NIH Public Access
Author Manuscript
J Neurosci Methods. Author manuscript; available in PMC 2015 December 30.

Published in final edited form as:
J Neurosci Methods. 2014 December 30; 238: 35–42. doi:10.1016/j.jneumeth.2014.09.018.

Multiple Frequency Audio Signal Communication as a Mechanism for Neurophysiology and Video Data Synchronization

Nicholas C. Topper[2], **S.N. Burke**[2], and **A.P. Maurer**[1,2]

[1]Evelyn F. McKnight Brain Institute; Arizona Research Laboratories Division of Neural Systems, Memory and Aging, University of Arizona, Tucson, AZ 85724

[2]McKnight Brain Institute, Department of Neuroscience, University of Florida, Gainesville, Florida 32610

Abstract

BACKGROUND—Current methods for aligning neurophysiology and video data are either prepackaged, requiring the additional purchase of a software suite, or use a blinking LED with a stationary pulse-width and frequency. These methods lack significant user interface for adaptation, are expensive, or risk a misalignment of the two data streams.

NEW METHOD—A cost-effective means to obtain high-precision alignment of behavioral and neurophysiological data is obtained by generating an audio-pulse embedded with two domains of information, a low-frequency binary-counting signal and a high, randomly changing frequency. This enabled the derivation of temporal information while maintaining enough entropy in the system for algorithmic alignment.

RESULTS—The sample to frame index constructed using the audio input correlation method described in this paper enables video and data acquisition to be aligned at a sub-frame level of precision.

COMPARISONS WITH EXISTING METHOD—Traditionally, a synchrony pulse is recorded on-screen via a flashing diode. The higher sampling rate of the audio input of the camcorder enables the timing of an event to be detected with greater precision.

CONCLUSIONS—While On-line analysis and synchronization using specialized equipment may be the ideal situation in some cases, the method presented in the current paper presents a viable, low cost alternative, and gives the flexibility to interface with custom off-line analysis tools. Moreover, the ease of constructing and implements this set-up presented in the current paper makes it applicable to a wide variety of applications that require video recording.

Correspondence: Andrew P. Maurer, Department of Neuroscience, University of Florida, P.O. Box 100244, Gainesville, Florida 32610., drewmaurer@ufl.edu.

Keywords

alignment; behavior; brain; *in vivo* recording; microcontroller

1. Introduction

Systems-level approaches to understanding the function of the brain often involve deriving neurophysiological correlates with behavior. Such experiments include the behavioral outcome of brain stimulation to restore motor control (Popovic et al., 1993), the syncopation of finger movements relative to scalp EEG (Jirsa et al., 2002; Naeem et al., 2012), and the neural correlates of behavior (Olsson et al., 2005; Lansink et al., 2007). Therefore, precise and accurate synchronization between behavioral monitoring equipment and neurophysiology data acquisition components is required in order to derive meaningful conclusions regarding neurobehavioral relationships. While commercial software is available to capture video relative to the same internal clock as the neurophysiology data, it is often expensive and limits the user in their selection of video cameras, the overall number of recording devices that can be used, or the number of computers that the software can concurrently operate on. Furthermore, these acquisition systems may not be as reliable as advertised (Blair et. al., 1997). Researchers have managed to circumvent these limitations through the use of readily available video cameras synchronized to physiological data by simultaneously capturing a common electrical pulse that is sent to the neurophysiology acquisition machine and used to drive an LED that is captured by a recording camera (Stark et al., 2012; Ludwar et al., 2012). An on screen illuminated LED, however, has several limitations including the potential to interfere with animal behavior, the requirement of sacrificing frame space, and the potential for misalignment when a symmetrical, repeating stimulus is used. Moreover, recording equipment, such as headstage tethers, can potentially block the light from the view of the camera as an animal moves about causing a false “blink” to the camera.

As neurophysiology equipment is becoming more affordable, cost-effective, high-resolution methods for aligning behavioral parameters to the acquired neural data are necessary. A previously unexplored means to synchronize behavioral and neural data capitalizes on the use of a common audio signal. Many consumer grade camcorders feature a microphone input that writes to the audio track of the video file at a relatively high sampling rate (~48kHz) compared to video sampling rates (~60Hz). An audio signal can be sent to the microphone input though a hard-wired connection, such that no sound would be audible in the location of the experiment. Therefore, this method provides a more precise, less obstructive signal that can be used for off-line synchronization. This paper outlines a method in which video and physiology data can be synchronized using pulse signals recorded on the audio track of a video file (Figure 1).

2. Material and methods

In order to achieve synchrony between two separate recording devices, the off-line processing algorithm must compensate for sampling frequency differences between the two devices, differences between recording start times, and missing data points (Hoogeboom, 2003). Therefore, the method presented in the current paper, uses a common audio pulse that is synchronously delivered to the camcorder as well as the neurophysiology data-acquisition system. The synchrony pulse has two components: a low-frequency "counting" pulse embedded with an audible, pseudo-random frequency square wave. As the "counting" pulse contains explicit information as to when each input occurred, the algorithm can be seeded with the approximate lag between initial start times and calculate the ratio of sampling frequencies between the two devices. Moreover, the use of iterative matching between common points can account for offset drift due to missed sample points or changes in sampling rate. At the end of analysis, an index is created that temporally links data acquisition samples to the video file with an alignment beyond the precision of the ~60Hz camcorder sampling frequency.

2.1 Apparatus Set-Up

In order to test and improve upon methods of synchronization, an experimental set-up was constructed that consisted of neural data acquisition equipment (Intan RHD2000 Evaluation board; Intan Technologies, Los Angeles, CA), a microcontroller (Arduino Uno R3; Ivrea, Italy) and a video camera with microphone input (Sony Handycam HDR-CX380; Kōnan Minato, Tokyo, Japan; see Figure 2). Pulses of frequency modulated signal transmitting a common pulse containing temporal information (see below) were sent from analog pin 5 of the Arduino Uno R3 Micro-controller to the Intan RHD2000 Evaluation board and the Sony Handycam HDR-CX380. The Arduino analog pin was connected to the left and right stereo microphone input of the camcorder via a modified audio extension cable (Scosche Industries, Oxnard, California). The signal pulse output from the Arduino Uno was also sent to ADC input 8 of the Intan RHD 2000 Evaluation board in order for the signal to be collected relative to the data acquisition system's internal timestamps. Moreover, the Arduino was powered via the +5V output of the evaluation board, and shared a common ground with the Intan system and the sleeve of the audio jack. The Arduino Uno micro-controller acted as a pulse generator, creating the common signal pulse.

2.2 Recording

Prior to recording, the Sony camcorder was set to record at 60 frames per second, progressive, with a default audio sampling frequency of 48,000 Hz. The Intan Evaluation board was programmed to sample at 20,000Hz. The start of each recording session began with powering the Intan system. Once powered, the Arduino Uno controller is initialized and begins generating a synchrony pulse. In testing, the data acquisition and video recordings were initialized and ended asynchronously, with the start times varying 2–3 seconds on average.

