34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

A Real-Time Detector System for Precise Timing of Audiovisual Stimuli

Andreas Henelius, Sharman Jagadeesan and Minna Huotilainen

Abstract— **The successful recording of neurophysiologic signals, such as event-related potentials (ERPs) or event-related magnetic fields (ERFs), relies on precise information of stimulus presentation times. We have developed an accurate and flexible audiovisual sensor solution operating in real-time for on-line use in both auditory and visual ERP and ERF paradigms. The sensor functions independently of the used audio or video stimulus presentation tools or signal acquisition system. The sensor solution consists of two independent sensors; one for sound and one for light. The microcontroller-based audio sensor incorporates a novel approach to the detection of natural sounds such as multipart audio stimuli, using an adjustable dead time. This aids in producing exact markers for complex auditory stimuli and reduces the number of false detections. The analog photosensor circuit detects changes in light intensity on the screen and produces a marker for changes exceeding a threshold. The microcontroller software for the audio sensor is free and open source, allowing other researchers to customise the sensor for use in specific auditory ERP/ERF paradigms. The hardware schematics and software for the audiovisual sensor are freely available from the webpage of the authors' lab.**

I. Introduction

In the fields of cognitive and naturalistic neuroscience, the recording of event-related potentials (ERPs) and event-related magnetic fields (ERFs) are common techniques used in the study of brain activity. Brain signals, such as an electroencephalogram (EEG) or a magnetoencephalogram (MEG), are recorded during the presentation of different types of stimuli. All waveforms, such as ERPs and ERFs, used to study the response to the audiovisual stimuli are most dependent on precise temporal knowledge of the stimuli. The ERPs and ERFs are studied on the millisecond time scale, and hence even small perturbations in stimulus presentation times can potentially have a large effect on the studied phenomena.

The time of occurrence of stimulus presentation needs to be communicated to the signal acquisition system so that the stimulus events are recorded in time-synchrony with the biosignals. Traditional measurements in cognitive neuroscience using auditory or visual stimuli are made without the use of external sensors, and the data analysis is performed using only markers sent by e.g. the stimulus presentation software to the signal acquisition system. In this scenario there is a risk of signal degradation due to imprecise timing of the markers with respect to the stimuli. In most cases, the triggers sent by the software do not reflect the true stimulus presentation time, e.g. due to processing latency in the stimulus presentation computer or due to the speed of the communications bus. It is also possible that an auditory or visual stimulus is not presented at all for some reason, although the software sends a trigger for the stimulus. These factors lead to more noise in the ERP/ERF waveforms. It is also recommended in guidelines for ERP research that the method by which the ERPs are time-locked to the stimuli should be reported [1]. Robust hardware sensors remedy these problems, providing the researcher with the true time of occurrence of the auditory or visual stimuli to which the ERPs or ERFs are time-locked.

All authors are with the Brain Work Research Centre,
Finnish Institute of Occupational Health, Helsinki, Finland.
Corresponding author: Andreas Henelius
Email: andreas.henelius@ttl.fi
Address: Finnish Institute of Occupational Health
Topeliuksenkatu 41 a A, 00250 Helsinki, Finland

This study was supported by the SalWe Research Programme for Mind and Body (Tekes – the Finnish Funding Agency for Technology and Innovation grant 1104/10)

Many new ERP paradigms use complex auditory stimuli, e.g. [2], [3]. Instead of simple tones, the stimuli consist of multipart words, music as well as continuous and natural sounds. The stimuli are hence becoming more complex, but the underlying need for exact timing still persists. The audio sensor described here was developed primarily for use in auditory ERP/ERF paradigms with multipart auditory stimuli. Multipart stimuli consist of several parts, with a low signal amplitude between the parts, e.g. a sound of the form "ta–ta". This is in contrast to monopart stimuli, such as a simple tone ("beep"). The audio sensor operates in real-time and utilises a programmable digital microcontroller making it possible to adjust the function of the sensor for specific auditory paradigms. The sensor measures voltage fluctuations in the electric audio signal from the sound source and outputs a square wave of adjustable width when the voltage exceeds a set level. The main adjustable parameters are the detection threshold, the width of the output pulse from the sensor and the dead time of the sensor. The adjustable dead time removes the need for offline detection of erroneously marked auditory stimuli that arises if only simple thresholding is used. The analog photosensor circuit is based on a photodiode and can be used on-line during the presentation of stimuli to detect changes in light intensity on the screen occurring in synchrony with visual stimuli. The circuit outputs a waveform reflecting the changes in light intensity on screen.

