

A graphical user interface for infant ERP analysis

Jussi Kaatiala · Santeri Yrttiaho · Linda Forssman ·
Katherine Perdue · Jukka Leppänen

Published online: 22 November 2013
© Psychonomic Society, Inc. 2013

Abstract Recording of event-related potentials (ERPs) is one of the best-suited technologies for examining brain function in human infants. Yet the existing software packages are not optimized for the unique requirements of analyzing artifact-prone ERP data from infants. We developed a new graphical user interface that enables an efficient implementation of a two-stage approach to the analysis of infant ERPs. In the first stage, video records of infant behavior are synchronized with ERPs at the level of individual trials to reject epochs with noncompliant behavior and other artifacts. In the second stage, the interface calls MATLAB and EEGLAB (Delorme & Makeig, *Journal of Neuroscience Methods* 134(1):9–21, 2004) functions for further preprocessing of the ERP signal itself (i.e., filtering, artifact removal, interpolation, and rereferencing). Finally, methods are included for data visualization and analysis by using bootstrapped group averages. Analyses of simulated and real EEG data demonstrated that the proposed approach can be effectively used to establish task compliance, remove various types of artifacts, and perform representative visualizations and statistical comparisons of ERPs. The interface is available for download from <http://www.uta.fi/med/icl/methods/eeg.html> in

a format that is widely applicable to ERP studies with special populations and open for further editing by users.

Keywords Brain development · Electroencephalography · Event-related potential · Infant

Measurement of event-related brain potentials (ERPs) is one of the key techniques for examining typical and atypical neurodevelopment in human infants (Nelson & McCleery, 2008). Measurement of ERPs from infants is, however, very susceptible to various artifacts that arise from limitations in infants' ability to follow instructions (or engage in the behavior of interest), remain vigilant, and maintain a steady posture over extended periods of time (Hoehl & Wahl, 2012). In this article, we show that the majority of the most common artifacts and some potentially important but unrecognized sources of artifacts (e.g., systematic stimulus-dependent movements) in infant ERP studies are most readily detected through offline analysis of participant behavior from video records, prior to any analysis of the electrophysiological signal itself. After the data have been parsed on the basis of video records, epochs retained for analysis can be further processed by applying various signal processing routines to the electrophysiological signal (e.g., to detect noisy channels and low-frequency drift). After discussing the need for such a two-stage approach and the various reasons why video records are underutilized in current analytic approaches, we introduce a new graphical user interface that has been designed for an easy implementation of the proposed approach in the MATLAB environment (Mathworks, Inc., Natick, MA). In the Results and Discussion section of this article, we evaluate the proposed preprocessing routines by using example data from 7-month-old infants and simulated data.

Electronic supplementary material The online version of this article (doi:10.3758/s13428-013-0404-4) contains supplementary material, which is available to authorized users.

J. Kaatiala · S. Yrttiaho · L. Forssman · J. Leppänen (✉)
Tampere Center for Child Health Research, University of Tampere
School of Medicine, 33014 Tampere, Finland
e-mail: jukka.leppanen@uta.fi

K. Perdue
Boston Children's Hospital and Harvard Medical School, Boston,
MA, USA

Task compliance and artifacts in infant EEG

Technical aspects of the recording system, the laboratory environment, and electrode contact can all contribute to the sensitivity of the recording to noise and artifact. For example, cleansing and abrading dead skin cells and oils make the recording less sensitive to skin potentials and low-frequency noise (Kappenman & Luck, 2010; Sinisalo, Mäki, Stjerna & Vanhatalo, 2012). Yet even an ideally configured recording setup with well-prepared electrode contact is not immune to artifacts that arise from noncompliant behavior (i.e., the child is not engaged in the behavior of interest), muscle activity (electromyography [EMG]), and perhaps most important, various types of movement. Thus, the detection and removal of artifacts is always an important part of ERP analysis.

Movement is the single most common source of artifacts in infant ERP studies (Hoehl & Wahl, 2012). Movement artifacts in infant studies include gross movements of the upper body (e.g., leaning backward or forward), head turning, limb movements, saccadic eye movements, blinks, and more subtle movements of the mouth (sucking) and facial muscles (e.g., raising the forehead). In some cases, movement-related artifacts can be relatively easily distinguished from the EEG on the basis of their larger amplitude (gross movements) or unique shape (such as saccades and blinks). This is not always the case, however. Eye movements and blinks are characterized by consistent, temporally confined effects on EEG in adults, but the effects can be much more variable and harder to separate from background EEG in infants (Fujioka, Mourad, He & Trainor, 2011). Also, because infants are easily distracted by electrodes in the facial area, the electrodes that are most sensitive and optimal for detecting eye movements are typically left out from the recording montage in infant ERP studies. Sucking of a pacifier can result in a low-frequency artifact in EEG that is not necessarily distinguishable from the background EEG. Finally, subtle movements of facial muscles may affect EEG. For example, modest pushing of the tongue toward the incisors in a closed mouth results in scalp potentials that have a strong gradient between the mastoid region and frontal scalp areas and may be mistakenly similar to the slow-wave cognitive potentials (Vanhatalo, Voipio, Dewaraja, Holmes & Miller, 2003).

Infants' movements in ERP studies are often sporadic and not systematically related to the events of interest. This is important since subtle changes in the EEG signals that are unrelated to the event of interest can be expected to be canceled out in the process of averaging across several events. The possibility exists, however, that some movements are systematic and occur in a relatively regular pattern in response to specific stimuli. It is known, for example, that the dishabituation of attention in infants (i.e., recovery of interest in a stimulus after some aspect of the stimulus has been changed) is associated with increased high-amplitude sucking

(Atkinson & Braddick, 1976). Such effects may occur in response to various types of stimuli. Infants may also spontaneously imitate others' facial gestures (Wörmann, Holodyski, Kärtner & Keller, 2012). Although the time course of such imitative responses is not well characterized, studies in adults have shown that differential EMG activity in response to viewing pictures of facial gestures can be observed 300–400 ms after stimulus onset (Dimberg & Thunberg, 1998) and, therefore, systematic EMG responses to specific stimuli may well occur within the time frame of typical ERP analyses. Finally, a common observation in event-related studies with infants is that infants may look up toward the parent's face regularly during the testing session. This behavioral pattern is known as social referencing (Sorce, Emde, Campos & Klinnert, 1985), and it may be particularly evident in 10- to 12-month-old infants.