2.3 Synchrony Pulse Generation

While a classical method of synchrony involves a repeating pulse of fixed duration and frequency, this is prone to misalignments. For example, if different numbers of pulses are recorded by the video acquisition and data acquisition as a result of one system starting recording late, then synchronizing the data to the first pulse would incorrectly shift the data. As these pulses are symmetrical, the information regarding the correct alignment can only be surmised. In order to avoid such errors, other researchers have employed a pseudo-random pulse to drive a light-emitting diode in order to disambiguate adjacent pulses (Hoogeboom, 2003). Pseudo-random noise, the computer-programmed random switching between low and high binary states, breaks the potential symmetry allowing the signal to be aligned either visually or through the use of a shift-and-correlate method. Being random, however, removes any possibility to embed information into the common pulse. Therefore, in order to advance upon the pseudo-random noise method, the Arduino microcontroller was programmed to generate a signal with an information bearing component, a binary count to convey temporal information, as well as a random component in order to add entropy to the signal (Fig 3). For instance, each pulse series began with one second of silence, followed by a fixed-frequency square wave at 294 Hz, lasting one second. This initialization segment was then followed by a signal with two components. The low frequency component provided an iterative count in the base 2 number system with 16 bits, where each 1 bit is represented by a 60 millisecond square wave window (16.667 Hz). For greater entropy, each 1 bit produces a square wave tone with a randomly determined frequency, ranging between 200 and 300 Hz. Therefore, the low frequency counting signal will produce unique integer values for 65,535, iterations with increased entropy generated by the randomly selected high-frequency components.

The predictable nature of the low frequency information facilitates inspection of the data and visually determining the offset between recording start-times. During development, this was helpful for debugging, and confirming synchrony between datasets. For ease of implementation, the Arduino source code that was used to generate the synchrony signal is available at (https://github.com/ntopper/KitchenSync). ***This software is available under the GNU General Public License, Version 3.***

2.4 Overview of Software Execution

Following data acquisition, data processing begins by paring the audio track from the video using ffmpeg (http://www.ffmpeg.org/). This audio stream is converted into an array of amplitudes in order to be aligned to the analog channel on the acquisition system that recorded the same common pulse (contained as a vector of voltages). A custom written Python code (https://www.python.org/) was constructed that initially parses the audio information from the video file and extracts the Intan analog channel. For the purposes of finding the appropriate alignment between the common pulses, a correlation-matching algorithm (see Using Moving Window Correlation to find Corresponding Data Points) was written using the SciPy library (https://www.scipy.org/). Once the alignment algorithm is completed, the user is provided with an index that relates the pulses of the signal on the neurophysiology data acquisition system to the frames of the video. Therefore, with an array of Intan timestamps and their corresponding video frames, it is possible to interpolate neurophysiological events to camcorder frames for behavioral analysis. This index array is stored as a text file, and as a pickled NumPy array object, a data file that is easily imported by the SciPy or Matlab framework. The Python code for synchronization requires the following parameters: the relative path to the recorded synchrony pulse analog data, the relative path to the recorded video file, the sampling rate of data acquisition, an estimated offset, a confidence interval of this offset, and an optional offset recalculation interval (see Figure 4). This Python script has been made available here. The analog pulse data can be a pickled NumPy array file, or a One File per Channel flat binary file of unsigned 16 bit integers (see Intantech.com Application Note: Data File Formats). The expected video format is MP4. MTS files recoded in the AVCHD format can be concatenated using MTS File Joiner (http://sites.google.com/site/mtsfilejoiner/) and then re-encoded in the mp4 format using MEncoder (http://www.mplayerhq.hu/).

The estimated offset of the video and neurophysiology data streams and associated confidence intervals are expected in seconds, and are used as an estimation to limit the range of the initial offset calculation. For some experimental set-ups, a single alignment iteration may be satisfactory to guarantee behavior-neurophysiology synchrony. If, however, there is a potential for one signal to "drift" relative to another, the algorithm conducts an iterative alignment procedure. In this situation, a user can specify an additional parameter, the offset recalculation interval. This parameter denotes how often, in seconds, the time lag between traces is recalculated in order to compensate for offset drift (see section 3.3).

2.4.1 Determining Relative Sampling Frequencies—The Intan data acquisition system and Sony camcorder audio recording have two different sampling frequencies, 20kHz and approximately 48kHz respectively, therefore, it is necessary to determine the precise relationship between the two. At the onset of the algorithm, the user inputs two estimated sampling rates as an initial "seed". If the sampling rate of the audio recording is not provided, the code will attempt to extract it from the meta-data of the isolated audio file.

In order to calculate the relative sampling frequencies, two common pulses are identified in the Intan and audio traces using a template matching algorithm (described below). It should be noted that, as this alignment method is based on finding common pulses, the neurophysiology acquisition and video recordings do not need cover the same duration of time. The alignment is achieved by taking two audio sub-samples, spaced at a fixed distance apart, and determining the corresponding location of these two sub-samples along the axis of the synchrony pulse data recorded by the acquisition system. The distance, in samples, between the two audio subsamples and the distance, in samples, between the two corresponding data acquisition samples represent the same interval of time. Thus, the relationship between these two sampling frequencies can be expressed as follows:

$$\frac{\Delta audiotemplatestart}{\Delta correspondingwindowstart}=\frac{AudioSamplingFrequency}{AquisitionSamplingFrequency} \quad 1]$$

Next, the ratio between the camera frame rate and the sampling frequency of the audio track is determined. As the video and audio are recorded by the camcorder in parallel, the duration of each recording represents the same interval of time. The relationship between the average video frame rate, and the average audio sampling frequency is expressed as a ratio of the duration of each recording:

$$\frac{VideoDuration(frames)}{AudioDuration(samples)}=\frac{VideoSamplingFrequency}{AudioSamplingFrequency} \quad 2]$$

With these two values, the ratio between the video frame rate and the data acquisition sampling frequency can be determined:

$$\frac{VideoSamplingFrequency}{AudioSamplingFrequency}\cdot\frac{AudioSamplingFrequency}{AquisitionSamplingFrequency}=\frac{VideoSamplingFrequency}{AquisitionSamplingFrequency} \quad 3]$$

2.4.2 Composition of Sample to Video Frame Index—With two parameters, the relative sampling rate, and the time lag between recordings, an index relating data acquisition sample n to a fraction of a video frame can be expressed as follows:

$$VideoIndex=(n-offset_n)\cdot\frac{VideoSamplingFrequency}{AquisitionSamplingFrequency} \quad 4]$$

Where *n* is the index of a data acquisition sample, *Video Intex* represents the correct temporal position of sample n relative to the recoded frames of the video file, $off\ set_n$ is the time offset between data sets calculated at, or near sample *n*. Thus, by precisely calculating two parameters, the relative changing frequency ratio and the time offset calculated at or near each sample, we can construct an index array accurately aligning the recorded physiological data with the time scale of the video file. Our algorithm more accurately determines the values of these parameters using common information recorded on both the audio track of the video file, and the analog input of our data acquisition system, along with a method of matching corresponding data points between both sample arrays (Hoogeboom, 2003). When calculated with floating point precision, this sample-to-frame index array can link data acquisition samples along the time scale of the video recording to a sub-frame level of precision, as it was constructed using the higher sampling frequency of the recorded audio information.