To the best of our knowledge, there is no readily available device offering the possibility to adjust the dead time of the audio sensor making it directly suitable for ERP/ERF paradigms with multipart stimuli. The current sensor solution was independently developed in our lab, but two commercial products were brought to our attention afterwards. Examples of two commercial products offering similar functionality as the audiovisual sensor presented here are the Serial Response Box by Psychology Software Tools Inc and the StimTracker system by Cedrus Corporation. The Serial Response Box does not permit on-line recording of visual events on the screen using a photosensor and requires use of the E-Prime stimulus presentation software. Neither does it permit audio signals from the sound source to be passed through. The StimTracker system allows pass-through of the audio signal for stimulus detection and permits the use of up to four photosensors. This system is also independent of the stimulus presentation software. However, neither of these systems permit the dead time of the audio sensor to be adjusted.

II. Hardware Design

A. Audio Sensor

The audio sensor is housed in a plastic case (Fig. 1), with the following connectors: **(1)** audio input, **(2)** audio output, **(3)** a communications port providing power to the sensor and output signal from the sensor to the measurement device. The audio input signal can be any mono or stereo signal, e.g. from the audio card of a stimulus computer, or the signal from a microphone monitoring the overall sound level in the measurement room. However, a signal obtained directly from the sound source will be cleaner, as a microphone is susceptible to ambient noise in the measurement room. The audio sensor is equipped with a liquid crystal display (LCD) displaying status messages to the user. A turnable knob connected to a potentiometer is used to adjust the dead time of the sensor. A pushbutton controls the backlight of the LCD screen. The audio sensor operates using a split supply of $\pm$ 5 V direct current (DC), and a +10 V DC power supply can hence be used to power the circuit. The audio sensor consumes less than 50 mA when operating with the LCD screen off. It is possible to operate the circuit using a suitable battery, allowing the sensor to be used also in electrically shielded environments without introducing alternating current (AC) sources that could interfere with signal acquisition.

The audio sensor uses a Teensy 2.0 microcontroller (PJRC.com LLC, Sherwood, OR, USA). The Teensy 2.0 uses an 8-bit Atmel AVR ATmega32U4 microcontroller. A flowchart of the operation of the audio sensor is shown in Fig. 2. The left and right channels of the stereo audio signal are processed independently, using identical processing stages. In the first stage, the signal is passed through an analog precision full-wave rectifier (absolute value circuit), since the analog-to-digital converter (ADC) of the microcontroller can only read positive voltages in the range 0 V to 5 V. In the second stage, the rectified left and right audio signals are summed and fed to an analog input of the microcontroller. The audio channels are summed so that the audio sensor may be triggered by a mono stimulus from either audio channel. Also, it is thus necessary to sample only one signal with the microcontroller. The signal is digitised at 10-bit resolution with a sampling rate of approximately 50 kHz according to our tests, after setting the ADC prescaler to 16 [4]. The detection of audio pulses and corresponding output of a marker pulse is realised in software after this stage.

The output of the audio sensor is a square wave with a width of 20 ms (adjustable), which allows for reliable detection of the square wave at EEG sampling rates of 500 Hz. For lower sampling rates it is advisable to use a longer pulse width for the square wave. The default amplitude of the marker signal is 5 V. This signal is possible to feed to a high-level voltage input found on several signal acquisition devices. Alternatively, the level of the output signal can be lowered and then recorded as a bipolar channel on the signal acquisition device.