The effects of infants' systematic behavioral responses to events of interest are possibly not that detrimental for very short-latency evoked potentials occurring within the first 100 or 200 ms after stimulus onset. However, the analysis periods in typical infant ERP studies are often substantially longer (up to 1.5 s) and, thus, may well capture some of the regular movement responses discussed above.

Existing approaches for behavior monitoring and artifact removal

Currently, the most common and, perhaps, also most practical method of monitoring noncompliant behavior and artifacts is through video monitoring of the infant during the testing session. This method is less precise than other methods such as eye tracking, EMG, or electro-oculogram (EOG), to register specific types of artifacts. However, video records can give valuable information about a broad spectrum of infant behaviors, including behaviors that are not easily monitored by other systems (e.g., regular patterns of movement in response to stimuli). Also, some of the other monitoring systems may be impractical due to infants' intolerance to electrodes near the eyes (EOG) or in facial areas (EMG).

Typically, real-time monitoring of infant behavior is used to detect periods of recording when the infant is complying with the requirements of the task (e.g., attending on the screen), and the events of interest can be presented to the infants. The experimenter may also mark periods of data when the infant is noncompliant or there is movement that is likely to contaminate the recording. Although such records are relatively easy to obtain, they are also prone to errors, due to the limited time available for making the markings during the testing session, variable delays between the actual behavior and the manually made records of that behavior, and difficulties in detecting certain behaviors real time. The experimenter may not, for example, notice sporadic or systematic stimulus-dependent

movements of the head, mouth, or other parts of the body. If the records of infant behavior are not retrieved for offline data processing stages, potential sources of artifact may go unnoticed in data processing. As was discussed above, some of the potential artifacts that go unnoticed during the actual testing session may result in scalp potentials that are indistinguishable from the background EEG signal and, therefore, remain unnoticed in the offline analysis stage if the analysis is based on visual inspection or automatic processing of the electrophysiological signal alone.

A preferred solution is, therefore, to code the video records offline. Even typical digital video recordings and editing software provide sufficient temporal accuracy for most behavioral analyses (30 frames per second) as well as slow-motion playback options for detailed behavioral analysis. In more advanced systems, higher sampling at 60 frames per second or more can be achieved.

Although video–EEG integration is an established routine in clinical EEG recordings and some researchers have used offline analysis of videos in infant ERP studies (see Hoehl & Wahl, 2012, and references therein), our experience is that the potential benefits of coanalyzing video and EEG are underutilized in current analytic approaches, most likely due to impracticalities and limitations in video–EEG synchronization. The integration of video and electrophysiological data may be either impossible or very time consuming in many of the existing software packages, and if the integration is possible, it is not optimized for analyzing data from event-related experimental designs. For example, software packages may enable simultaneous recording of video and physiological signals, but the two streams of data are not necessarily synchronized to the level required in event-related studies (the asynchrony can be up to seconds), and it is typically not possible to segment the video with the physiological data. As it is, synchronization has to be conducted manually by matching event markers in the video stream (e.g., mirror image of the display shown to the participant) with those in the physiological signal (event markers from the computer). For this reason, a substantial amount of time is required for simply finding and matching the events of interest in the two streams of data before actual analysis or coanalysis of the data can commence. The time required for the analysis is further increased by the fact that tens or even hundreds of trials may be presented in physiological studies with infants and because lengthy video files may be cumbersome to process in video-editing software. Finally, because the matching process is entirely manual, it is error prone.

Summary

The recording of event-related potentials in infants is prone to several different sources of artifacts. Some of these artifacts can be avoided by infant-friendly laboratory setups, paradigms, and

other established routines (Hoehl & Wahl, 2012). However, several of the artifacts are unavoidable (most notably, movement artifacts) and are therefore carried on to the EEG signal if not removed by adequate procedures. Because some of the artifacts discussed above are potentially not distinguishable from the background EEG signal (especially in infants), artifact detection cannot ideally be based on the electrophysiological signal alone and needs observational data on infant behavior during the testing session. Although such data are available in video records taken during the testing session, these records are not routinely retrieved in the data analysis stage, due to difficulties in synchronizing EEG and video data efficiently and attaining a sufficient level of accuracy for event-related designs.

Materials and methods

The primary goal of the present project was to design a widely applicable user interface that enables efficient detection and removal of artifacts on the basis of observational data and integrates this stage of data analysis seamlessly with other EEG preprocessing stages. To achieve this, we created a graphical user interface for the MATLAB environment. The interface is designed for a two-stage approach. In the first stage of this approach, the video and EEG data are synchronized on the basis of minimal user input, and the video can be inspected on a trial-by-trial basis to reject artifact-contaminated epochs. Information about rejected trials is automatically saved in the format that can be easily retrieved in subsequent stages of EEG processing. In the second stage, the interface calls MATLAB and EEGLAB (Delorme & Makeig, 2004) functions for further preprocessing of the signal, including different filtering options, baseline correction, artifact detection based on EEG signal, interpolation, and rereferencing. Finally, a separate analysis interface is included that allows for calculation and visualization of average event-related potential and time-frequency responses using EEGLAB functions.

The interface and specific instructions for its use can be downloaded from <http://www.uta.fi/med/icl/methods/eeg.html>. In this article, we describe the technical solutions used in the interface and the impact of proposed approaches on the quality of infant ERPs.

Video-based artifact detection

Synchronizing video with EEG

Our goal was to create a user interface that offers a generic solution for synchronized viewing of video and EEG (or any other physiological signal) at the level of single epochs. In addition to providing the user with a convenient way to

observe events in the video and EEG, we wanted to optimize the interface for viewing and marking of data from event-related designs—that is, to allow automatic advancement from one event to the next and automatic storing of information about accepted and rejected events for subsequent processing stages. Given the differential temporal resolution of current digital video and EEG systems (in our setup, 30 and 250 Hz, respectively), synchronization at the level of individual EEG samples was not possible. Synchronization at the level of single video frame (± 33 ms) was considered sufficient, however, for visual inspection of the typical artifacts.

Our method requires the user to collect both EEG and video footage of the participant during the recording situation. EEG and video are not required to be started at the same time as the synchronization is performed in the offline data analysis stage. However, it is critical that both the EEG and video are collected at constant sampling rates (please note that this is not necessarily the case for all commercial EEG software packages) and that the video covers all the events. The best solution is to record video and EEG continuously during the recording so that video and EEG are not stopped until the end of the recording. Also, stimulus onset times must be marked onto the EEG-trace for all stimuli that are of interest. Our program will read these onset times directly from the EEG collected by the Electrical Geodesics recording systems (i.e., data exported in .raw format) or data converted to EEGLAB .set format. In addition to the event onset times from the EEG trace, our method requires that the onset of the first stimulus should somehow be observable or the user should know the onset point in the video. Common ways to achieve this in visual ERP studies is to use observable luminance contrast between the prestimulus period and the stimulus period (e.g., a change from gray to white background that appears as a flash in the video recording) or set up a mirror that shows an image of the stimulus display.