2.4.3 Using Moving Window Correlation to find Corresponding Data Points—In order to appropriately align the two data traces, an iterative "shift-and-correlate" method was used (Hoogeboom, 2003). This method involves identifying a common pulse that was recorded on both devices. While this could be a pseudo-random binary pulse (Hoogeboom, 2003), the current method uses a highly structured binary pulse with random frequency elements embedded within the lower frequency pulse. For one of the data streams, a subsample is isolated which becomes the template. If both devices were recording at different sampling frequencies, the template must be re-sampled to match the approximate sampling rate of the opposing recording device.

A moving window is iteratively shifted along the data set opposing the template. At each position, the Pearson's correlation coefficient is calculated between the template and the data (Fig 5; middle panel). The point at which this R-value is greatest reveals the position of the sample along the opposing dataset corresponding to the sample at the beginning of the template. In Figure 5, the sample offset between the beginning of an audio template and data acquisition window with the greatest correlation was used to create an approximate alignment of the two recordings (Fig 5; bottom panel).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

2.4.4 Using Moving Window Correlation to Construct a Time-lag Array—Rather than calculating the sample-to-frame index using a single time offset, the current algorithm calculates an array of time lags throughout the duration of the recording. This allows one to compensate for offset drift that may be caused by various factors such as missing data points, or inconsistency in the sampling frequencies of our hardware. To construct this offset array, first the audio signal pulse array is re-sampled to match the sampling rate of data acquisition, by calculating the relative sampling frequency ratio, as described above. At each *n*th acquisition sample, *i*, as specified by the user, the time offset between audio and data acquisition is recalculated. First, a data acquisition subsample, starting at sample *i* is isolated. This subsample is treated as a static template, and the audio sample corresponding to the Intan data acquisition at sample *i* is determined using the moving window correlation method, as described above. The difference between the index of sample *i*, and the index of the beginning of the corresponding window reflects the time offset at sample *i*. This offset is appended to an array of offsets, and repeated *n* times. This yields an offset array with a calculated offset for each data acquisition sample. With these two parameters, the ratio between data acquisition sampling frequencies, and the iteratively calculated time-lag array, a sample-to-frame index can be constructed, as described in equation 4.

3. Results

3.1 Evaluating the Calculated Sampling Rate Ratio

The accuracy of the calculated sampling rate frequency ratios is critical to the accuracy of the constructed sample-to-frame index. To test the accuracy and precision of the sampling frequency ratios calculated, six videos were recorded, with a frame rate setting of 60 frames/s, progressive, and a data acquisition sampling frequency of 20,000 Hz. For each sample, the ratio of video to data acquisition sampling frequencies was calculated and multiplied by 20,000 Hz, yielding an approximation of the calculated video frame rate. The average frame rate approximation across the six tests was calculated to be 59.939 frames/s (standard deviation: 7×10^{-4} frames/sec). While the camcorder was set to record at 60 frames/s, this is often a ‘shorthand’ for the NTFC standard of 59.94 frames/s, a legacy frame rate consistent with the standard field refresh rate from the early days of color television still persistent in the video industry (Hewlett Packard, 1994). Therefore, the Sony camcorder used in testing does not record precisely at the manufacturer specified frame rate of 60 frames/s, but rather at the more common standard of 59.94 frames/s. Rather than assuming the manufacture specified frame rate to be true, the method described in section 2.4.1 yielded a more precise measure of framerate, which is of great importance for proper synchrony.

3.2 Audio Pulse Input vs Video Pulse

This off-line synchrony technique differs from other methods in that it takes advantage of the higher sampling rate of the audio input of the camcorder to make more accurate time-lag calculations. Traditionally, a synchrony pulse would be recorded on-screen via a flashing diode. The typical consumer grade camera frame rate of ~60 frames/second therefore limits the accuracy of the recording of the pulse event to about 16.7ms. By recording at a higher frequency, the audio track is able to detect precisely when, within a frame of video, a pulse event occurred.

In order to compare the LED and audio methods of synchronization, a series of videos were recorded, in which an Arduino micro-controller simultaneously illuminated an LED, and generated a synchrony pulse signal as described above. A Sony Handycam HDR-CX380 camcorder recorded the pulsing LED at a frame rate of 59.94Hz, and the audio signal was recorded on the microphone input of the camcorder at 48,000Hz. Both the LED and audio data streams were recorded as separate voltages by the Intan RHD2000 data acquisition system at 20kHz. OpenCv (http://opencv.org/) was used to create an array of pixel intensity per frame at the location of the LED in order to extract pulse-time information.

LED voltage and audio signal voltage were synchronized to the data extracted from the camcorder video file (Figure 6) using the algorithm described above. The LED voltage was aligned in accordance to the audio data, revealing the true temporal position of the LED pulse event along the time scale of the video recording. While both pulses were generated simultaneously, the higher sampling rate of the audio input of the camcorder enabled the pulse to be detected with greater precision. In this way, the audio track was able to reveal the time at which the event occurred within the exposure of a single frame. The sample to frame index constructed using the audio input correlation method described above enables video and data acquisition to be aligned at a sub-frame level of precision.

While a pulsing LED as a form of synchrony is less precise than the proposed method, one could feasibly use a pulsing LED in conjunction with the audio track synchrony method to confirm that the software/hardware combination to be used in experimentation and analysis has proper synchrony between devices. Included with the software provided are instructions to replicate the experiment performed above, and software perform statistical analysts on the latency between pulses detected by video, and by an acquisition system. Analysis on the test dataset showed that pulses were detected by the acquisition system an average of ~0.005 seconds into the first video frame in which they were detected, with a standard deviation of ~0.006 seconds. The aligned video and audio traces had a correlation coefficient of 0.9493.

3.3 Sampling Rate Variability and Temperature

While testing the software implementation of a single point alignment (as opposed to an iterative alignment method), a common pattern was observed in the calculated time lag arrays. That is, there was a consistent, iterative drift between the start times of each alignment point. Visually inspecting our time-lag curves (figure 7), we suspected that they may be related to the camera’s internal temperature altering the frequency of the internal Sony Handycam quartz oscillator. There is a well-documented association between crystal oscillators and temperature (Walls and Gagnepain, 1992) to the point that many electronics account for this effect via the implementation of temperature compensated crystal oscillators (Ueno, 1996). Therefore, four recordings were taken while simultaneously recording the external temperature of the camcorder with a no-contact infrared thermometer (Infrared Thermometer - MLX90614; Sparkfun, Boulder, Colorado), All of the recordings, save one, began with the camera at room temperature. Prior to one of the recording sessions, the camera was placed into Steel Compact Refrigerator Mini Fridge (Midea, Shunde, Guangdong, China) for four hours, cooled to 11 °C. The thermometer was positioned to measure at a relatively warm part of the camera, behind the LCD screen near the HDMI output port. The synchronization pulses were also collected for the duration of the recordings in order to calculate the offset. During analysis, one dataset was too short in duration and therefore dropped from analysis.

