An open-source toolbox for measuring dynamic video framerates and synchronizing video stimuli with neural and behavioral responses

Benjamin G. Schultz\textsuperscript{a,b,*}, Emmanuel Biau\textsuperscript{a}, Sonja A. Kotz\textsuperscript{a}

\textsuperscript{a} Basic & Applied Neurodynamics Laboratory, Department of Neuropsychology & Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, the Netherlands
\textsuperscript{b} Institute for Logic, Language, & Computation, Faculty of Humanities, University of Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Oscillations
Video presentation
EEG
MATLAB toolbox
Visual stimuli
Steady-state evoked potentials
Entrainment

ABSTRACT

Background: Researchers rely on the specified capabilities of their hardware and software even though, in reality, these capabilities are often not achieved. Considering that the number of experiments examining neural oscillations has increased steadily, easy-to-implement tools for testing the capabilities of hardware and software are necessary.

New method: We present an open-source MATLAB toolbox, the Schultz Cigarette Burn Toolbox (SCiBuT) that allows users to benchmark the capabilities of their visual display devices and align neural and behavioral responses with veridical timing of visual stimuli. Specifically, the toolbox marks the corners of the display with black or white squares to indicate the timing of onset of static images and the timing of frame changes within videos. Using basic hardware (i.e., a photodiode, an Arduino microcontroller, and an analogue input box), the light changes in the corner of the screen can be captured and synchronized with EEG recordings and/or behavioral responses.

Results: We demonstrate that the SCiBuT is sensitive to framerate inconsistences and provide examples of hardware setups that are suboptimal for measuring fine timing. Finally, we show that inconsistencies in...
framerate during video presentation can affect EEG oscillations. 

Conclusions: The SCiBuT provides tools to benchmark framerates and frame changes and to synchronize frame changes with neural and behavioral signals. This is the first open-source toolbox that can perform these functions. The SCiBuT can be freely downloaded (www.band-lab.com/scibut) and be used during experimental trials to improve the accuracy and precision of timestamps to ensure videos are presented at the intended framerate.

1. Introduction

Research into neural oscillatory activity has recently received increased attention. Several papers have examined how these oscillations may correspond with dynamic properties of video stimuli. However, there are several issues surrounding the hardware and software setups used when presenting videos that are rarely discussed. One such issue is whether the configuration can consistently produce the frames per second (FPS) specified by the video. Similarly, video production may not be synchronized with neural and behavioral responses and frame rates may show temporal drift. Here we present the Schultz Cigarette Burn Toolbox (SCiBuT) that borrows a convention from motion picture films where a changeover cue is used to inform projectionists of when to change film reels; a mark (known colloquially as a “cigarette burn”) appears in a corner of the screen and acts as a cue for a change of frames or to mark an event. This mark can be read by an external visual sensor (i.e., a photodiode) to ascertain the veridical timing of frame onsets and changes. The present paper describes how the SCiBuT can be used to examine the consistency and latency of visual stimuli in various conditions (e.g., different hardware arrangements) and presents human neural data to demonstrate that such temporal inconsistencies are perceptually relevant.

1.1. Entrainment to dynamic stimuli

Entrainment is a process whereby neural oscillations adapt to and synchronize with a regular external stimulus. In the auditory domain, an example is the tick-tock of a clock that makes a sound every second at a frequency of 1 Hz. As the perceiver listens to the clock, their neural oscillations entrain to this regularity and a Fourier transformation of Electroencephalography (EEG) recordings reveals that these oscillations occur at a rate of 1 Hz (e.g., Nozaradan et al., 2011). These auditory entrainment mechanisms indicate the synchronization of neural ensembles in primary auditory cortex when tracking the temporal structure of a sensory input (Ding et al., 2016; Gross et al., 2013; Nourski et al., 2009; Schroeder and Lakatos, 2009; Zoefel and VanRullen, 2016). Similarly, if a person watched the pendulum of a clock in silence, then neural oscillations would occur at a rate of 1 Hz (the time between each tick and tack) and/or 0.5 Hz (the time between ticks and also between tocks) in the visual cortex. These neural oscillations in response to stimuli with a constant rate are called steady-state evoked potentials (also, stead-state visual evoked potentials) and are well established in the literature (cf. Norcia et al., 2015).

Entrainment to dynamic visual stimuli (e.g., videos) has also been examined with regard to entrainment to rhythms (Friedman-Hill et al., 2000), Herrmann, 2001; Keitel et al., 2017; Krolak-Salmon et al., 2003; Williams et al., 2004), multisensory integration of audiovisual speech (Romero et al., 2015; Schepers et al., 2013), and other dynamic properties of video stimuli (e.g., Avanzini et al., 2012; Biau et al., 2015; Park et al., 2016; Press et al., 2011). In the context of multimodal speech perception, studies show oscillatory responses in low frequency bands during audiovisual speech processing (Park et al., 2015, 2016). Although auditory stimulus presentation appears to be temporally stable (e.g., Nozaradan et al., 2011), video presentation requires additional considerations to account for possible artefacts introduced by hardware and software. The fundamental concern of the present article is the framerate at which a video is presented, which may or may not remain constant.