To synchronize video with a continuous EEG datafile in our user interface, the user must upload the EEG datafile and a corresponding continuous video file and then manually find and mark the frame for the first stimulus on the video (this manual marking is required for the first stimulus only, to minimize user input). The marking on the video file is used for calculating the time difference between the first trial in the EEG and the first trial in the video in order to match the timescales of the two data sources (see Fig. 1). The actual synchronization of the video and EEG occurs automatically when the EEG data are segmented according to the following three steps: (1) The time stamps included in the original EEG file are used to calculate the onset times of each event from the beginning of the EEG recording (a corresponding time stamp for the first epoch in the video has been given by the user when the video is uploaded); (2) the first epoch on the recording time line is selected in both the EEG and video, and the offset of the two is subtracted from the video time line,

taking into account the differences in sampling rates; and (3) the minimum and maximum time value of each EEG epoch in the EEG time line (i.e., the starting and ending points of the epochs) are taken and used to find the corresponding starting and ending points for all epochs in the video data. In this step, the first frame of the epoch in the video is defined to be the first video frame after the start of the corresponding EEG segment. Similarly, the last frame of the epoch in the video is the last video frame before the end of the EEG segment. The purpose of the last step is to form vectors that serve as a reference for the video frames that are included in a particular EEG epoch.

The synchronization method in our interface requires that the sampling rates should be constant for the video and EEG signals throughout the recording session or that potential drift in either of the two signals is not large enough to affect the synchronization. Researchers can test the accuracy of the automatic synchronization process by comparing event markers not only for the first trial, but also for the last trial in the recording session. Problematic levels of drift would lead to a misalignment of the video and EEG for the last trial (i.e., the onset of the event in the EEG does not match with the onset of the event in the video).

Viewing segmented videos, marking rejected trials, and storing information about rejected trials for subsequent processing stages

After synchronization, the segmented EEG can be viewed together with the segmented video. The interface displays EEG and video in two separate windows (Fig. 2). The EEG window displays EEG signals at each electrode location according to the place of the electrode in the recording array. Using the controls in the bottom and top-right corner, the user can browse the data epoch by epoch. When the user browses EEG epochs, the video will automatically move to the start of the corresponding epoch in the video. Within the selected epoch, the user can play the video frames back and forth. When playback reaches the end of the epoch, the video returns to the first frame, and hence, the video is repeated. The user can remove an epoch from the analysis by pressing the button “remove epoch” in the EEG window controls.

Important to our approach was that the data of the removed epochs should be stored in an accessible format for subsequent stage of EEG preprocessing. Our solution here was to store the identifiers of the events (i.e., event or trial numbers) that were rejected during the video-viewing stage and retrieve these data in subsequent preprocessing stages. Thus, after the bad segments have been removed, the user can export the information about the segments that were removed. The most convenient way for storing this information is to export the segmentation information as a function call that can be “read” into the analysis in subsequent EEG preprocessing stages, as

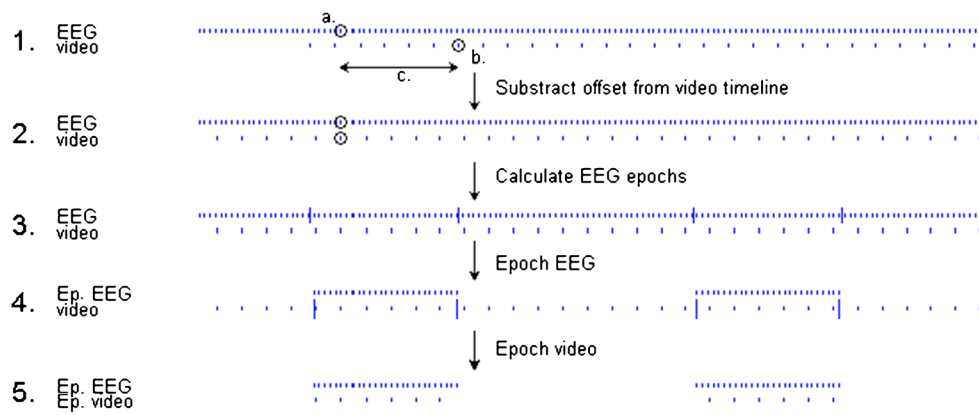


Fig. 1 Method of video-EEG synchronization. In phase 1, the EEG and video start at different times but proceed with constant rate and continuously. The time point of the first stimulus time in the EEG (**a**) and the time point of the first stimulus in the video (**b**) are not aligned in time. The offset between **a** and **b** is calculated and subtracted from all sample times in the video timeline. In phase 2, the video and the EEG are

described below. Function calls can be collected later on as a common script file, which makes preprocessing with different parameters and workflows easy and preserves the video analysis information.

Preprocessing of the EEG signal

The video analysis can presumably help to eliminate a substantial portion of artifacts and noise from the signal, but it does not eliminate the need for further preprocessing of the

electrophysiological signal. Inspection of data in our example data set (described below) showed that single-channel data on the trials that were judged as analyzable on the basis of the video analysis were frequently contaminated by high-amplitude changes and substantial low-frequency drifting, with the latter most likely arising from cephalic skin potentials that are typical in high-impedance recordings (Kappenman & Luck, 2010).

In the following, we describe various routines that are implemented in our interface for the purposes of offline signal

processing. Inspection of data in our example data set (described below) showed that single-channel data on the trials that were judged as analyzable on the basis of the video analysis were frequently contaminated by high-amplitude changes and substantial low-frequency drifting, with the latter most likely arising from cephalic skin potentials that are typical in high-impedance recordings (Kappenman & Luck, 2010).