B. Photosensor

The photosensor consists of a photodiode that is fastened to the computer screen. In our implementation, the photodiode was encapsulated in Blu-Tack, allowing only the light from the screen area covered by the photosensor to reach the photodiode. The Blu-Tack was coloured black to avoid distracting the subject. In an experiment, a small area of the screen (e.g. a corner region) is dedicated to the photosensor, and the colour of this area is varied, producing a signal for the photosensor. The intended use is for the sensor area on the screen to be normally black, but set to white when a stimulus is presented. This creates a change in brightness that is converted into a voltage fluctuation. The time the output signal is high equals the time the sensor area is white. This makes it possible to measure how long a stimulus was visible on the screen.

The photosensor is realised using an analog circuit consisting of two stages, shown in Fig. 3. In the first stage, the current from the reverse-biased photodiode operating in photoconductive mode is transformed into a voltage signal using a transimpedance amplifier. In the second stage, the voltage signal is fed to a Schmitt trigger circuit with preset threshold values, generating a sharp increase in voltage. The Schmitt trigger will keep the output high as long as the voltage exceeds the threshold. The output of the photosensor is hence dependent on the light intensity of the image shown on screen. The photosensor has no adjustable parameters and is a hardware-only solution. The photosensor operates using a split supply of $\pm$ 10 V.

III. Software Design

The software used in the audio sensor was developed in C++ using the Teensyduino [5] add-on for Arduino. This choice was made to make the source code for the audio sensor easily modifiable by other researchers. The software development tools required for programming the microcontroller are free and available for GNU/Linux, Mac OS X and Microsoft Windows. The Teensy microcontroller is programmed by connecting it directly to the computer's USB port, requiring no additional hardware.

The three adjustable parameters of the audio sensor are **(1)** the detection threshold, **(2)** the duration of the output pulse and **(3)** the maximum total stimulus duration (dead

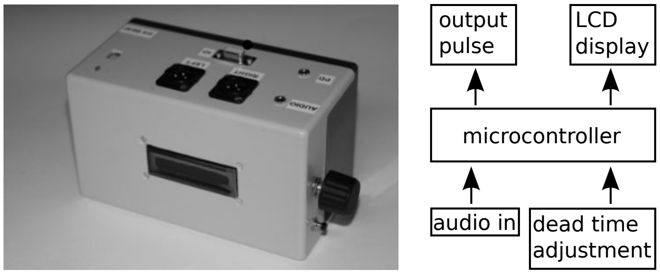

Fig. 1: The audio sensor.

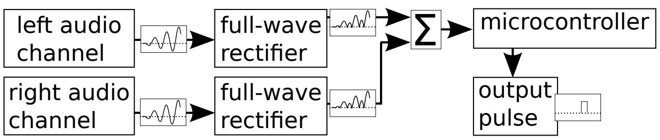

Fig. 2: The operation of the audio sensor.

time). The first two parameters are only possible to adjust as parameters in the software, but the dead time of the sensor is possible to adjust by a hardware knob. The setting of the knob is then read by the microcontroller and transformed into a dead time value in milliseconds.

The software of the audio sensor operates in two phases: **(1)** a setup phase and **(2)** a monitoring phase. In the setup phase, the software allows the user to change the dead time of the sensor for a period of 10 seconds (adjustable in the source code). The current dead time in milliseconds is shown to the user in real-time on the LCD screen as the user turns the adjustment knob. If no user input has occurred for 10 seconds, the software moves on to the monitoring phase. In the monitoring phase, turning the adjustment knob no longer changes the delay to avoid accidental changes during recording, until the value is re-read during the next initialisation of the audio sensor. The audio sensor polls the input (the summed audio channels) and compares the voltage to the detection threshold. If the threshold is exceeded, the sensor outputs a square wave marker signal. After outputting the marker signal, the sensor sleeps for a period corresponding to the dead time, preventing any input during this time (e.g. from the second part of a multipart audio stimulus) to trigger the sensor.

IV. Results

Examples of the operation of the audio sensor and photosensor are shown in Fig. 4 and Fig. 5.