There was a −0.78 Pearsons correlation R-value between the average calculated offset, and average temperature. A polynomial fit for both datasets was also calculated. The discrete derivative of the calculated offset polynomial fit, and the temperature polynomial fit showed a statistically significant relationship ($r = 0.98$, $p < 0.001$). While it was not directly possible to measure the temperature or frequency of the quartz oscillator itself, measurements of the external camera temperature and offset drift did in fact reveal a parabolic relationship. Figure 8 shows the relationship between calculated offset drift (from the minimum calculated offset), and the measured temperature of the camera. As expected, these tests all had a critical temperature at which the calculated offset begins to climb back towards the initial value. The average temperature recorded from the exterior of the camera at the lowest calculated offset for each of the three trials was 27.47°C (standard deviation: ±0.453°C). Given this relationship, internal changes in temperature most likely account for the variability in the sampling frequency. While it may be feasible to control the internal temperature of the camcorder with modification of hardware, the current algorithm is able to detect and account for this offset drift by iteratively re-calculating the signal pulse lag throughout the duration of the video.

3.4 A Note on Accuracy of Alignment

The method proposed uses events recorded by sound to synchronize data with events recorded as video, and it is therefore important that the audio track to be properly synchronized with the video. The required level of accuracy depends on the nature of the analyses to be performed, and this is to be determined by those performing the video analysis. Because of the intended application of consumer grade camcorders, it is reasonable to expect at least a “lip synchrony” level of accuracy. In other words, a camcorder recording a human talking will typically have enough accuracy that there will not be a noticeable latency between lip movements and sound. For those using consumer grade equipment who depend on a great deal of precision, there are some factors that should be taken into account.

One limitation in using low cost video equipment is the way Complementary Metal–Oxide–Semiconductor (CMOS) sensors typically found in digital camcorders record images. A single frame recoded by a CMOS sensor chip does not represent a single snapshot in time. Images are recorded by a “rolling shutter”, the sensor chip scans the image row-by-row (typically) or column-by-column (Hedborg, 2011). Each scanline of pixels is exposed at different times. This means different part of the image will have been recorded at slightly different times. This is why fast moving objects are often skewed, or flashing lights are sometimes visible in only part of a video frame. The rolling shutter effect is a well-known issue, and is generally considered a compromise made by those who choose to record using CMOS image sensor cameras as opposed to more expensive Charge-Coupled Device (CCD) cameras.

Without knowing exactly what analysis is to be performed on a given video, the proposed method simply calculates a “sample to frame” index along the timescale of the video file recorded. This allows flexibility for the analyst to compensate for any issues that they may have with the recorded images as they deem appropriate. The extent of issues caused by the rolling shutter depends on a number of application specific factors, such as the frame rate, resolution, exposure time, the velocity of the subject, the nature of the video analyses performed, and the needed level of precision. Other factors, such as artifacts caused file format converters, could cause a misalignment. With each hardware and software combination, it is recommend that those implementing the current software perform a test similar to that in section 3.1,

4. Conclusion

To synchronize data recorded between two devices, one must account for differences in sampling frequency, the time offset between recordings, and any missing data points or variability in sampling frequency. The method presented in the current paper more precisely calculates these required parameters and, the iterative offset drift is mitigated by calculating a sampling rate ratio between recording devices. Moreover, offset drift caused by missing data points, or by small variability in sampling frequency can be taken into account by iteratively re-calculating the offset throughout the duration of the video and neurophysiology recording. Furthermore, with the new method of data synchrony, the erroneous alignment of pulses is not likely because each cycle of pulses contains a pseudo-random high-frequency square wave signal nested within a unique, low frequency “counting” signal. Finally, this method can be extended to add multiple cameras that can be synchronized relative to neurophysiological data. It should be noted, however, that while this method enables users to align electrophysiology events along the timescale of a video with sub-frame accuracy, in most cases the resolution of the behavior determined by the camera will not exceed the rate that the video frames are collected.

Conventional methods of video synchrony with non-specialized equipment require that part of the video frame is sacrificed to a pulsing light. This light can potentially interfere with the experiment, and elements of the experiment may inadvertently block the light. This also limits movement of the camera to keep the light within the frame in the event that more sophisticated methods of pan-and-tilt video-tracking were to be implemented. Synchrony methods with a stationary pulse-width and frequency may also not take into account any cumulative error or missing data points, and a fixed-interval pulse without an appropriate level of entropy may lead to the matching up of erroneous pulses. Furthermore, synchrony methods that depend on the detection of events recorded by the image sensor of the camcorder limit precision to the frame rate of the camcorder.

An alternative solution to some of the limitations of the conventional LED method would be the use of specialized scientific cameras. Many of these cameras are capable of high frame rates, the output of pulse signals marking the beginning of a frame, and even the on-line analysis and recording of data. High speed recordings are excellent for experiments involving low-duration recordings, but are not particularly viable for long-term recordings characteristic of behavioral studies, due to limitations in data storage space. Consumer grade camcorders are capable of recording at resolutions within the realm of behavioral variability (<20 ms), and are well suited for long-term experiments. While on-line analysis and synchronization using specialized equipment may be the ideal situation in many cases, this method presents a viable, low cost alternative, and gives the flexibility to interface with custom off-line analysis tools.

Acknowledgments

This work was supported by the McKnight Brain Research Foundation. NIH Health Grant NS070464, and the UF Research Opportunity Seed Fund 00115480. We would like to thank Reid Harrison for helpful comments on an earlier version of the manuscript.

References

Blair HT1, Lipscomb BW, Sharp PE. Anticipatory time intervals of head-direction cells in the anterior thalamus of the rat: implications for path integration in the head-direction circuit. Journal of Neurophysiology. 1997; 78(1):145–59. [PubMed: 9242269]

Hewlett Packard. Cable Television System Measurements Handbook. Santa Rosa, California: Hewlett-Packard Company; 1994.

Hedborg, J.; Ringaby, E.; Forssén, P.; Felsberg, M. Structure and Motion Estimation from Rolling Shutter Video. IEEE International Conference on Computer Vision Workshops; Nov 17–23, 2011; 2011.

Hoogeboom, PJ. Off-line synchronisation of measurements based on a common pseudo-random binary signal. National Aerospace Laboratory NLR;

Jirsa VK, Jantzen KJ, Fuchs A, Kelso JA. Spatiotemporal forward solution of the EEG and MEG using network modeling. IEEE Trans Med Imaging. 2002 May; 21(5):493–504. [PubMed: 12071620]

Lansink CS, Bakker M, Buster W, Lankelma J, van der Blom R, Westdorp R, Joosten RNJMA, McNaughton BL, Pennartz CMA. A split microdrive for simultaneous multi-electrode recordings from two brain areas in awake small animals. Journal of Neuroscience Methods. 2007; 162:129–138. [PubMed: 17307256]

Ludwar, BCh; Evans, CG.; Cropper, EC. Monitoring changes in the intracellular calcium concentration and synaptic efficacy in the mollusc Aplysia. J Vis Exp. 2012 Jul.(65):e3907. [PubMed: 22824826]

Naeem, M.; McGinnity, TM.; Watson, D.; Wong-Lin, K.; Prasad, G.; Kelso, JA. Inter-brain Mutual Information in Social Interaction Tasks 2012 Second International Workshop on Pattern Recognition in NeuroImaging; 2012 July, 25–28;

Olsson RH 3rd, Buhl DL, Sirota AM, Buzsaki G, Wise KD. Band-tunable and multiplexed integrated circuits for simultaneous recording and stimulation with microelectrode arrays. IEEE Trans Biomed Eng. 2005; 52(7):1303–11. [PubMed: 16041994]

Popovic, DB.; Stein, RB.; Jovanovic, KL.; Dai, R.; Kostov, A.; Annstrong, WW. Sensory Nerve Recording for Closed-Loop Control to Restore Motor Functions IEEE Transactions On Biomedical Engineering; 1993 October; p. 1024-31.