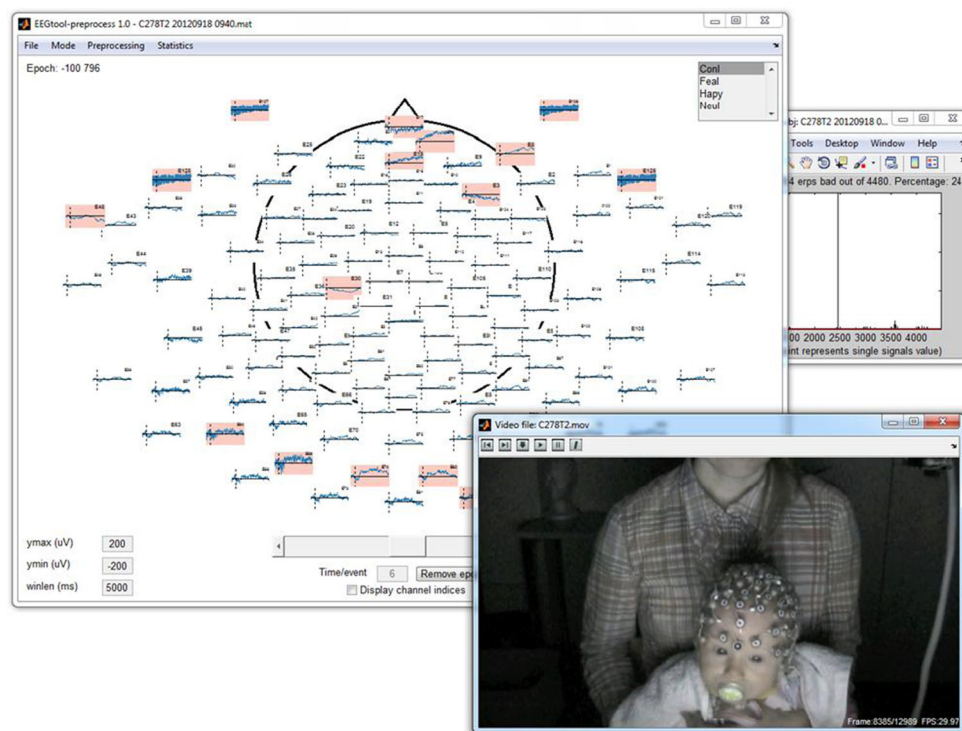


Fig. 2 An overview of the interface used for viewing video records of the participants together with the epoched EEG

processing. All of these routines have been previously used in ERP studies, although some of the proposed preprocessing steps are less commonly included in the analysis of infant ERPs.

Many infant researchers perform artifact detection *manually* on the basis of visual inspection of the signal. Such analyses can be performed in our graphics user interface by manually marking channels as bad (details of this approach are given in the user manual). In a *semiautomatic* strategy (<http://sccn.ucsd.edu/wiki>), the researcher uses automatic artifact detection algorithms for artifact detection and later “accepts” or “rejects” the detected artifacts manually. In our interface, this strategy can be implemented by using the artifact detection algorithms in the graphical user interface and manually accepting or rejecting the detected artifacts epoch by epoch. A fully automatic artifact detection is implemented in a scripting mode in our interface, using any of the algorithms described below. Instructions for implementing each of these steps are given in the user manual for the interface.

Channels with high impedance

High impedance values can serve as a proxy for a poor contact between the electrode and the skin and may help to identify channels that are particularly susceptible to noise (Kappenman & Luck, 2010). For this reason, we developed a function that will automatically retrieve electrode impedances in the offline analysis stage and allows for rejection of channels with impedance values above a user-defined threshold.

High-pass filtering and detrending

Slow shifts of voltage over a period of seconds is a common problem in high-input impedance recordings and is mainly caused by cephalic skin potentials (i.e., varying differences in conductance across the skin epithelial layer; see Kappenman & Luck, 2010; Sinisalo et al., 2012; Stjerna, Alatalo, Mäki & Vanhatalo, 2010). In our experience, this problem is common in high-density recordings with infants, causing a clearly visible linear trend in the signal on individual epochs. In the present approach, we included two methods that can be used to remove low-frequency drifts: a *high-pass filter* and *signal detrending*. The high-pass filter function calls *pop_eegfilt* EEGLAB function (see Rousselet, 2012, for a discussion of other filtering options in EEG preprocessing). The *detrend* function is a MATLAB function that removes a linear trend from the data from every epoch (Craston, Wyble, Chennu & Bowman, 2009; Martens, Korucuoglu, Smid & Nieuwenstein, 2010). Both of these functions offer effective solutions for removing low-frequency activity from the data, but they should be applied with caution and awareness of the potential

problems associated with severe high-pass filters. For a more extensive discussion of these problems and possible recommendation for filter types and settings, the reader is referred to recent publications by Kappenman and Luck (2010) and Rousselet (2012).

Automatic artifact detection

The most common method for automatically detecting and removing artifacts is based on amplitude thresholds under the assumption that artifacts can be separated from the background EEG on the basis of high-amplitude fluctuations within a relatively short period of time (Hoehl & Wahl, 2012). In the present approach, these methods can be used by setting a maximum threshold or by setting a maximum value for the difference of minimum and maximum during the ERP time window. In addition to these methods, we have implemented a third method that is based on setting a threshold for the root mean square (RMS) of the signal (Palmu, Kirjavainen, Stjerna, Salokivi & Vanhatalo, 2013). Each of three artifact detection methods treats each EEG channel autonomously. This provides a method that gives the user a clear understanding of which signals are marked as “bad” and “good.” The user can later check or uncheck signal markings on the basis of visual inspection.

Other signal-processing routines

In high-density recordings, the number of trials retained for analysis can be very low if all recording channels are required to be artifact-free. To minimize data loss, researchers routinely replace individual bad channels by using various interpolation methods, when the number of individual bad channels is not excessive (Junghöfer, Elbert, Tucker & Rockstroh, 2000). In the present approach, the user has the option to retain channels for analysis when the number of bad channels does not exceed a user-defined threshold (e.g., 10 %) and interpolate artifact-contaminated channels by using the EEGLAB *eeg_interp* function.

The rereferencing option in the interface calculates a new reference for the EEG by calling the EEGLAB *reref* function. The user has the option to leave selected channels out of the calculation of the average reference.