When presenting videos at a predetermined framerate, researchers expect video presentation to be consistent and to accurately reflect the framerate specified by the video file but, in most cases, this might not be achieved. Moreover, researchers assume that every frame is presented even though frames can sometimes be dropped for various reasons based on the hardware and/or software. Different sources of instability can disrupt the actual framerate, specifically, the hardware within the computer (e.g., the graphics processing unit; GPU), the capabilities of the monitor (e.g., the refresh rate), the video settings, and the software used to present videos. Additionally, there can be issues in the analyses if the realized framerate does not match the expected framerate; neural responses may not align with specific time points within the video, entrainment to regular visual stimuli may occur at a different rate, or it could become impossible to remove neuro-visual artefacts introduced by frame changes through filtering (e.g., notch filters removing the frequency, or harmonics, of frame changes). Any inconsistency in the framerate can affect oscillatory responses in the brain, and lead to artefacts in the frequency domain. Studies investigating oscillatory brain activity in the brain during visual stimulation with highly-controlled flickering light-emitting diodes reported entrainment in the occipital cortex at the frequency of the light onsets up to approximately 60 Hz (Herrmann, 2001). Entrainment can also occur at the frequency of the framerate when presenting visual stimuli using a computer (Williams et al., 2004). Quasi-rhythmic visual stimulation also affects neural entrainment suggesting that the temporal structure of visual events corresponds to specific oscillatory responses (Keitel et al., 2017). Therefore, it is imperative that veridical visual presentation times can be monitored. To the knowledge of the authors, no open-source tools for measuring framerates and video onsets are currently available for the scientific community (but see Poth, et al., 2018 for measuring static images). Moreover, there are no commercially available tools for accurately synchronizing video events with behavioral responses and EEG. Here, we describe technical details in layperson terms to guide non-technical readers in their hardware/software decisions and to evaluate experimental setups used in reported research.

There are multiple possible sources of temporal variability when presenting videos. First, we discuss the computer screen, also called monitors or displays. Monitors fall into several categories each with advantages and disadvantages with regard to color accuracy and physical limits in timing capabilities. Because the SCiBuT only measures timing accuracy, we focus on the temporal characteristics of monitors for the present study. The two main specifications that describe the temporal properties of a monitor are latency and refresh rate. Latency refers to how long it takes the monitor to show visual information after it has been received from the computer. The refresh rate is the number of times per second that a monitor updates what is shown on the screen. However, the refresh rate is distinct from the framerate; the framerate is how often the computer sends an entire frame to the monitor whereas the refresh rate includes updating the screen even if the frame being displayed is identical to the previous one. Also, the monitor will only display new information in the next refresh cycle after the information has been received and the specified latency period has passed. Finally, monitors may vary in how they process extra frames that have not been displayed before the next frame arrives, and this level of detail is not provided by the manufacturer, likely due to intellectual property protection.

To complicate the matter further, computers have certain constraints regarding how many frames can be sent to the monitor per
second even if a framerate is specified. Specifically, computers can have an “integrated graphics processor unit” (iGPU) whereby the central processing unit (CPU) and random access memory (RAM) are used to process graphics (i.e., what is displayed on your monitor) or a “dedicated graphics processor unit” (dGPU) that uses its own RAM (i.e., Video RAM). Dedicated GPUs usually provide more consistent framerates than integrated GPUs because the CPU and RAM resources are used for multiple computer applications simultaneously and might be required for other tasks that are occurring without the user’s knowledge. In contrast, the video RAM from a dedicated GPU is solely used for graphics processing (unless specified otherwise by the user) resulting in a more consistent framerate when presenting static images, displaying videos, or dynamically changing images (e.g., video games). Many experiments do not report GPU and monitor specifications, making it difficult to assess whether the hardware could achieve the expected framerates.

A related source of framerate inconsistency is the number of pixels per second that are sent by the computer and displayed by the monitor. For a color image, each pixel is 24 bits and higher resolution videos contain more pixels. For example, high definition videos contain 1920 (width) by 1080 (height) pixels meaning that each frame contains 2,073,600 pixels adding up to 49,766,400 bits per frame and, for a video playing at 24 FPS, a total of 1,194,393,600 bits per second (a little over a gigabit). For this reason, videos are often “compressed” using encoders (i.e., “codecs”) to save space and then decoded when playing back the video. If the video information is not decoded by the computer quickly enough as a result of large file size and/or slow decoding, the frame might be presented late or dropped entirely. Thus, the computer hardware and monitor must both be capable of producing these frames at a consistent rate. In the present experiment we keep the pixel resolution constant, but researchers should be aware that the pixel resolution will affect the capabilities of a given video setup.

The final source of framerate noise discussed in the present study is the software used to present videos and, relatedly, the decoders (i.e., “codecs”) used to read video files of various formats (e.g., mp4, mov, avi). It is possible that different software packages perform sub-optimally for video presentation but this has not yet been tested. Experiment software that can present videos includes the Psychtoolbox (Brainard, 1997), Presentation (Neurobehavorial Systems, Berkeley, CA), e-Prime (Psychology Software Tools, Pittsburgh, PA), Python’s PsychoPy2 (Peirce et al., 2019), MATLAB (Mathworks, Natick, MA), and others. The present study does not systematically test all these configurations but, instead, provides a free toolbox that can be used by researchers to benchmark their own systems and report the system capabilities alongside their results. Moreover, the timestamps from the visual sensor can be synchronized with behavioral and EEG data to provide accurate measurements of stimulus onsets and framerates within each trial.

1.2. Aims & hypotheses

We introduce the SCiBuT to precisely record video onsets and framerates via visual sensors that can be synchronized with EEG recordings and responses via an analogue input box. We also provide scripts and a user guide to perform behavioral experiments using an Arduino microcontroller. To demonstrate the necessity of the SCiBuT, we investigated the effect of inconsistent framerates using silent video examples encoded at different framerates (up to 90FPS) that depicted various kind of visual information (e.g., action scenes, slow pans, a bouncing ball, and flicker between black and white for each frame) while recording the EEG signal of participants. Framerate consistency was compared between using the integrated and dedicated GPU on the same system. We further examined the effect of inconsistent framerates on neural entrainment by examining neural entrainment to the flicker video in all conditions. We hypothesized that framerates are less consistent when using an integrated GPU compared to a dedicated GPU. We further hypothesized that framerate consistency decreases as the framerate increases. Finally, we hypothesized that neural entrainment to videos is more evident when the framerate is consistent.