Olsson, Roy H., III; Buhl, Derek L.; Sirota, Anton M.; Buzsaki, Gyorgy; Wise, Kensall D. Band-Tunable and Multiplexed Integrated Circuits for Simultaneous Recording and Stimulation With Microelectrode Arrays. IEEE Transactions On Biomedical Engineering. 2005 Jul; 52(7):1303–11. [PubMed: 16041994]

Stark E, Koos T, Buzsaki G. Diode-probes for spatiotemporal optical control of multiple neurons in freely-moving animals. Journal of neurophysiology. 2012 Apr.:349–363. [PubMed: 22496529]

Ueno Y. Temperature-Compensated Crystal Oscillators. Electronics and Communications in Japan. 1996; 79(2):507–19.

Walls FL, Gagnepain JJ. Environmental Sensitivities of Quartz Oscillators. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control. 1992 Mar; 39(2):241–9.

Highlights

- Current methods for aligning neurophysiology and video data are either prepackaged, requiring the additional purchase of a software suite, or use a blinking LED with a stationary pulse-width and frequency.
- A cost-effective means to obtain high-precision alignment of behavioral and neurophysiological data can be obtained by generating an audio-pulse embedded with two domains of information, a low-frequency binary-counting signal and a high, randomly changing frequency.
- The sample to frame index constructed using the audio input correlation method described in this paper enables video and data acquisition to be aligned at a sub-frame level of precision.
- The higher sampling rate of the audio input of the camcorder enables the timing of an event to be detected with greater precision than traditional LED methods.

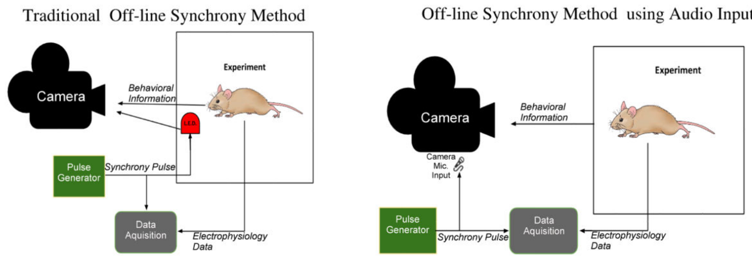

Figure 1.

Conventional methods of off-line video data synchrony make use of a signal pulse which simultaneously illuminates an on-screen light emitting diode, while being recorded on a data acquisition system. The temporal position of pulse events recorded in the video and data acquisition system can be used to correlate behavioral and physiological data recorded on both devices. This method requires that frame space is sacrificed to the LED, and has the potential to interfere with animal behavior. Furthermore, a low entropy pulse runs the risk of erroneous alignment, and the precision of physiological data along the time scale of the video is limited by the frame rate of the camcorder. The method proposed in this paper circumvents the use of an on-screen LED by recoding a synchrony signal along the audio track of the video file and as an analog signal recorded. The higher sampling rate of audio recording can be taken advantage of to achieve a higher precision alignment between the two datasets. The signal generated carries temporal information, and can be used to calculate a set of parameters used to construct an index placing individual samples along the time scale of the video recording.

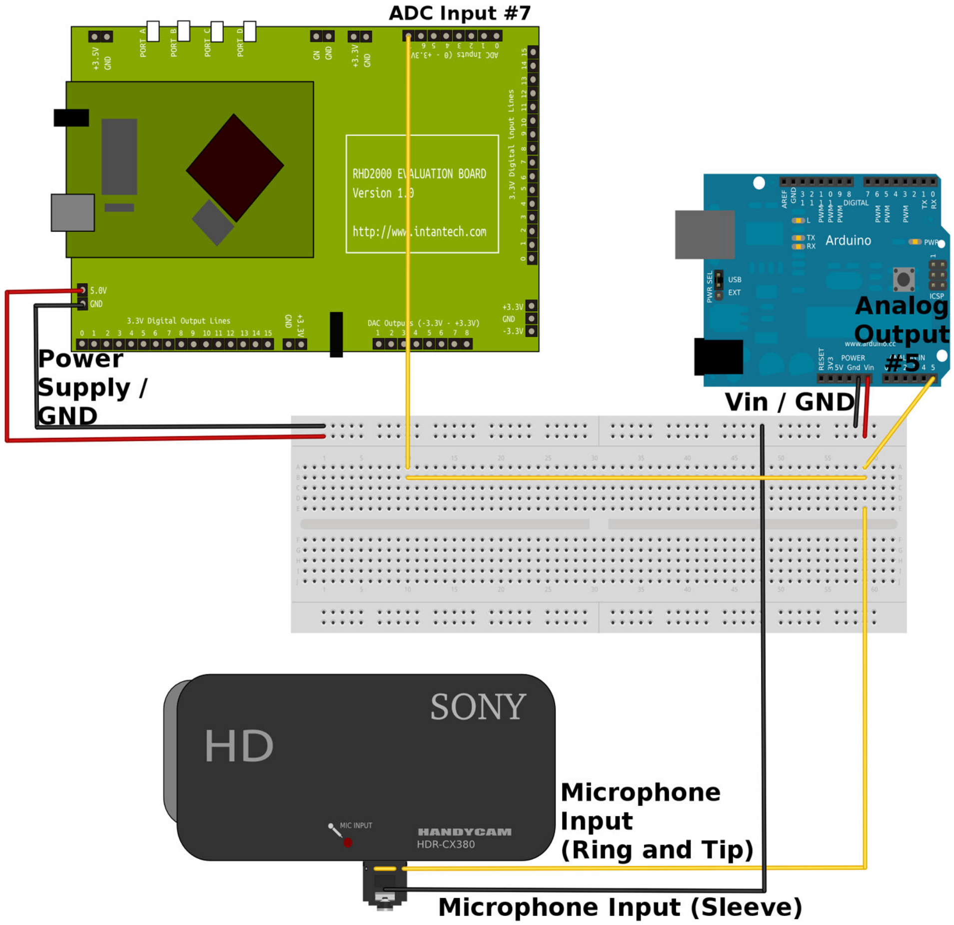

Figure 2.

Schematic of Hardware set-up for audio synchronization of neurophysiology and video data. A synchrony signal is generated on pin 5 of an Arduino micro-controller (see Fig 3), and recorded on both ADC channel 8 of an RHD2000 evaluation system as an analog voltage, and on the microphone input of a Sony Handicam as an audio voltage.

