by checkerboard stimuli matched to the dynamic range of the preceding content. The same general pattern as the initial analysis was observed, despite the reduced amount of data available for calculating the individual ERPs. This result reached a significance of p=0.06.

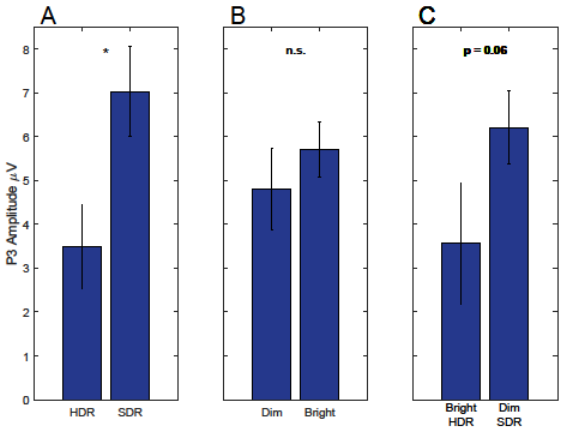

Figure 7. *Comparison of P300 amplitudes across experimental conditions. A) Mean P300 response amplitudes to checkerboard stimuli following HDR or SDR content. B) Mean P300 response amplitudes to dim and bright checkerboard stimuli. C) Mean P300 response amplitudes to checkerboard stimuli matched to the preceding HDR or SDR content. Statistical comparisons were performed using a non-parametric permutation test. Asterisk indicates a significant result at the p < 0.05 level.*

Taken together, these results indicate that the checkerboard stimuli more effectively captured attention following SDR vs. HDR content, independent of the relative brightness of the checkerboard stimulus to the preceding content. It has been

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-126

previously shown that increased attentional engagement in a secondary task prior to the onset of a 'pop-out' stimulus will reduce the amplitude of the P300 response [29]. Applying this terminology to the context of this experiment, the secondary task corresponds to the viewing of the movie content, and the checkerboard is the 'pop-out' stimulus. This interpretation suggests an overall increased level of engagement during HDR content viewing relative to SDR content. This interpretation and its relationship to the results of the analysis of the alpha power time-courses are discussed further below.

Discussion

In this study, we explore alpha power, SSVEPs and the P300 ERP component of EEG signals measured during viewing of SDR and HDR cinematic content. The strength of this approach, compared to traditional methods like rating scales, lies in the viewer's ability to have a natural and undistracted viewing experience of cinematic content, interspersed with collection of physiological measurements elicited by a strong artificial stimulus, the reversing checkerboard. Interactions between these natural and artificial stimuli warrant further investigation.

These methods of signal collection have other advantages in that they do not require any knowledge of the experiment on the part of the subject, and are therefore not prone to certain bias and noise inherent to numerical or adjective-based ratings tests. The subject need not even be informed that an experiment is taking place. Complete absence of awareness that an experiment is being conducted would only be possible with certain types of physiological measurement techniques, such as thermal imaging and some forms of pupillometry. With most physiological measurement techniques, the subject will know that some kind of experiment is taking place, but it still possible to not know what is actually being studied. For example, in the case of studying a subject watching movies while using slightly intrusive equipment, such as heart-rate monitors, they will certainly know an experiment is occurring, but they will not know what experimental variable is being manipulated, or even whether the variable pertains to image, sound, content, etc. Thus the act of the experiment will have a limited effect on the perceptual responses, and observations can be interpreted knowing that certain experimental goals and parameters were unknown to the participant.

Content used for this experiment was mapped either to SDR, to simulate typical consumer viewing equipment, or to HDR on a stimulus display with a 0.005-1000 cd/m^2 dynamic range. While this provided a noticeably expanded color gamut and luminance to vision experts, surveys of experimental subjects following testing revealed that several subjects did not notice significant brightness, contrast, or color differences during the experiment. Because HDR can easily extend far beyond 1000 cd/m^2, we believe our findings represent a lower bound of emotional and attentional states that can be evoked by the technology.

A reduction in alpha-band power is associated with tasks that drive attention and engagement. We found that alpha-band power in the EEG spectra was inversely correlated to median luminance levels, and that the correlation was stronger for the HDR viewing condition. This increased modulation may reflect a higher level of visual engagement.