1.3. The Schultz cigarette burn toolbox

The SCiBuT (www.band-lab.com/scibut) contains schematics and MATLAB code to facilitate the accurate measurement of video onsets and framerates using any setup by recording voltage changes produced by a photodiode (see Fig. 1). Photodiodes convert light into an electrical current with sub-microsecond latency (Goushcha and Tabbert, 2017). The electrical current can be read by any external device that measures changes in electrical charge (e.g., oscilloscopes, analog input boxes, and microcontrollers). The present experiment used an analog input box to record the continuous readings from the photodiode at a rate of 1000 Hz, thus providing veridical timestamps for frame onsets. The first step for using the SCiBuT is to add the “cigarette burns” to the desired video stimuli using the SCiBuT addCB.m function. These cigarette burns change the corner of the screen from black to white for each consecutive frame to provide a cue that the frame has changed. In the event of faster frame rates, a second corner can be activated that changes every two frames, a third corner that changes every four frames, and a fourth frame that changes every eight frames. Thus, we can provide 4-bit resolution to examine whether multiple consecutive frames (up to 16) were missed. Given the slow frame rates tested here (24FPS to 90FPS) and the high sampling rate (1000Hz), we used the 2-bit resolution allowing us to measure up to four consecutive missing frames.

The second step is to construct the photodiodes to and attach them to the corners of the screen (see Fig. 1) and cover the corners to ensure participants cannot see the “cigarette burns”. If recording using an analog input box, ensure the additional sensor channels are connected to the analog input box and that the software is able to read the channels (i.e., add physical channels and test for signal changes when the screen changes color). If recording using an Arduino microcontroller, ensure the cables are configured correctly (see Supplementary Materials) and that the provided scripts record changes when the screen changes color. Note that only data recorded from an analog input box are reported in the present study as this validates the first step.
synchronization of neural responses with frame onset times; the temporal precision of Arduino microcontrollers when recording analog signals at 1000 Hz has previously been verified (see D’Ausilio, 2012; Schubert et al., 2013; Schultz and van Vugt, 2016).

Once data have been collected, video onsets and offsets (i.e., frame changes) can be extracted through signal processing techniques provided in the custom-made MATLAB scripts that are available for download from the Basic and Applied NeuroDynamics Laboratory website (www.band-lab.com/scibut). Furthermore, these triggers can be used to measure the temporal delay of video or visual stimulus onsets (relative to timestamped onsets or triggers) produced by any hardware and/or software configuration. Frame changes (changes from low to high or high to low) were estimated by the first moment that the normalized voltage (ranging from 0 to 1) cross over the midpoint of 0.5 and then pinpointed by finding the point at which the previous sample (n-1) no longer had a lower voltage (for low-to-high changes) or a higher voltage (for high-to-low changes).

2. Methods

2.1. Participants

Six participants (4 females, mean age = 32.2 years, SD = 2.7 years, range = 27–37 years) volunteered after giving informed consent. All participants had normal or corrected-to-normal vision and no hearing deficits. The effects examined in this study could be detected using a small sample size following on from previous studies using psycho-physical approaches to visual entrainment (cf. Norcia et al., 2015). We had strong directional hypotheses and only examined the frequency and sub-harmonic of the video framerate. Moreover, we used a repeated-measures design based on well-established principles in vision science including strong measurement, strong theory, and effective control of error variance (see Ross, 2009; Smith and Little, 2018). We further report effect sizes and these were moderate-to-large for our hypothesis tests. Trial and error variance (see Ross, 2009; Smith and Little, 2018) are used to measure the temporal delay of video or visual stimulus onsets (relative to timestamped onsets or triggers) produced by any hardware and/or software configuration. Frame changes are estimated by the first moment that the normalized voltage (ranging from 0 to 1) cross over the midpoint of 0.5 and then pinpointed by finding the point at which the previous sample (n-1) no longer had a lower voltage (for low-to-high changes) or a higher voltage (for high-to-low changes).

2.2. Stimuli

Seven 20-second silent videos were used covering a range of scenes and initial construction methods (i.e., filmed scenes, animations, and custom-made videos) with the same number of pixels (1280 × 720) and presented in .avi format using MJPG compression. The “flicker” video displayed changes between black and white frames for each subsequent frame at framerges of 24FPS, 30FPS, 60FPS, and 90FPS. To ensure that framerate effects were not driven by our method of video construction and to prevent neural entrainment carryover between trials, one custom-made video (at all FPS) and five movie scenes (at 24FPS, 30FPS, and 60FPS) were also presented (see Appendix A).

2.3. Apparatus

The stimulus presentation computer (Intel i7-6700 CPU @ 3.40 GHz, 32 GB, running 64-bit Windows 7) contained a dedicated GPU (NVIDIA GeForce GTX 1080 GPU). The monitor was a 27-inch iiyama G-MASTER (GB2760HSU-B1) TN display with a 1ms response time, a refresh rate of 144Hz, and a native resolution of 1920 × 1080 pixels. The monitor was connected to the video output of the dedicated GPU for the dedicated condition and connected to the motherboard for the integrated condition. Stimulus presentation was performed using a custom-made MATLAB (64-bit, Version 7.12.0, R 2011a, The MathWorks, Natick, MA, USA) that called VideoLAN Client (VLC) media player to play the videos (Videolan Client, 2017; http://www.videolan.org/). VLC was allowed to drop frames or skip late frames to keep the average framerate consistent. EEG and sensor data were collected using BrainVision Recorder (Brain Products, GmbH, 2017) software on an Intel Xeon E5-1650 PC (3.5GHz, 32GB RAM) running Windows 7. EEG data were preprocessed using the Letswave toolbox for MATLAB (version 6; http://nocions.webnode.com/letswave) and framerate data were analyzed using the SCiBuT.