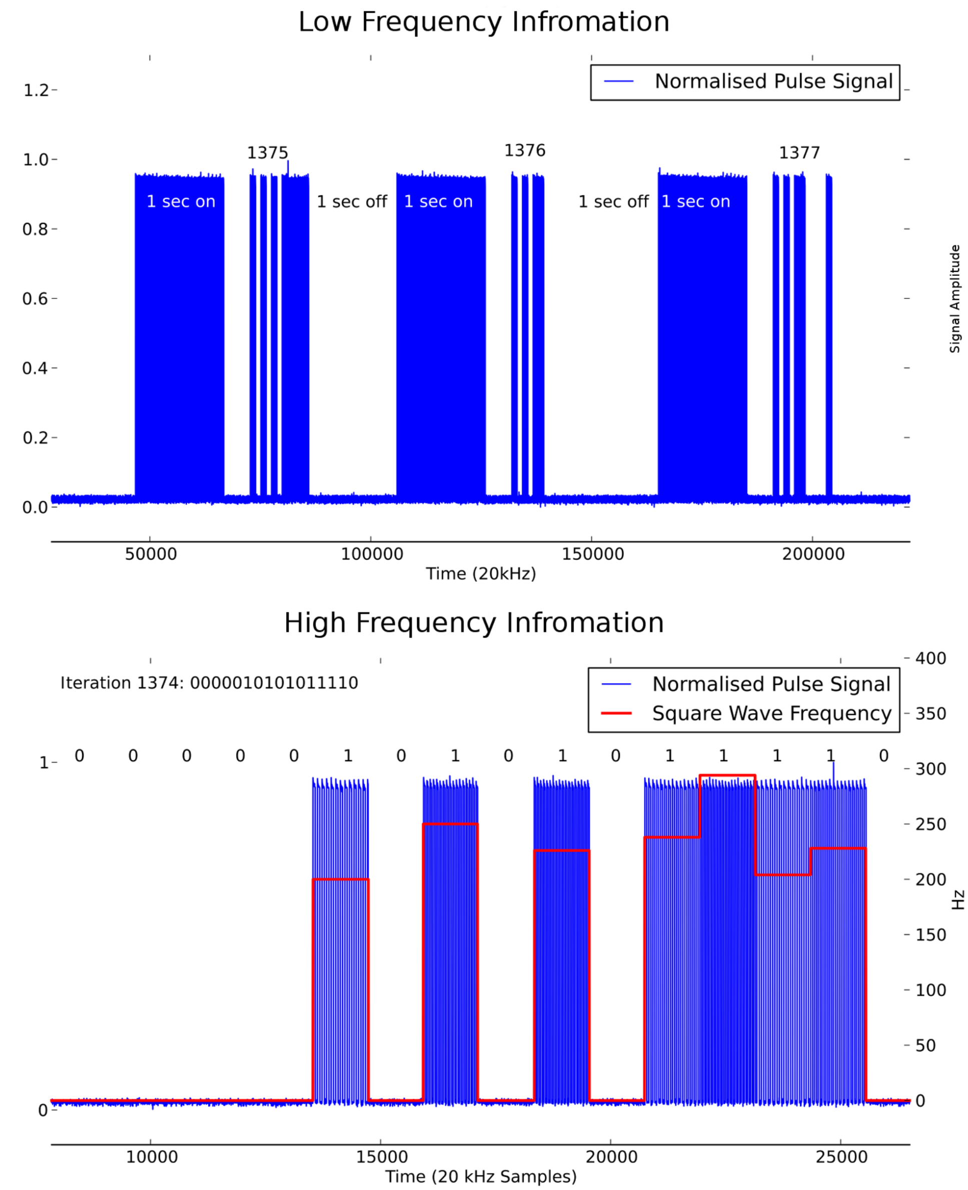

Figure 3.

Examples of low and high frequency information in audio pulses that are used for video and neurophysiology synchronization. Each iteration of low frequency information carries a sequential “number” represented in the base 2 number system. Each “1” bit is represented as a 60 millisecond square wave window. The frequency of these square waves are modulated randomly to create greater entropy.

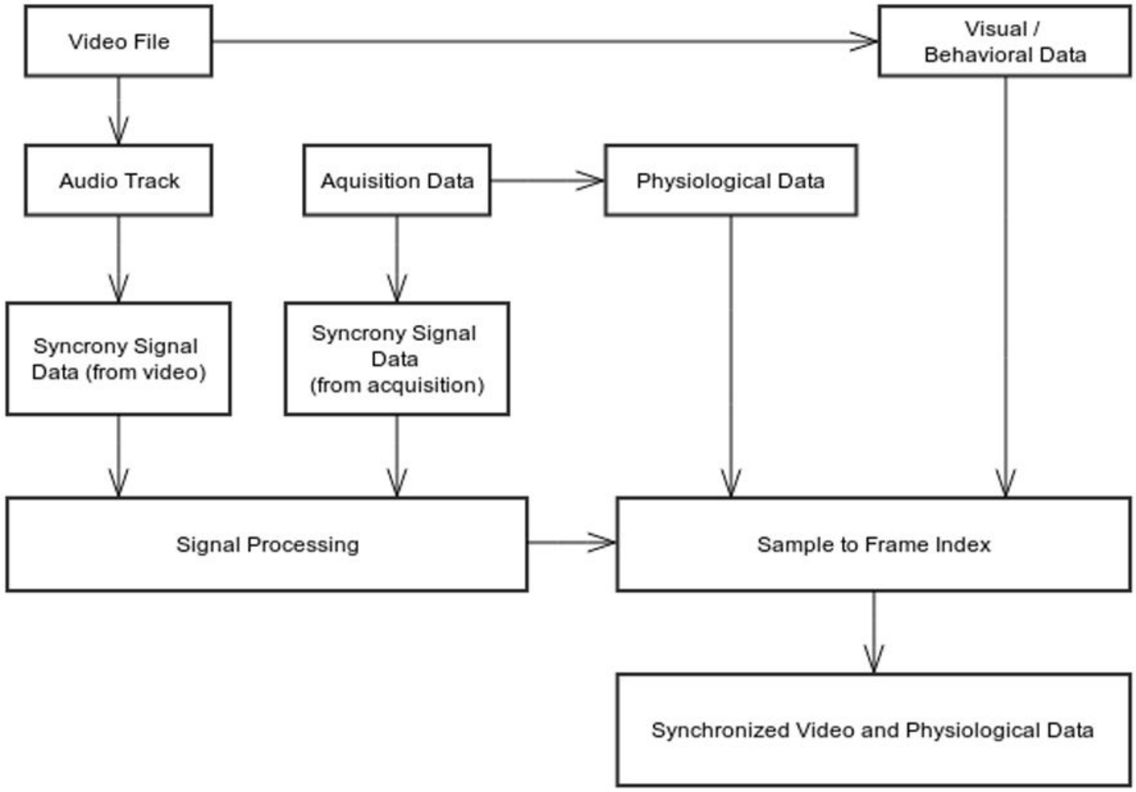

Figure 4.

From the video file, the audio track is isolated, and converted into an array of amplitudes per sample. From the acquisition data, the recorded synchrony signal is isolated as an array of voltages per sample. These two signal traces are processed, and used to generate a sample to frame index. This index can be used to align the physiological data collected from the acquisition system with the information collected from the frames of the video file.