Analyzing and visualizing ERPs

A separate graphical user interface is included for visualizing condition-specific ERP waveforms for individual participants or a group of participants and for extracting conventional amplitude and latency-based metrics for ERP analyses. In addition, we have included a function that uses bootstrapping, or sampling with replacement, to calculate 95 % confidence intervals (CIs)

for condition-specific group ERPs. The CIs are calculated as follows: (1) n acceptable data epochs (from a participant group and a given experimental condition) are pooled into a single repository of epochs; (2) an ERP bootsample is calculated by averaging n epochs drawn randomly and with replacement from the epoch repository; (3) altogether, N bootsamples or ERPs (e.g., 1,000) are acquired in this way; (4) the amplitude values of the ERP bootsamples are sorted in ascending order for each timepoint; (5) the grand-average ERP is calculated as the mean of all ERP bootsamples; (6) the confidence intervals for the ERP are calculated using the basic bootstrap intervals (centered percentile bootstrap intervals) as

$$(2\Theta - \theta * 1 - \alpha; 2\Theta - \theta * \alpha),$$

where Θ is the mean ERP (i.e., mean of all epochs in the epoch repository) and $\theta * \alpha$ is the ERP from the sorted bootsample ERPs corresponding to the percentile of $100 \times \alpha$.

Bootstrapping has been extensively used in adult studies (e.g., Rousselet, Husk, Bennett & Sekuler, 2008), but less in ERP studies with infants. There are several benefits of presenting average waveforms with bootstrapping CIs. First, CIs are important for visualizing the amount of uncertainty in ERPs and the robustness of potential differences in ERPs between experimental conditions (Allen, Erhardt & Calhoun, 2012). Second, the latency or peak values of certain ERP components are often difficult to determine for single-participant averages (especially in infant ERP studies, where low trial counts are a common problem), whereas the time course of condition differences are readily observed in bootstrap group responses as periods of nonoverlapping CIs between conditions of interest (Rousselet et al., 2008). Finally, participants with low trial counts (e.g., fewer than 8 or 10 epochs per condition) are typically excluded from analyses in infant studies to attain sufficient signal-to-noise ratios. In doing so, the researcher also rejects all the acceptable epochs from those participants with a low total number of epochs. Therefore, a large number of perfectly valid epochs from different participants have to be rejected. Because bootstrapping analyses are based not on a single-participant average but, rather, on randomly selected epochs on the basis of multiple participants in the entire data set, they provide a means for statistical representation of data without excluding participants with low trial counts.

Results and discussion

To test the interface, we applied the proposed analytic approach to real and simulated ERP data. Our first goal was to evaluate whether video records can be used to establish task compliance and to remove artifacts in simple visual ERP

studies with infant participants. Second, we evaluated different signal preprocessing functions (i.e., detrending and artifact detection functions) with respect to their performance in extracting a known ERP waveform from background EEG (i.e., an ERP waveform that was computationally superimposed on resting state EEG). Finally, we examined the benefits of bootstrapping-based approaches in data visualization and analysis.

Example data sets

Real ERPs were obtained from an ongoing longitudinal study ($N = 127$; mean age at 7-month testing, 214 days; range, 207–243) in which 7-month-old infants viewed a sequence of nonface control stimulus or a neutral, happy, or fearful facial expression at the center of a computer screen (Peltola, Hietanen, Forssman & Leppänen, 2013). EEGs were recorded by using a Net Amps 300 amplifier and 128-channel EGI Hydrocel sensor nets (Electrical Geodesics Inc., Eugene, OR) with a 250-Hz sampling rate and a 0.01- to 100-Hz band-pass. EEG was collected together with video (Canon ZR960 digital video camera and QuickTime or iMovie software) and corneal-reflection eye-tracking (Tobii TX300, Tobii Technology, Stockholm, Sweden) data. The present analyses were confined to the first 800 ms of each trial, representing the period when the face alone was presented on the screen (later during the trial, a peripheral distractor stimulus was shown as described in Peltola et al., 2013). Approval for the project was obtained from the Ethical Committee of Tampere University Hospital, and informed, written consent was obtained from the parent of each child.

Simulated data sets were created by computationally adding an ERP response at randomly determined times to 2-min periods of resting state EEG. The number of ERP stimuli added was varied from 5 to 30, to simulate common numbers of trials in infant ERP studies. The ERP was an experimentally derived ERP from sixteen 7-month-old infants as reported in Righi et al., (in press) and shown in Fig. 4. The ERP was 900 ms long and was from the occipital response to a visual stimulation, recorded over nine channels. The ERP was superimposed on resting state EEG from 81 healthy infant participants, 5 months ($n = 15$), 7 months ($n = 34$), and 12 months ($n = 32$) of age. The ERP and the resting state EEG were recorded with an EGI HydroCel Geodesic sensor net with 128 electrodes and a NetAmps 200 amplifier (Electrical Geodesic Inc., Eugene, OR). Electrode impedances were less than 50 k Ω at all locations during the recordings. The code to add the ERPs to the resting state EEG was developed in MATLAB (The Mathworks, Natick, MA). All infants were recorded with informed consent from the parent(s), as overseen by the IRB of Boston Children's Hospital.

Evaluation of video-based artifact detection

The first 30 infants who had simultaneous recordings of video (at a constant 30/s frame rate) and EEG (at 250 samples per second) were selected for testing video-based artifact detection. Of the 30 infants, 1 infant was excluded prior to any preprocessing because of a technical difficulty in stimulus presentation and the infant's intolerance for the EEG sensor net, and 2 infants were excluded because of technical difficulties with the video recording (i.e., difficulty in locating the onset of the first stimulus in the video record). A total of 27 infants were retained for preprocessing. Of these, 3 infants completed part of the trials (i.e., 32–35 trials), and 24 infants all of the trials (i.e., 48 trials).

We evaluated our approach to video analyses in several ways. First, we asked two independent raters to mark all of the 27 example files and examined the interrater coherence of the markings. Second, we examined the assumption that visual inspection of the videos can be used to establish task compliance in simple visual ERP studies (i.e., to ensure that the infant is looking at the stimulus of interest) by comparing visual analyses with eye-tracking data. Third, we examined whether video-based artifact removal reduced noise and, thereby, improved the quality of the EEG (i.e., reduced the “magnitude,” or power, of EEG as assessed by RMS voltage).

Interrater agreement

The time required to code the video of 1 participant with 48 trials varied between 10 and 15 min. The primary coder rejected 44 % of the trials (557 out of 1,267 trials) from the analysis on the basis of visual inspection of the video records. The events in video leading to EEG rejections were the following: gaze shifted away from the central stimulus, saccades, blinks, sucking the pacifier, contraction of the oral or other facial muscles, prominent tongue movements, excessive body movements, infant or parent touching the electrodes, parent moving the infant, and infant being outside the angle of view of the video. Saccades or head/body movements appeared to be among the prevalent causes for event rejections. To examine the reliability of the coding, we asked an independent coder to code the same videos. The second coder had no prior experience in coding ERP-related artifacts. The second coder was given a one-page written explanation of the trial selection criteria along with one example file (different from the coded data set). The second coder rejected 40 % of the trials (505 out of 1,267 trials). The two coders agreed on 82 % of the trials, and the Cohen's kappa was .63. In general, a Cohen's kappa between .60 and .75 is regarded as “good agreement.”