Characteristic SSVEPs elicited by the inverting checkerboard stimulus were not found to differ between HDR and SDR contexts, or between bright and dim checkerboard stimuli. It is possible that the SSVEP is not subject to attentional modulation by the content, or that differences in state are obscured by a rapid or strong response to the stimulus.

However, the transition from cinematic content to the checkerboard stimulus also elicited robust P100 and P300 responses. We analyzed the P300 response and found it to be significantly reduced following HDR content, possibly due to increased cognitive engagement driven by the improved display parameters. To rule out a direct effect of brightness in the different P300 responses, we measured average power of the P300 for the dim and bright checkerboard conditions. These did not differ significantly, meaning brightness alone could not account for the reduced amplitude following HDR viewing. Furthermore, the amplitude of the transition from HDR to the bright checkerboard trended lower than the SDR-to-dim transition. Although that difference was not statistically significant due to the small size of the data set, more data that supported the result could be taken as additional evidence that the difference observed in P300 responses is not due directly to luminance, but rather higher-level attentional processes.

Together, these results demonstrate that properties beyond effects generated by low-level processing (e.g. visibility and brightness) can be measured with EEG. These data support the existence of signals that correlate well to modulation of viewer engagement and attention by visual stimuli.

Acknowledgements

We would like to thank Timo Kunkel and Jaclyn Pytlarz for the color volume plots, Robin Atkins and Suzanne Farrell for the display mapping algorithm, and Timo Kunkel for setup of the custom PRM-4200.

References

[1] B. Keelan, "Predicting multivariate image quality from individual perceptual attributes," IS&T 2002 PICS conferences, 2002.

[2] W. Chen, et al., "Quality of experience model for 3DTV," Proc. SPIE 8288, Stereoscopic Displays and Applications XXIII, 82881P, 2012.

[3] J.A. Redi, et al., "How Passive Image Viewers Became Active Multimedia Users," Visual Signal Quality Assessment, 2015.

[4] P. Hanhart, et al., "Subjective quality evaluation of high dynamic range video and display for future TV," IBC 2014.

[5] C. Neveu and L. Stark, "Hysteresis in accommodation," Ophthalmic and Physiological Optics, vol. 15, no. 3, pp. 207-216, 1995.

[6] J. L. Blin., "SAMVIQ–Subjective assessment methodology for video quality," Rapport Technique BPN, vol. 56, pp. 24, 2003.

[7] S. Farrell, "Best in Show: considerations for the creation of a perceptual quality metric," SMPTE Annual Tech Conference and Exhibition, 2015.

[8] J. Clark, "What is the 'cinema feel?'", Red Shark News, March 6, 2015. <http://www.redsharknews.com/post/item/2365-what-is-the-cinema-feel>

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-126

[9] S. Bentin and L. Deouell, "Structural encoding and identification in face processing: ERP evidence for separate mechanisms," Cognitive Neuropsychology, vol. 17, pp. 35-54, 2000.

[10] B. Nhan and T. Chau, "Classifying affective states using thermal infrared imaging of the human face," IEEE Trans. Biomedical Engineering, vol. 57, no. 4, 2010.

[11] L. Palmateer et al., "Characterization of electronic displays: current methods to human centered approaches as EEG brainwave monitoring," China Display, 2011.

[12] L. Lindemann and M. Magnor, "Assessing the quality of compressed images using EEG," IEEE 18th ICIP, 2011.

[13] S. Arndt, et al., "A physiological approach to determine video quality," IEEE Int. Symposium on Multimedia, 2011.

[14] L. Acqualagna, et al., "EEG-based classification of video quality perception using steady state visual evoked potentials," J. Neural Eng., vol. 12, 2015.

[15] S. Scholler, et al., "Toward a direct measure of video quality perception using EEG," IEEE Trans. Image Processing, vol. 21, no. 5, 2012.

[16] E. Calore, et al., "Analysis of brain activity and response during monoscopic and stereoscopic visualization," SPIE Proc., pp. 8288, 2012.