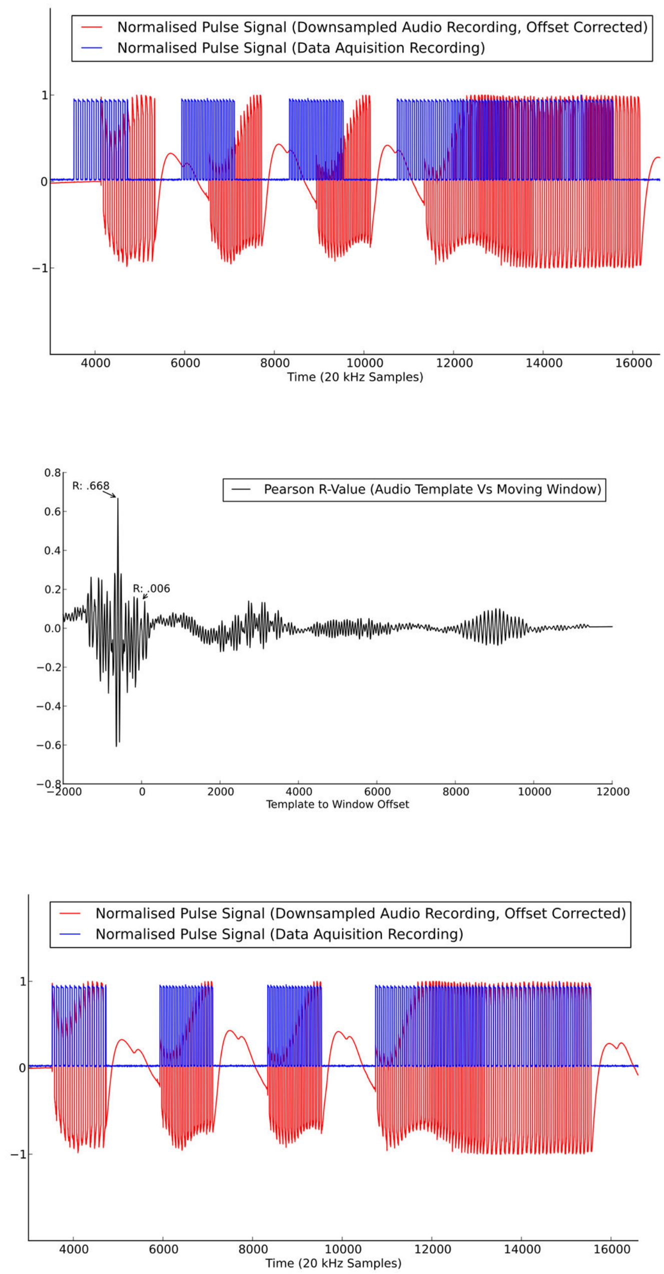

Figure 5.

A synchrony pulse is recorded on both the audio track of the video file, and as an analog signal on the Intan RHD 2000 evaluation system (Top Panel). Both of these traces are re-sampled to approximately the same sampling rate to enable a sub-sample of one trace to be shifted along the other. The Pearson correlation coefficient (R) between the sub-sample and the opposing dataset at that point is then calculated (center panel). The temporal position of the sub-sample yielding the greatest R-value reveals the point along the opposing dataset that corresponds to the point at the beginning of the sub-sample. This information can be used to shift the dataset into an approximate alignment (bottom panel).

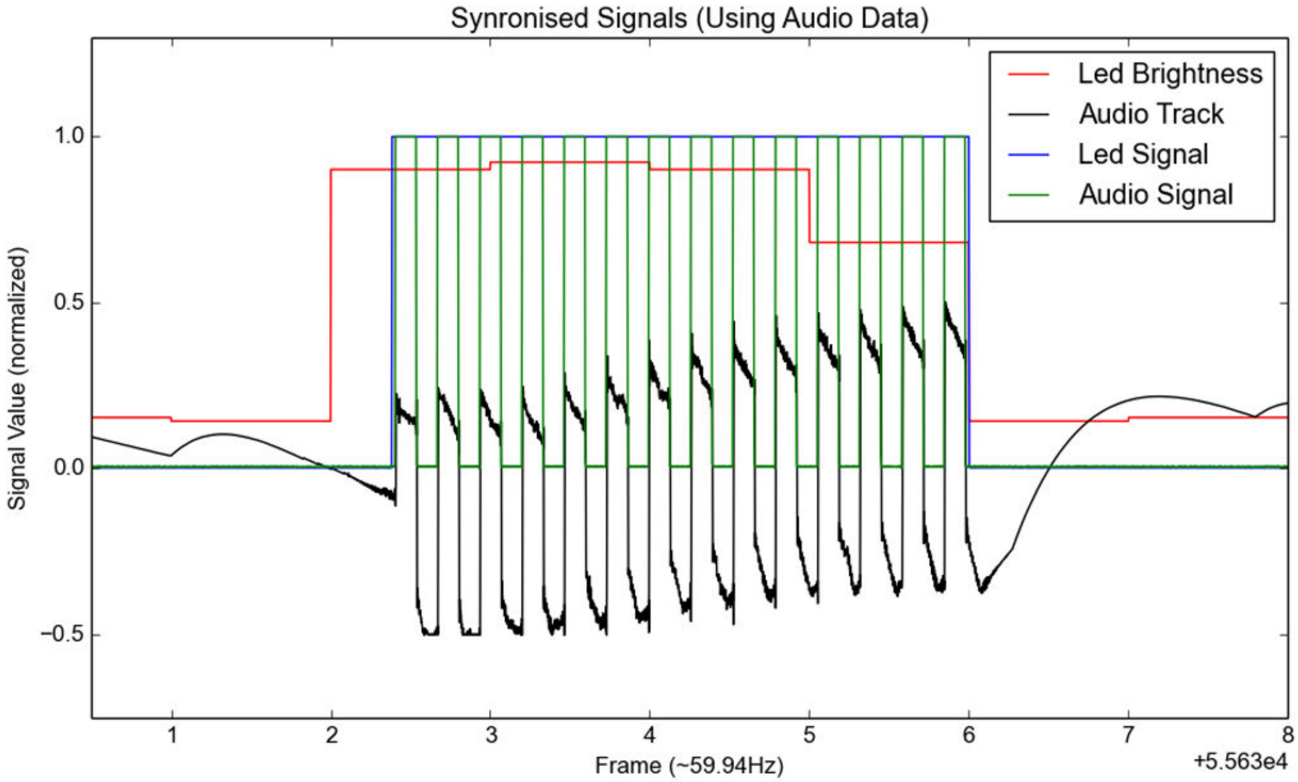

Figure 6.

A signal generator simultaneously illuminated an LED to generate a signal pulse, as described in Fig 3. These signals were recorded by the camcorder both as visible light, as an audio signal, and by the Intan data acquisition system as analog voltages. These signals were aligned using the method described in this paper see equation 4. By aligning these signals in accordance with the recorded audio signal, the true temporal position of the LED event is relieved. In this way, our method allows us to align data recorded on a data acquisition system with a video file with a sub-frame level of precision. Note that, due to a lower sampling frequency, the audio track signal was detected before the illuminated LED appears in a video frame.