Comparison of video judgments with eye tracking

To test the assumption that visual inspection of the videos can be used to establish task compliance in simple visual ERP studies (i.e., to ensure that the infant is looking at the stimulus of interest), we selected data from the first five 7-month-old infants who had simultaneous recordings of video and eye-tracking data with Tobii TX300 corneal-reflection eyetracker (Tobii technology, AB). For these analyses, we selected trials that fulfilled the criterion of task compliance on the basis of the video analysis (i.e., the infant was judged to be looking at the stimulus) and examined whether the point of gaze fell within the coordinate limits corresponding to the stimulus (i.e., the face stimulus shown in the middle of a computer screen). Of the 120 trials selected for the analysis (24/participant), 54 were accepted on the basis of the video analysis for ERP. Valid eye-tracking data were available for 36 of the 54 accepted trials (i.e., eye gaze data available for 90 % or more of the trial time). On these 36 trials, point-of-gaze was within the face area, on average, 99.2 % of the time (range, 91 %–100 %). This analysis supports the assumption that video records can be used to establish task compliance, at least in relatively simple tasks that require attention to a single stimulus or spatial location.

EEG noise reduction

Our third analysis examined whether video-based artifact detection was effective in reducing noise in the EEG. We predicted that some of the artifacts detected during video inspection (e.g., gross movements, touching the electrodes), although not all (subtle movements), are associated with large voltage changes and add to the overall amplitude changes in the recording. To assess the effects of video-based artifact removal on segmented EEGs, we used the RMS of EEG voltage as a measure of the overall level of noise in the EEG segments. The RMS provides an overall measure of the “magnitude,” or power, of the signal, irrespective of frequency (Kappenman & Luck, 2010). To remove the effect of DC offset on RMS voltage, the EEG segments were baseline corrected against a 100-ms prestimulus period before calculating RMS voltage for raw and video-corrected EEG segments. The mean RMS before video editing was 85.8 ($SD = 72.8$; range, 28.4–285.8), and the mean RMS after video-based artifact removal was 60.0 ($SD = 39.4$; range, 24.1–161.4). A reduction of RMS was observed in all but one of the example cases, $t(26) = 3.3$, $p = .003$. This difference in the RMS value confirms that visual inspection of the video records reduces noise in the EEG, most likely because artifacts such as movement typically induced high amplitude changes in EEG. It is noteworthy, however, that the mean value and variability of the RMS remains high for epochs that survived video analysis. This aspect of the data underscores the

importance of further analyses focused on the EEG signal itself.

Evaluation of EEG signal preprocessing functions

High-impedance channels

Figure 3 shows the scalp distribution of electrode impedance values (kOhm) for a 128-channel high-density EEG recording in the example data set. The scalp distribution was calculated on the basis of median impedance values from all available impedance logs in the example data set. The distribution shows the expected pattern of higher impedance values for electrodes around the ears and for electrode 17 above the nose (an electrode that usually has poor scalp contact in the infant EGI nets) and, perhaps surprisingly, slightly increased impedance values around the vertex (possibly owing to the fact that the electrode net was loose around the scalp for some participants). It is noteworthy that there is no indication of higher impedance values for the last row of posterior electrodes in this analysis.

Although the regional differences in impedance values were relatively small in our data, it may be advisable to exclude electrodes with consistently poor scalp contact from the recording montage, given their susceptibility to noise (Isler, Tarullo, Grieve, Housman, Kaku, Stark & Fifer, 2012; Nyström, 2008). Exclusion of these electrodes from the montage is important because they distort average reference and interpolation even when not included in the area of interest in the ERP analysis. It is also noteworthy that although some of the other variation in the impedance level is relatively small and clearly under the acceptable range (i.e., <50 kOhm), higher impedance level may be a proxy for poor or unstable contact, and even slight variations in contact are possibly not trivial, given their possible effects on the regional distribution of the signal-to-noise ratio (Kappenman & Luck, 2010).

Detrending and artifact correction

To evaluate our detrending and artifact detection methods, we used these functions to “recover” the ERP waveform that was added to the resting state infant EEG in our simulated data and compared the recovered ERP with the true ERP. This procedure was used to test the impact of the number of trials (5–30), number of participants (20 vs. 81), and processing method on the accuracy of the recovered ERP.

Four data-processing streams were tested. The first processing stream was a simple average over all of the trials. The second processing stream was a trial-by-trial detrending before averaging. The third was artifact rejection, implemented by rejecting trials that had a range of more than 150 μ V. The final processing stream was threshold rejection with the same parameters and added

detrending. In all processing streams, baseline correction with a period of 100 ms was used, and the response was averaged over the nine occipital channels that contained the simulated ERP.

The accuracy of the processing methods was evaluated by calculating the root mean squared error (RMSE) between the true ERP and the simulated ERP. This metric provides a general measure of the error over the whole time window, and a smaller RMSE indicates that the method is a more accurate reconstruction of the original ERP. The RMSE averaged over participants is reported for all numbers of trials and processing methods, in the 20 and 81 participant groups.

The impact of data-processing methods and trial numbers on the recovered ERP is shown in Fig. 4. The dependence of the mean RMSE on the number of trials is shown in panels A and B for the 20- and 81-participant sample groups. Generally, RMSE decreases as the number of trials increases. For both participant groups, the simple average performed the worst and had the highest RMSE, and the combination of the detrending and threshold rejection performed the best. Detrending and threshold rejection appear roughly equivalent for the 20-participant group, but when more participants are available, threshold rejection seems to do better than simple detrending.

The reconstructed group average ERPs are shown in Fig. 4c, d, along with the true ERP used to create the simulation. The number of trials in this average was 15. The figure shows the impact of detrending on the reconstructed ERP, especially in the later part of the ERP, where the detrending leads to values that are closer to zero and, therefore, further from the true values. Despite the low RMSE values for the detrend plus threshold rejection condition, the mean ERP appears to be most faithful for the threshold rejection only condition. This result may be caused by the detrending procedure introducing a bias, leading to lower RMSE in individual participants but an overall offset when the participants are averaged. In particular, it appears that the offset is caused by the fact that the detrending procedure removes the slight negative trend that is present in the original ERP waveform.