[17] Y. He, et al., "Assessing the impact of crosstalk on 3D image quality using EEG," SID IDW, pp. 3D2/VHF2, 2013.

[18] Y. Kuroki, et al., "Effects of motion image stimuli with normal and high frame rates on EEG power spectra: comparisons with continuous motion image stimuli," JSID, vol. 22, no. 4, 2015.

[19] J. Kamienkowski, et al., "Fixation-related potential in visual search: A combined EEG and eye tracking study," J. Vision, vol. 12, no. 7, pp. 4, 2012.

[20] J. P. Dmochowski et al., "Audience preferences are predicted by temporal reliability of neural processing," Nature Communication, vol. 5, pp. 4567, 2014.

[21] C. Goldman, "This is your brain on Jane Austen, and Stanford researchers are taking notes," Stanford Report, Sept. 7, 2012.

[22] S. Szczepanski, et al., "Dynamic changes in phase –amplitude coupling facilitate spatial attention control in fronto-parietal cortex," PLOS Biology, vol. 12, no. 8, 2014.

[23] O. Jensen and A. Mazaheri, "Shaping functional architecture by oscillatory alpha activity: gating by inhibition," Frontiers in Human Neuroscience, vol. 4, 2010.

[24] H. Van Dijk, et al., "Pre-stimulus oscillatory activity in the alpha band predicts visual discrimination ability," The Journal of Neuroscience, vol. 28, no. 8, pp. 1816-1823, 2008.

[25] S. P. Kelly, et al., "Visual spatial attention tracking using high-density SSVEP data for independent brain-computer communication," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 13, no. 2, pp. 172-178, 2005.

[26] W. S. Pritchard, "Psychophysiology of P300," Psychological Bulletin, vol. 8, no. 3, pp. 506, 1981.

[27] C. C. Duncan, et al., "Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400," Clinical Neurophysiology, vol. 120, no. 11, pp. 1883-1908, 2009.

[28] E. Donchin and M. G. H. Coles, "Is the P300 component a manifestation of context updating?", Behavioral and Brain Sciences, vol. 11, no. 3, pp. 357-374, 1988.

[29] J. Polich, "Updating P300: an integrative theory of P3a and P3b," Clinical Neurophysiology, vol. 118, no. 10, pp. 2128-2148, 2007.

[30] E. Reinhard, et al., High Dynamic Range Imaging: Acquisition, display, and image-based lighting, Burlington MA: Morgan Kaufmann, 2006.

[31] J. McCann and A. Rizzi, The Art and Science of HDR Imaging, Hoboken NJ: Wiley Press, 2011.

[32] F. Dufaux, et al., High Dynamic Range Video: From Acquisition to Display and Applications, Cambridge MA: Academic Press, 2016.

[33] A. Pennington, "Pixar and ILM keynote at the IBC Big Screen," TVB Europe, Sept 2015. <http://www.tvbeurope.com/pixar-ilm-keynote-ibc-big-screen/>

[34] P. Hanhart, et al., "Subjective evaluation of higher dynamic range video," Proc. SPIE, pp. 9217, 2014.

[35] R. Oostenveld, et al., "FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data," Computational Intelligence and Neuroscience, 2010.

[36] T.P. Jung, et al., "Removing electroencephalographic artifacts by blind source separation," Psychophysiology, vol. 37, no. 2, pp. 163-178, 2000.

[37] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," Journal of Neuroscience Methods, vol. 134, no. 1, pp. 9-21, 2004.

[38] D. Lehmann, "Multichannel topography of human alpha EEG fields," Electroencephalography and Clinical Neurophysiology, vol. 31, no. 5, pp. 439-449, 1971.

[39] F. Di Russo, et al., "Cortical sources of the early components of the visual evoked potential," Human Brain Mapping, vol. 15, no. 2, pp. 95-111, 2002.

Author Biography

Dan Darcy received his PhD from UCSD, Evan Gitterman received his BS from Stanford University, Alex Brandmeyer received his PhD from the Radboud University Nijmegen, Scott Daly received his MS from U of Utah, and Poppy Crum received her PhD from UC Berkeley. All are currently members of the science and imaging groups at Dolby Laboratories in San Francisco, CA.