2.4. Procedure

Participants were seated in a sound attenuated booth, about 30 cm from the monitor. In each trial, the participant pressed the spacebar to launch the video. To minimize movements contamination, they placed their chin on a headrest fixed to the table and were asked to gaze at the central part of the screen and avoid blinking at the maximum during the video presentation. Participants were able to start each trial at their leisure and were instructed to take as much time as needed to rest between trials. Participants were presented all 33 videos three times within each block (see Appendix A). At the end of the first block (with either the dedicated or integrated GPU, counter balanced), participants could have a longer rest, while the experimenter changed the display connection. Once ready, the participants pressed space bar to begin the second block. In total, the experiment lasted approximately 50 minutes.

2.5. EEG analysis

Sensor and electrophysiological data were recorded at 1000 Hz from 64 active electrodes (ActiCap, Brain Vision Recorder, Brain Products) according to the 10–20 international standard and impedances were kept below 10 kΩ. The amplifier was powered via a battery to avoid electrical noise stemming from the power line frequency (50Hz-60H). Four electrodes were used to detect ocular artefacts, placed above and below the left eye and on the outer canthus of the left and right eyes. The ground electrode was fixed to the right collarbone and the reference electrode was placed at the mastoid. A highpass 1 Hz Butterworth filter was applied to the EEG data, ocular artefacts were removed using independent components analysis (Keil et al., 2014). Following on from previous studies (cf. Norcia et al., 2015) steady-state evoked potentials were measured using fast Fourier transforms (FFT) with baseline correction (signal-to-noise ratio) from 2 to 18 seconds after the first video onset. EEG channels representing the occipital cortex were averaged (Oz, POz, PO3, PO4) and only EEG data of the Flicker video were analyzed to examine the effects of consistent and inconsistent framersates on neural entrainment.

2.6. Statistical analysis

Separate linear mixed-effects models (LMM) were conducted for the dependent variables framerate accuracy (veridical FPS minus video FPS), framerate variability (the coefficient of variation defined as the standard deviation of FPS divided by the mean FPS), and the number of missed frames (defined as frame onsets that occurred more than half a period after the video-specified onset time). The fixed factors were GPU (dedicated, integrated) and Video FPS (4; 24FPS, 30FPS, 60FPS, 90FPS) with the maximal random effects structure justified by the experimental design (see Barr et al., 2013), that is, Video FPS nested within GPU, GPU nested within Video, Video nested within Trial, and Trial nested within Participant. To examine neural entrainment to framersates, we measured FFT power at the fundamental frequency (i.e., the video framerate) and the first subharmonic (i.e., half the video framerate) for EEG epochs showing the Flicker video. These data were subjected to a LMEM with fixed factors GPU, Video FPS, and Harmonic (2; sub-harmonic, fundamental) with a maximal random effects structure with Harmonic nested within Video FPS, Video FPS nested within GPU, GPU nested within Trial, and Trial nested within Participant.

LMEMs were performed using the lmer function of the lme4 library (Bates et al., 2015) or the lme function of the nlme library (Pinheiro
et al., 2015) for the R package of statistical computing (R Core Team, 2013) using Satterthwaite’s method of approximation for degrees of freedom. F-statistics, significance values, and effect sizes (generalized eta squared; $\eta^2_G$, where 0.02 = small, 0.13 = medium, and over 0.26 = large; Bakeman, 2005) are reported. Two-tailed pairwise contrasts were computed using generalized linear hypothesis testing for Tukey’s Honestly Significant Difference contrasts, using the glht function in the multcomp library (Hothorn et al., 2008).

3. Results

3.1. Framerate accuracy

For framerate accuracy, there were significant main effects of GPU [$F(1, 827) = 75,868.00, p < .001, \eta^2_G > 0.99$] and Video FPS [$F(3, 827) = 5313.50, p < .001, \eta^2_G > 0.99$], and a significant interaction between GPU and Video FPS [$F(3, 827) = 22,232.53, p < .001, \eta^2_G > 0.99$]. Pairwise comparisons yielded significant differences between all levels of the interaction ($ps < 0.001$) except between 24FPS and 30FPS for the dedicated GPU ($p > 0.99$). As shown in Fig. 2a, the dedicated GPU produced more FPS on average relative to the expected FPS for 60FPS compared to 24FPS and 30FPS, and for 90FPS compared to all other FPS conditions. This indicates that the dedicated GPU may have presented some frames too slowly and then the framerate increased and overcompensated to catch up. In contrast, the integrated GPU produced relatively slower framersates and was unable to compensate.

3.2. Framerate variability

The coefficient of variation was positively skewed and, therefore, data were log transformed prior to analysis. As variance of the coefficient of variation of FPS was not equal between levels of GPU, a non-linear mixed-effects model was fit to the data with variance allowed to vary across levels of GPU. There were significant main effects of GPU [$F(1, 268) = 1274.86, p < .001, \eta^2_G = .97$] and Video FPS [$F(3, 268) = 266.84, p < .001, \eta^2_G = .87$], and a significant interaction between GPU and Video FPS [$F(3, 268) = 45.31, p < .001, \eta^2_G = .87$]. Pairwise comparisons revealed significant increases in variability as the video framerate increased for the dedicated GPU ($ps < 0.001$). The integrated GPU demonstrated significant variability increases from 24FPS to 30FPS, from 24FPS to 60FPS, from 24FPS to 90FPS ($ps < 0.007$). Interestingly, the integrated GPU demonstrated significant variability decreases from 30FPS to 60FPS and 30FPS to 90FPS ($ps < 0.001$), and 60FPS and 90FPS did not significantly differ ($p = .09$). The integrated GPU demonstrated significantly greater variability than the dedicated GPU for all video FPS conditions ($ps < 0.001$). As shown in Fig. 2b (and Fig. 2a), even though the integrated GPU demonstrated lower variability in the 60FPS and 90FPS conditions compared to the 30FPS condition, the accuracy did not
approach the correct framerate. In other words, the integrated GPU consistently produced inaccurate framerates for higher framerates.