Because detrending removes all linear trends from the data, it is important to determine whether these trends are independent of the experimental manipulations of interest in the study or whether there is a systematic relation between the trends and one or more of the experimental conditions. If the trends are independent, they can be safely removed from the data without introducing a bias in the phenomenon of interest in the study. If the trends are systematically associated with the experimental manipulations, the situation is more problematic and, at a minimum, further analyses are required to establish whether any of the observed phenomena can be attributed (i.e., is correlated) to the detrending procedure. To assist in such analyses, we created a function that calculates the trend

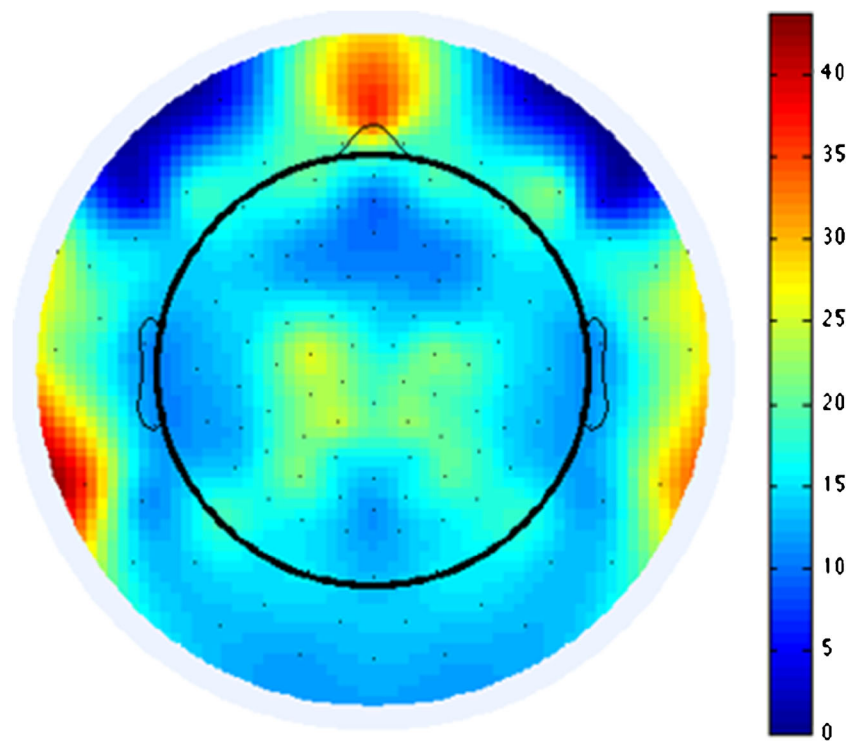


Fig. 3 Scalp distribution of electrode impedance values (kOhm) for a 128-channel high-density EEG recording

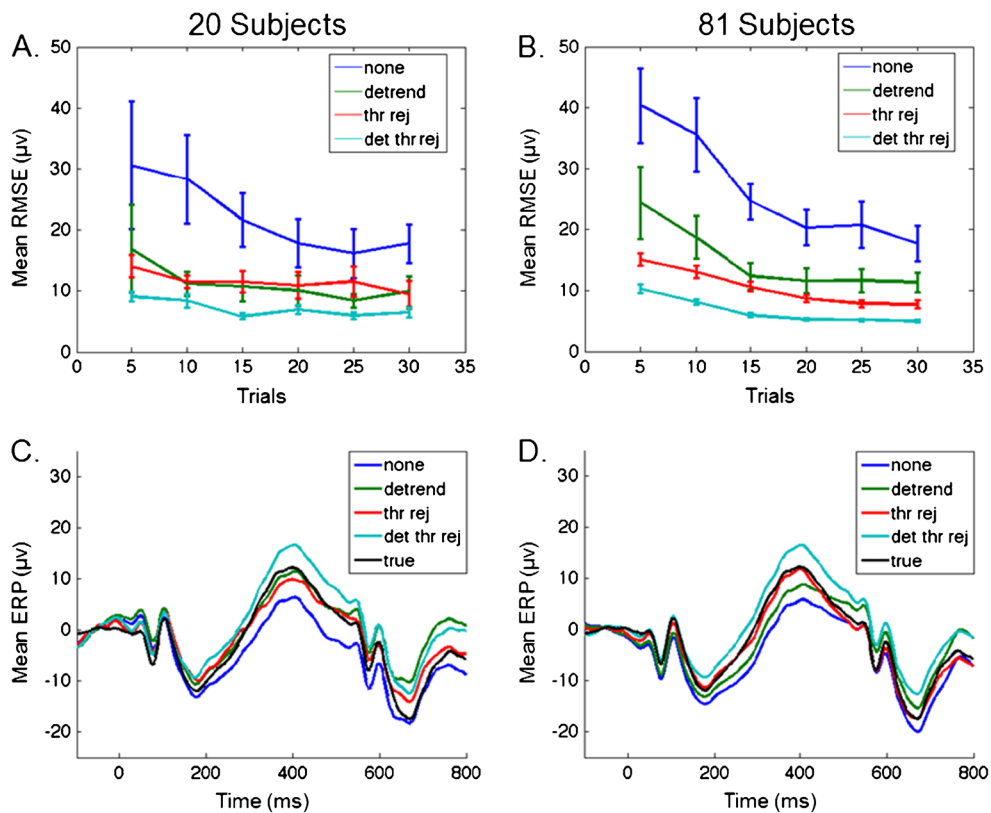


Fig. 4 Effects of sample size and various preprocessing methods on ERP extraction as assessed by comparing extracted ERPs with the original (“true”) ERP waveforms. The accuracy of the processing methods was

evaluated by calculating the root mean squared error (RMSE) between the extracted ERP and true ERP (a, b) and by comparing the average of the extracted ERPs with the “true” ERP (c, d)