In order to achieve high temporal accuracy, the internal latency of the audio sensor and photosensor must be small. The internal latency of the audio sensor was measured as follows. A square wave from a signal generator, with an amplitude several magnitudes above the sensor's detection threshold, was fed to the audio input port of the audio sensor. The input waveform from the signal generator and the output waveform from the audio sensor were simultaneously measured using an oscilloscope. The time difference between the waveforms represents the internal processing time of the audio sensor. The result are shown in Table I. The results show, that the internal latency of the audio sensor corresponds to a precision of about 1 sample, even at a very high EEG sampling rate of 40 kHz. The detection threshold of the audio sensor also influences the delay, as a higher detection threshold makes the sensor less sensitive, requiring a higher input signal level before a marker signal is output.

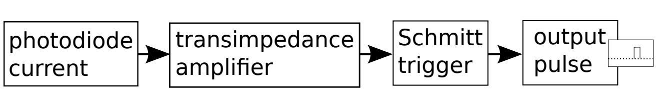

Fig. 3: The operation of the photosensor.

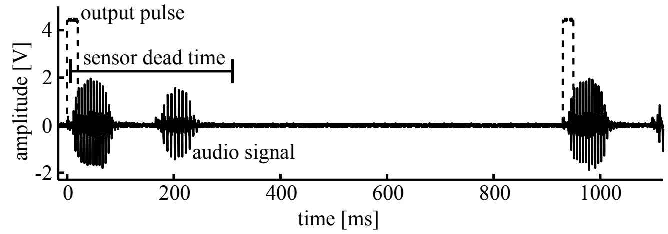

Fig. 4: Detection of audio pulses by the audio sensor. The sensor correctly detects only the first part of a two-part audio stimulus at $t = 0$ ms, and triggers again at $t = 950$ ms when encountering a new audio pulse after the dead time of 250 ms.

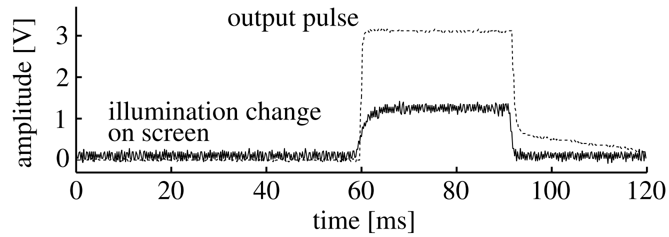

Fig. 5: Detection of illumination changes on an LCD monitor by the photosensor. The photosensor outputs a square wave when the light intensity exceeds a threshold.

The internal processing latency of the photosensor was measured as follows. The photodiode was subjected to a light pulse from a bright white LED fastened close to the photodiode. It was ensured that no other light reached the photosensor. The voltage over the LED and the square output wave of the photosensor circuit were simultaneously measured using an oscilloscope. The time difference between these two waveforms gives an upper limit for the internal latency of the analog photosensor circuit. The results are shown in Table I. The internal latency of the photosensor is negligible in comparison to the refresh rates of computer monitors, in the order of 100 Hz. The photosensor was also tested with a Samsung P2770H LCD monitor, having a reported response time of 1 ms and a refresh rate of 60 Hz, providing a temporal resolution of 17 ms between successive frames. The photosensor functioned well with this monitor (Fig. 5).

TABLE I: Average internal latency of the sensors ($N = 5$).

Sensor	Internal latency [µs]
Audio sensor	22.7 (6.8)
Photosensor	1.4 (0.2)

Standard deviation in parenthesis.

V. Discussion

In modern neuroscience, the auditory and visual stimuli used in experiments are becoming more variable and natural, resulting in the use of different types of solutions for presenting the stimuli. Yet, the timing accuracy of the stimulus presentation with respect to the data collection is in many cases required to be on the millisecond scale.

External hardware audio sensors and photosensors make experiments more robust, since researchers do not need to rely solely on the output of the stimulus presentation software for timing information of stimulus onset. External sensors also allow researchers to verify that the stimuli actually were presented to the subject. The hardware sensor solution described here provides researchers working with auditory and visual ERPs and ERFs a simple and effective method of increasing the accuracy and reliability of the marker signals required for successful ERP/ERF analyses. It is independent of the stimulus presentation software and signal acquisition system and can therefore be integrated into virtually any ERP/ERF measurement paradigm. The internal latencies of both sensors are low and they can be considered to operate in real time and reliably represent the true time of occurrence of the stimuli.