3.3. Missed frames analyses

For the proportion of missed frames, there were significant main effects of GPU \(F(1, 821.01) = 92,137.87, p < .001, \eta^2 = .99\) and Video FPS \(F(3, 821.01) = 18.99, p < .001, \eta^2 = .57\), and a significant interaction between GPU and Video FPS \(F(3, 821.01) = 7.87, p < .001, \eta^2 = .36\). For the dedicated GPU, pairwise comparisons only yielded significant differences between 90FPS and the other three Video FPS conditions \(p < 0.001\) indicating that more frames were missed for the 90FPS condition (see Fig. 2c). For the integrated GPU, the 30FPS condition missed significantly fewer frames than the 60FPS and 90FPS conditions \((p < 0.02\). Based on the extremely large effect size, the dedicated GPU missed decidedly fewer frames than the integrated GPU for all Video FPS conditions \(p < 0.001\).

3.4. Neural oscillation analysis

As shown in Fig. 3, the pattern of power of neural oscillations at various frequencies differed between the GPU and Video FPS conditions. Specifically, the dedicated GPU demonstrated a relatively clear pattern of activation for the 24FPS, 30FPS, and 60FPS conditions at the fundamental frequency, subharmonic, or both (see Fig. 3a, solid and dotted white lines). Conversely, the pattern of activation for 90FPS with the dedicated GPU and all Video FPS conditions with the integrated GPU were not localized at the fundamental frequency, subharmonics, or harmonics. Instead, activation spread into multiple frequencies that were not harmonically related to the Video FPS. To statistically justify this interpretation, EEG power the frequency spectrum was normalized according to the maximum power across frequencies ranging from 4 Hz to 120 Hz, and the normalized power at the fundamental frequency (i.e., the Video FPS) and first subharmonic (i.e., half of the Video FPS) for each Video FPS were subjected to the LMEM (see Fig. 4). These were chosen because participants viewing the flicker video could entrain to the time between white screens (and/or black screens) at the subharmonic or to each screen change (black-to-white and white-to-black). An analysis of the non-normalized power and spaghetti plots that show within-subject effects are presented in Appendix B.

Normalized power was positively skewed and, therefore, data were log transformed prior to analysis. As variance of the normalized power was not equal between levels of GPU, a non-linear mixed-effects model was fit to the data with variance allowed to vary across levels of GPU. There were significant main effects of GPU \(F(1, 80) = 28.71, p < .001, \eta^2 = .25\) and Video FPS \(F(3, 80) = 3.04, p = .03, \eta^2 = .10\), and Harmonic failed to reach significance \(F(1, 80) = 3.56, p = .06, \eta^2 = .04\). There was a significant interaction between GPU, Video FPS, and Harmonic \(F(3, 80) = 4.95, p = .003, \eta^2 = .15\) and no other interaction reached significance \((p > 0.05)\). As shown in Fig. 4, pairwise comparisons demonstrated that the dedicated GPU produced significantly higher amplitudes than the integrated GPU for both Harmonics and all Video FPS conditions \((p < 0.002)\) except for 90FPS for the fundamental \((p = .98)\). For the first subharmonic, the dedicated GPU showed greater power for 60FPS compared to 24FPS \((p = .001\) 90FPS compared to 24FPS \((p = .003\); other \(p > .07\), and the integrated GPU showed greater power for 24FPS compared to all other Video FPS conditions \((p < .02\); other \(p > .69\). For the fundamental frequency, the dedicated GPU showed significantly greater power for 24FPS, 30FPS, and 60FPS compared to 90FPS \((p < .002\; other \(p > .92\), and the integrated GPU showed significantly greater power for 60FPS compared to 90 FPS \((p = .03\; other \(p > .22)\). Overall, these results are in line with those of the framerate accuracy and variability; as the framerate become more variable, power decreased at the stimulus frequency (and subharmonics) and shifted to other frequencies (see Fig. 3a and b). This interpretation was confirmed by significant negative Pearson correlations between the coefficient of variation and power for the first subharmonic \((r = -.30, p < .001)\) and fundamental frequency \((r = -.38, p < .001)\) indicating that power decreases at the expected frequency as framerate variability increases.

4. Discussion

We introduced the SCIbUt and demonstrated how it can be used to determine framerate consistency during video presentation. First, we showed that using a dedicated GPU results in greater accuracy, less variability, and fewer missed frames compared to an integrated GPU. Second, we showed that higher video framerates produce less accurate and more variable framerates. Third, we demonstrated that inconsistencies in the framerate can reduce neural entrainment to the stimulus frequency and, in fact, shifted power to a range of different frequencies that were unrelated to the frequencies of the content of the video (including harmonics and subharmonics). This poses a serious concern for the use of video stimuli in experiments because neural oscillations thought to represent underlying cognitive processes can be

---

Fig. 3. Heat maps of mean normalized power of neural oscillations for the Video frames per second (FPS) conditions (y axes) using the dedicated graphics processing unit (GPU; panel a) and integrated GPU (panel b) for the flicker video. The diagonal lines represent the following: The full white line is the fundamental frequency (i.e., one-to-one match of FPS and frequency), the dotted white line is the first subharmonic (i.e., the frequency is half of the FPS), the full black line is a harmonic of 1.5 times the FPS, and the dotted black line is double the FPS.
activated when framerates are variable and inaccurate, regardless of the video content (see Williams et al., 2004). Similarly, frames that contain crucial information (e.g., subliminal priming tasks) may not be delivered to participants or might be delivered at the incorrect point in time. The SCiBuT can be used to detect such inconsistencies, timestamp video onsets and events within a video, and ensure that all frames are presented within a given trial.