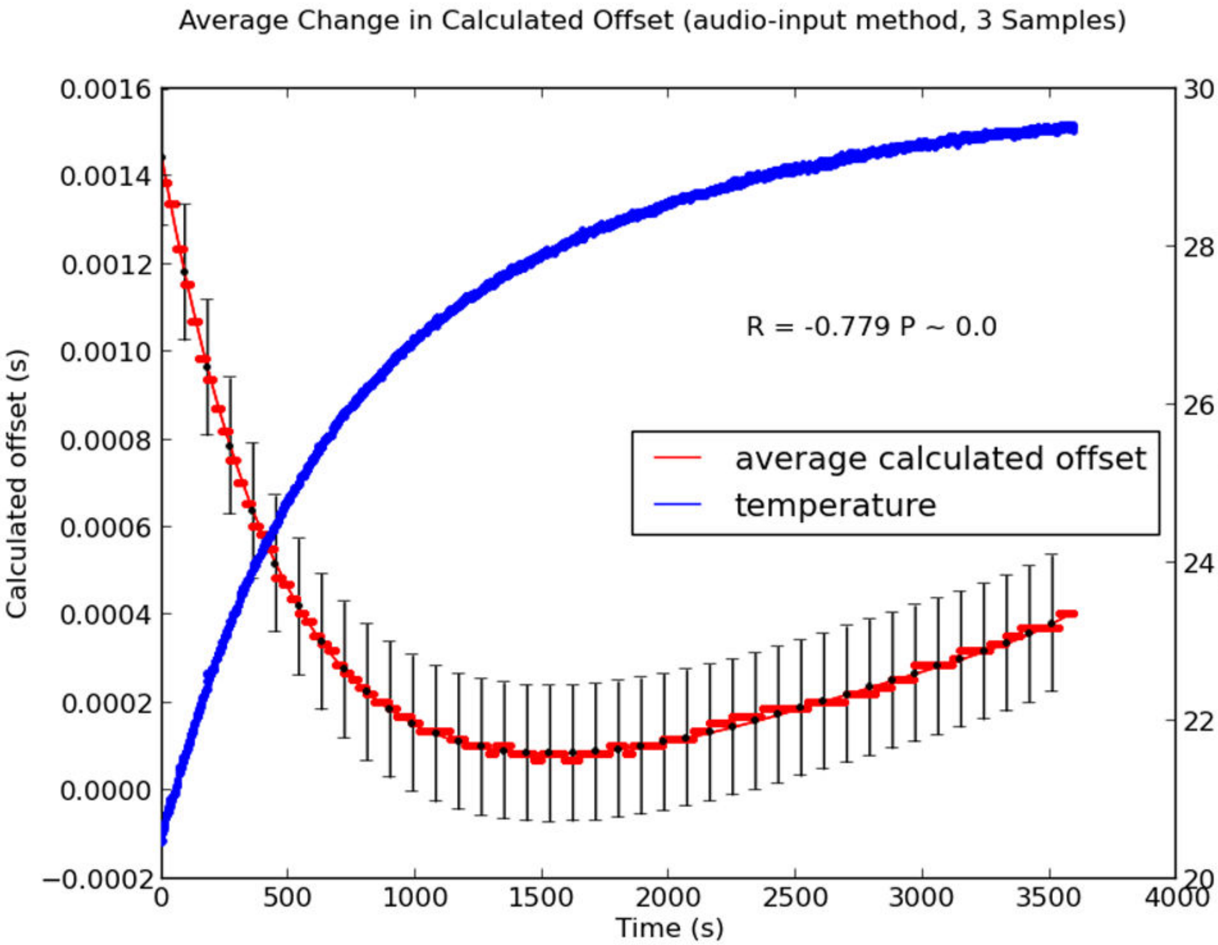

Figure 7.

The external temperature of a camcorder was recorded during several tests of the algorithm outlined in section 2.7. The average of calculated time lag arrays across 3 trials, and the average temperature yielded a high negative correlation (R = −0.78). The discrete derivative of a polynomial fit of the calculated offset array showed a much higher correlation to a polynomial fit of the average temperature (R = 0.98, p < 0.001). This algorithm was able to detect, and account for what appears to be an offset drift caused by self-induced thermal change within the camera.

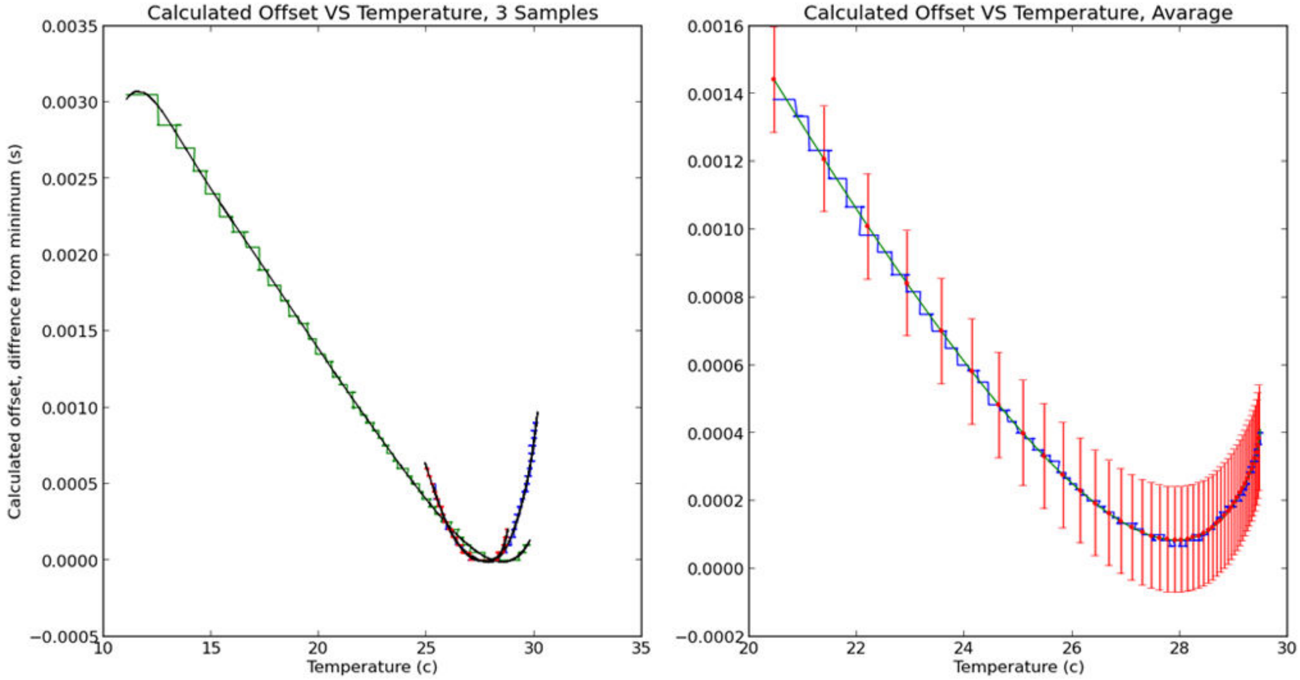

Figure 8.

The calculated offset, from the lowest value, was plotted as a function of temperature. This revealed a parabolic relationship between camera temperatures and offset. Note that in all three samples the calculated offset begins to climb back towards the intimal value at around the same temperature. The average temperature recorded from the exterior of the camera at the lowest calculated offset for each of the three trials was 27.47°C (standard deviation: ±0.453°C).