A second application of the described sensor solution is the validation of timing accuracies of stimulus presentation systems, i.e. quality control. The sensors can be used e.g. to quantify the jitter in evenly-spaced auditory and visual stimulus sequences and to determine the latency in a marker signal, corresponding to an auditory or visual stimulus, sent by the stimulus presentation programme with respect to the true time of stimulus occurrence as determined by the sensors. The sensors can also be used alongside markers from audio and video stimulus systems, supplementing them. The audio sensor can also be used to time an oral response by a subject, allowing the researcher to easily find the corresponding segment in the biosignal data. The audio sensor is also insensitive to silence in a sound file before the actual auditory stimulus and the sensor will report the true time of occurrence of the sound as observed by the subject.

The use of the audio sensor described here is not limited to neurophysiologic and naturalistic neuroscience, but can be used in any experimental setting requiring precise real-time monitoring of complex auditory stimuli. The software controlling the sensor's microcontroller can be easily tailored to be optimally suitable for different experimental paradigms.

The audio sensor could be improved by allowing all parameters to be adjusted by the researcher, without having to modify the software to control e.g. the length of the output pulse. The most flexible way would be to add pushbuttons to the microcontroller for selecting the parameter to be modified and for increasing or decreasing the value of the selected parameter. Instead of being adjusted by potentiometers, the parameter values could be stored in the microcontroller's non-volatile electrically erasable programmable read-only memory (EEPROM), which for the ATmega32U4 chip is 1 kB.

The photosensor could be improved by replacing the analog Schmitt trigger by a microcontroller-based solution, similar to the audio sensor. This would allow the threshold of the photosensor and the duration of the output pulse to be adjusted. Different thresholds may be needed for different monitors.

The total cost of the sensor solution presented here is approximately 100 €.

VI. Conclusion

All hardware schematics (audio sensor and photosensor), a bill of materials and software required for the operation of the audio sensor presented here are freely available from the homepage of the authors' lab: `http://www.brainworklab.fi/`. The software is released under the GNU General Public License version 3 (GPLv3) [6].

The microcontroller-based audio sensor provides a real-time audio monitoring solution. The adjustable dead time of the audio sensor makes the audio sensor most suitable for ERP/ERF paradigms with auditory multipart stimuli, eliminating the need for post-processing of markers in the data, saving time and reducing the risk of errors. The sensors make ERP/ERF measurements more reliable and increases the temporal accuracy with which stimuli can be detected. The open hardware and software architecture makes the low-latency hardware sensors presented in this paper a flexible solution for use in any research lab working with ERPs or ERFs, providing a simple and cost-effective way of increasing the reliability and accuracy of event-related measurements, especially when using multipart auditory stimuli.

VII. Acknowledgment

The authors thank P. Plaketti for assembling the sensors.

References

[1] T. Picton, S. Bentin, P. Berg, E. Donchin, S. Hillyard, R. Johnson Jr, G. Miller, W. Ritter, D. Ruchkin, M. Rugg *et al.*, "Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria," *Psychophysiology*, vol. 37, no. 2, pp. 127–152, 2000.

[2] E. Partanen, M. Vainio, T. Kujala, and M. Huotilainen, "Linguistic multifeature mmn paradigm for extensive recording of auditory discrimination profiles," *Psychophysiology*, 2011.

[3] M. Tervaniemi, S. Kruck, W. De Baene, E. Schröger, K. Alter, and A. Friederici, "Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus," *European Journal of Neuroscience*, vol. 30, no. 8, pp. 1636–1642, 2009.

[4] *AVR120: Characterization and Calibration of the ADC on an AVR*. [Online]. Available: http://www.atmel.com/Images/doc7766.pdf

[5] *Teensyduino*. [Online]. Available: http://pjrc.com/teensy/teensyduino.html

[6] *GNU General Public License*. [Online]. Available: http://www.gnu.org/copyleft/gpl.html