Some may argue that the effects of inconsistent framerates can be mitigated by increasing the number of repetitions and, subsequently, the signal-to-noise ratio. While this is correct in principle, the difficulty is that the framerate inconsistencies cover a wide range of frequencies and their harmonics and averaging these signals might simply produce a spread of activation (as shown in Fig. 3). A second concern is that EEG power is spread across lower frequency bands (i.e., less than 30 Hz) when using an integrated GPU (see Fig. 3b) and this effect was somewhat consistent across participants and trials. This likely occurred because the average framerates that were realized with the integrated GPU fell to these lower frequencies, albeit with higher variability. We recommend using dedicated GPUs for experiments that present dynamic visual stimuli with a monitor refresh rate above the "Nyquist frequency" (i.e. the refresh rate is at least double the desired video framerate). We further recommend that, should the framerate and (sub)harmonics of the framerate overlap with potential frequencies of interest, the video is presented at multiple different framerates; notch filters can then be implemented to remove the (sub)harmonics of the framerate before averaging to ensure that these artefact frequencies do not contaminate frequency responses that underlie perception and cognitive processes. Regardless, the capabilities of any system should first be benchmarked to ensure the chosen framerates are achievable and whether certain framerates are more consistent than others due when considering the monitor refresh rates.

The hardware configuration we used to test the SCiBuT was not the state of the art and, in the case of the integrated GPU, was expected to perform sub-optimally. This was done to show that expected framerates are achievable with non-specialized equipment, that is, a modern GPU and monitor with a high refresh rate. For optimal framerate consistency under high load, monitors with dynamic refresh rates (e.g., Gsync or Freesync) are recommended (see Poth et al., 2018 for more details). Given the absence of detail regarding computer hardware and monitor specifications for some experiments, it is likely that most researchers assume that the theoretical capabilities of a system are realized. We recommend that computer hardware, software, video file, and monitor specifications are not only reported but also benchmarked so readers can determine the margin of error for results. For example, the present results using the dedicated GPU are in line with those of Herrmann (2001) that demonstrated neural entrainment to the fundamental frequency of flickering LEDs up to approximately 60 Hz. The integrated GPU, on the other hand, only demonstrated entrainment for 24FPS. Without knowledge of the framerate inconsistencies within some systems (i.e., integrated GPUs), one may have incorrectly concluded that visual entrainment only occurs up 24FPS, even though analog methods have indicated that the limits are up to 60 Hz (Herrmann, 2001). The SCiBuT could prevent misleading conclusions by revealing the veridical timing offered by a given configuration.

The current experiment highlights how small differences in hardware and software configurations might increase the latency or variability of frames when presenting visual stimuli. Therefore, the latency and consistency of each configuration should be measured and possible sources of latency require further investigation. For example, we did not examine the effects of different video resolutions and subsequent increases in pixels per second that might also incur framerate inaccuracies. If researchers were to use videos with different pixel resolutions, differences in neural activity could be driven by artefacts resulting from stimulus properties. Furthermore, the information content within a frame or between frames within a video that has been compressed using a video codec may affect framerate accuracy because there is more information to be decoded. In experiments using dynamic images or videos, such effects might drive differences in neural activity that are thought to represent perceptual learning and neural responses to differences in entropy (e.g., predictive coding; see Huang and Rao, 2011). If certain compression methods influence framerate consistency, this problem could be circumvented using uncompressed video files or lossless compression. Future experiments could examine how these differences in video formats affect framerate consistency and, consequently, neural oscillations.

The SCiBuT could also be used to test asynchronies between
auditory and visual stimuli presented using videos. Many studies have examined multimodal synchrony perception (cf. Vroomen and Keetels, 2010) but the fidelity of veridical audiovisual synchrony could be influenced by the framerate and whether the frames are presented synchronously with the associated auditory stimulus. Moreover, the method of producing an auditory and visual stimulus simultaneously might differ between software and should be tested to ensure that the effects of perceived synchrony are free from artefacts resulting in framerate inconsistencies and asynchronous audiovisual presentation. This could be achieved by simultaneously measuring auditory and visual onsets with highly controlled multimodal stimuli that are in perfect synchrony and varying levels of asynchrony to see if these specified asynchronies are realized.

Software for presenting videos and/or dynamic visual stimuli should be benchmarked to ensure that frames are delivered at a consistent framerate without missing frames. A range of different programs have been used in previous experiments and, to the knowledge of the authors, the consistency of video presentation has not yet been tested (but see Poth, et al., 2018 for measuring the onset consistency of static images using C++, Python’s PsychoPy2, and Psychtoolbox3 for MATLAB). The present study used VLC media player to present stimuli but there are a variety of settings within VLC as well as other software (e.g., Media Player Classic) that may offer superior performance. The goal here was not to exhaustively test all possible configurations of hardware and software available to present video stimuli. Instead, we provide a tool for researchers to test their own configuration and adjust accordingly. The SCiBuT can test the performance of multiple systems and allows users to decide which software configuration(s) have an acceptable level of framerate accuracy, and framerate variability, and number of missed frames. The SCiBuT can further be used to determine the veridical commencement time of videos compared to the timestamps provided by experimental software.

5. Conclusion

The schematics, scripts, and sensor data are available online to download for free (www.band-lab.com/scibut), and any data submitted from other researchers will also be made available. The scripts in the SCiBuT can be used on any operating system and uses free software platforms (Python and Arduino IDE) to record the data. Moreover, the required hardware can be purchased for less than $50 USD. The SCiBuT provides a means to measure any visual stimulus presentation relative to EEG responses or other responses that generate a voltage signal. Experimenters can use the toolbox to measure temporal inconsistencies in their own setups and test the latencies of triggers instead of relying on the timestamps provided by experimentation software (e.g., PsychoPy2, Presentation, e-Prime, PsychToolbox). The SCiBuT allows sub-millisecond synchronization accuracy and precision of the stimulus with the subsequent neural response.

Author Contribution Statement

BGS developed the hardware and software, analyzed the data, and wrote the manuscript. EB collected the data, performed the literature review, and aided with the preprocessing of EEG data. SAK contributed to the preparation of the manuscript and provided the apparatus used to present and record data.

Open Practices Statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. The experiment was not preregistered.

Code Availability Statement

All code for the SCiBuT is freely available: www.band-lab.com/scibut

* Note: Custom-made videos using MATLAB

CRediT authorship contribution statement

Benjamin G. Schultz: Conceptualization, Formal analysis, Methodology, Software, Data curation, Project administration, Resources, Visualization, Writing - original draft. Emmanuel Biau: Investigation, Resources, Writing - review & editing. Sonja A. Kotz: Supervision, Writing - review & editing.

Declaration of Competing Interest

None.

Acknowledgements

The authors would like to thank the participants who kindly volunteered their time for this research and Rachel Brown, Michael Schwartz, and Laura Verga for their helpful comments on the manuscript. The authors hold no affiliations with any of the companies that developed the monitors, computer hardware, or software tested herein. E.B was supported by a postdoctoral fellowship from the European Union’s Horizon 2020 research and innovation program, under the individual Marie Skłodowska-Curie grant agreement (No. 707727).

Appendix A

Description of additional distractor videos used in the present experiment

The “bouncing ball” video contained a ball bouncing in an arc sine with the time between bounces randomly varying between 250 ms and 750 ms with a uniform distribution. The flicker and bouncing ball videos were generated using in-house MATLAB scripts (version 2016a, Mathworks, Natick, MA) at framers of 24FPS, 30FPS, 60FPS, and 90FPS. These videos were extracts from movies depicting one continuous shot in different styles (i.e., animations, action scenes, pursuit, walking, and slow pan; see Table A1). Movie videos were upsampled from 24FPS to 30FPS and 60FPS framers.

| Movie              | Time Start | Genre(s)                |
|--------------------|------------|-------------------------|
| The Adventures of Tintin | 1h21m38s   | Animation, Action, Adventure |
| The Avengers       | 1h56m55s   | Action, Adventure, Sci-Fi |
| Fantasia           | 1h59m47s   | Animation, Fantasy      |
| Goodfellas         | 0h31m32s   | Drama, Crime            |
| Kung Fury          | 0h18m30s   | Action, Comedy          |

Table A1

List of movie excerpts used in the experiment.
Using Adobe Premiere Pro CS3. Upsampling to higher rates (e.g., 90FPS) was not possible using this procedure, resulting in 3 FPS x 5 movies + 4 FPS x 2 custom-made videos (total = 33 videos, see Table A2). Following upsampling, the cigarette burns were applied to the upper left and right corner of the videos.

Appendix B

Analysis of FFT power for non-normalized data

Power was positively skewed and, therefore, data were log transformed prior to analysis. As variance of power was not equal between levels of GPU, a non-linear mixed-effects model was fit to the data with variance allowed to vary across levels of GPU. There were significant main effects of GPU \(F(1, 80) = 190.91, p < .001, \eta^2_G = .67\) and Video FPS \(F(3, 80) = 7.14, p < .001, \eta^2_G = .19\), and Harmonic failed to reach significance \(F(1, 80) = 3.56, p = .07, \eta^2_G = .03\). There was a significant interaction between GPU, Video FPS, and Harmonic \(F(3, 80) = 4.95, p = .005, \eta^2_G = .13\) and no other interaction reached significance \((ps > 0.05)\). As shown in Fig. 4, pairwise comparisons demonstrated that the dedicated GPU produced significantly greater power than the integrated GPU for both Harmonics and all Video FPS conditions \((ps < 0.05)\). For the first sub-harmonic, the dedicated GPU showed no significant differences between video framerates \((ps > 0.05)\) and the integrated GPU showed greater power for 24FPS compared to all other Video FPS conditions \((ps < 0.001; other ps > 0.62)\). For the fundamental frequency, the dedicated GPU showed significantly greater power for 24FPS, 30FPS, and 60FPS compared to 90FPS \((ps < 0.001; other ps > 0.94)\), and the integrated GPU showed significantly greater power for 60FPS compared to 90FPS \((p = .002)\) and for 24FPS compared to 90FPS \((p = .002; other ps > 0.21)\) (Figs. B1 and B2).

Overall, these results are in line with those of the framerate accuracy and variability; as the framerate became more variable, power decreased at the stimulus frequency (and subharmonics). This interpretation was confirmed by significant negative Pearson correlations between the coefficient of variation and power for the first subharmonic \((r = -0.55, p < .001)\) and fundamental frequency \((r = -0.58, p < .001)\) indicating that, as variability increases power at the expected power decreases.

Table A2

| Video                  | FPS Condition |
|------------------------|---------------|
|                        | 24FPS | 30FPS | 60FPS | 90FPS |
| Flicker*               | X     | X     | X     | X     |
| Bouncing Ball*         | X     | X     | X     | X     |
| The Adventures of Tintin | X | X | X | N/A |
| The Avengers           | X     | X     | X     | N/A   |
| Fantasia               | X     | X     | X     | N/A   |
| Goodfellas             | X     | X     | X     | N/A   |
| Kung Fury              | X     | X     | X     | N/A   |

Using Adobe Premiere Pro CS3. Upsampling to higher rates (e.g., 90FPS) was not possible using this procedure, resulting in 3 FPS x 5 movies + 4 FPS x 2 custom-made videos (total = 33 videos, see Table A2). Following upsampling, the cigarette burns were applied to the upper left and right corner of the videos.

Appendix B

Analysis of FFT power for non-normalized data

Power was positively skewed and, therefore, data were log transformed prior to analysis. As variance of power was not equal between levels of GPU, a non-linear mixed-effects model was fit to the data with variance allowed to vary across levels of GPU. There were significant main effects of GPU \(F(1, 80) = 190.91, p < .001, \eta^2_G = .67\) and Video FPS \(F(3, 80) = 7.14, p < .001, \eta^2_G = .19\), and Harmonic failed to reach significance \(F(1, 80) = 3.56, p = .07, \eta^2_G = .03\). There was a significant interaction between GPU, Video FPS, and Harmonic \(F(3, 80) = 4.95, p = .005, \eta^2_G = .13\) and no other interaction reached significance \((ps > 0.05)\). As shown in Fig. 4, pairwise comparisons demonstrated that the dedicated GPU produced significantly greater power than the integrated GPU for both Harmonics and all Video FPS conditions \((ps < 0.05)\). For the first sub-harmonic, the dedicated GPU showed no significant differences between video framerates \((ps > 0.05)\) and the integrated GPU showed greater power for 24FPS compared to all other Video FPS conditions \((ps < 0.001; other ps > 0.62)\). For the fundamental frequency, the dedicated GPU showed significantly greater power for 24FPS, 30FPS, and 60FPS compared to 90FPS \((ps < 0.001; other ps > 0.94)\), and the integrated GPU showed significantly greater power for 60FPS compared to 90FPS \((p = .002)\) and for 24FPS compared to 90FPS \((p = .002; other ps > 0.21)\) (Figs. B1 and B2).

Overall, these results are in line with those of the framerate accuracy and variability; as the framerate became more variable, power decreased at the stimulus frequency (and subharmonics). This interpretation was confirmed by significant negative Pearson correlations between the coefficient of variation and power for the first subharmonic \((r = -0.55, p < .001)\) and fundamental frequency \((r = -0.58, p < .001)\) indicating that, as variability increases power at the expected power decreases.

Fig. B1. Spaghetti plots of within-subject differences for power in the frequency spectrum.
Fig. B2. Spaghetti plots of within-subject differences for normalized power in the frequency spectrum.

References

Avanzini, P., Fabbri-Destro, M., Dalla Volta, R., Duperat, E., Rizzolatti, G., Cantalupo, G., 2012. The dynamics of sensorimotor cortical oscillations during the observation of hand movements: an EEG study. PLoS One 7 (5), e37534.

Bakeman, R., 2005. Recommended effect size statistics for repeated measures designs. Behav. Res. Methods 37, 379–384.

Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. J. Mem. Lang. 68, 255–278.

Bates, D.M., Maechler, M., Bolker, B., Walker, S., 2015. “Fitting Linear Mixed-Effects Models Using lme4”. Journal of Statistical Software 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Blei, D., Turrabia, M., Fuentemilla, L., de Diego Balaguer, R., Soto-Faraco, S., 2015. Speaker’s hand gestures modulate speech perception through phase resetting of ongoing neural oscillations. Cortex 68, 76–85.

Bing, N., Meleni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. 19, 158–164.

Friedman-Hill, S., Maldonado, P.E., Gray, C.M., 2000. Dynamic of striate cortical activity in the alert macaque: Incidence and stimulus-dependence of gamma-band neuronal oscillations. Cereb. Cortex 10, 1105–1116.

Goushcha, A.O., Tabbert, B., 2017. On response time of semiconductor photodiodes. Opt. Eng. 56, 097101.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., Garrod, S., 2013. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol. 11, e1001752.

Herrmann, C.S., 2001. Human EEG responses to 1–100 Hz flicker: resonance phenomena in continuous quasi-rhythmic sensory stimulation. Neuroimage 146, 58–76.

Knörig, A., Wettach, R., Cohen, J., 2011. Temporal envelope of time-compressed speech represented in the human auditory cortex. J. Neurosci. 29, 15564–15574.

Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E.S., Luck, S.J., et al., 2014. The steady-state visual evoked potential in vision research: a review. J. Vis. 15, 4.

Park, H., Ince, R.A.A., Schyns, P.G., Thut, G., Gross, J., 2015. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. Curr. Biol. 25, 1649–1653.

Park, H., Kayser, C., Thut, G., Gross, J., 2016. Lip movements entrain the observers’ low-frequency brain oscillations to facilitate speech intelligibility. eLife 5.

Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., et al., 2019. PsychoPy2: experiments in behavior made easy. Behav. Res. Methods 1–9.

Park, H., Bates, D., DeBoy, S., Sarkar, D., R Core Team, 2015. nlme: Linear and Nonlinear Mixed Effects Models. R Package Version 3.1-120. http://CRAN.R-project.org/package=nlme.

Poth, C.H., Forster, R.M., Behler, C., Schwazwecke, U., Schneider, W.X., Botsch, M., 2018. Ultrasonic temporal resolution of visual presentation using gaming monitors and G-Sync. Behav. Res. Methods 50, 26–38.

Press, C., Cook, J., Blakemore, S.-J., Kilner, J., 2011. Dynamic modulation of human motor activity when observing actions. J. Neurosci. 31, 2792–2800.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URLhttp://www.R-project.org/.

Romero, Y.R., Senkowski, D., Keil, J., 2013. Early and late beta band power reflects audiovisual perception in the McGurk illusion. American Journal of Physiology-Heart and Circulatory Physiology 113, 2342–2350.

Ross, J., 2009. Visual Perception 1950–2000. Inside Psychology: a Science Over 50 Years. pp. 243–252.

Schepers, I.M., Schneider, T.R., Hupp, J.F., Engel, A.K., Senkowski, D., 2013. Noise alters beta-band activity in superior temporal cortex during audiovisual speech processing. Neuroimage 70, 101–112.

Schubert, T.W., D’Ausilio, A., Canto, R., 2013. Using Arduino microcontroller boards to measure response latencies. Behav. Res. Methods 45, 1332–1346.

Schultz, B.G., van Vugt, F.T., 2016. Tap Arduino: an Arduino microcontroller for low-latency neurointerface response. PLoS One 7 (5), e37534.

Schultz, B.G., van Vugt, F.T., 2016. Tap Arduino: an Arduino microcontroller for low-latency neurointerface response. PLoS One 7 (5), e37534.