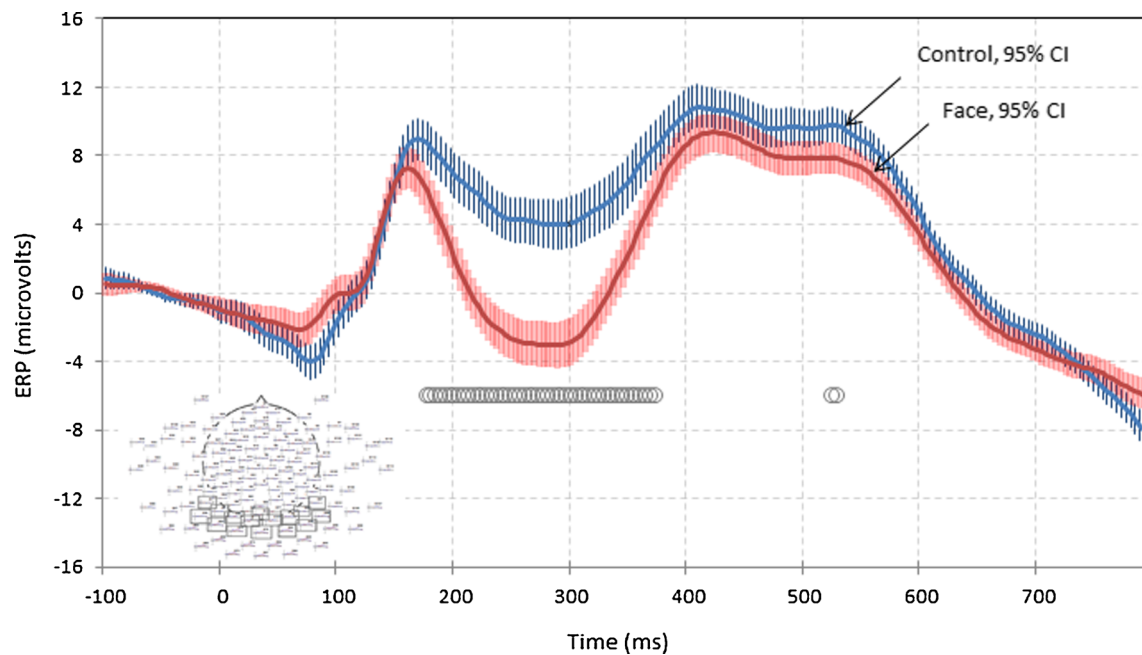


Fig. 5 A visualization of the average ERPs for nonface control stimulus and faces with 95 % confidence intervals (CIs). The CIs were calculated on the basis of 1,000 bootstraps using random selection of accepted epochs from all participants (a pool of 830 epochs/condition). Statistically

significant difference between the two conditions are shown in black circles along the *x*-axis and were identified by creating a time series of the supremacy of the lower bound of the control ERP over the upper bound of face ERPs

in each EEG epoch by using the MATLAB “polyfit” function and returns a coefficient for the slope of the trend line. The coefficients are averaged for each experimental condition and participant and saved into a .csv file.

Evaluation of data visualization and analysis functions

To evaluate our data visualization and analysis functions, we compared ERPs over occipital-temporal scalp regions for nonface control stimuli and facial expressions in our example experimental data set. Epochs surviving video analysis were further preprocessed by using an automated script with 30-Hz low-pass filter, detrending, baseline correction (−100 to 0), artifact detection (150 μ V threshold), interpolation (for trials with <10 % bad channels), and average referencing functions. Conventional analyses based on individual averages were conducted by including all participants with ≥ 5 trials per condition ($n = 78$, mean = 8.6 epochs/condition).¹ As was

expected on the basis of prior studies (Halit, Csibra, Volein & Johnson, 2004), the analyses showed differential mean amplitude for nonface control stimuli versus faces at the latency of the N290 (248–348 ms) and, to a lesser extent, P400 (348–596 ms) components, $F(1, 77) = 75.3$ and 11.1, $ps \leq .001$, partial $\eta^2 = .49$ and .13, respectively. The analyses also revealed differential amplitude for neutral/happy versus fearful facial expressions, with fearful expressions eliciting larger positivity at N290 and P400 latency ranges, $F_s(1, 77) > 11.9$, $ps \leq .001$, partial $\eta^2 = .07$ –.16. A bootstrap-based visualization of the ERP difference between nonface control stimulus and faces is presented in Fig. 5. This analysis adds to the data analysis by (1) providing CIs to the average ERP waveforms, (2) showing the time windows during which the CI of the control stimulus and face stimuli are nonoverlapping, and (3) allowing for a larger number of epochs to be used in the calculation of condition-specific ERPs.

A further analysis was conducted to examine whether any of the condition differences in the example analysis were affected by the preprocessing settings (in particular, signal detrending function). This analysis showed that, in the epochs that survived video analysis, the slope coefficients were significantly higher in the nonface control condition ($M = .03$, $SEM = .005$) than in the neutral ($M = .015$, $SEM = .005$), happy ($M = .009$, $SEM = .005$), and fearful ($M = .007$, $SEM = .005$) face conditions. The coefficients in the three face conditions did not differ

¹ A lowered criterion (≥ 5 trials/condition) was used to avoid excessive data loss. The effects of low trial counts on signal-to-noise ratio are compensated for by the relatively large sample size in the example data set ($n = 78$). Infants were excluded from the analysis for insufficient number of acceptable trials ($n = 35$), a problem in the application of the electrode net ($n = 1$), complete lack of EEG ($n = 2$) or video ($n = 7$) data due to the infant’s intolerance of the net, or technical problems, and withdrawal from the longitudinal study ($n = 4$).

significantly, all $ps > .05$. The difference between the nonface control condition and face conditions appeared to originate from a larger positive slow wave for control stimuli than for faces, especially in the later part of the analysis period. This slow wave was not visible in the primary analysis due to detrending (cf. the well-documented effect of high-pass filter on slow waves; Kappenman & Luck, 2010), but the trend is shown in the waveform that was calculated without detrending (supplementary Fig. 1). Importantly, however, the difference in the N290 response to nonface control stimulus and faces was evident in the primary and the control analysis, and the size of this effect (i.e., amplitude difference for nonfaces vs. faces) was not correlated with the difference in slope (i.e., trend coefficient for nonfaces vs. faces), $r(78) = -.13$, n.s., suggesting that the early ERP difference is independent of the condition differences in slope.

Conclusion

In this report, we have introduced a new interface for analyzing video records of infant behavior in event-related studies and integrating information obtained from the video analysis into EEG preprocessing. We have shown that video records can be successfully used to establish task compliance in visual ERP studies, detect artifacts, and reduce noise from the EEG recordings. Like any method that relies on user input and decisions, video analysis is prone to potential observer biases and errors. In our example data set, a minimal amount of training was sufficient to attain a satisfactory interrater agreement between two independent raters, but further work is required to establish whether the agreement can be improved. In our view, this method holds potential for higher levels of interrater agreement because the detection of certain behaviors is, by default, more accessible from videos, as compared with EEG signals. This is especially true with data collected from infants, given that EEG artifacts are often difficult to disentangle from the background EEG.

Our analyses also showed that, although the analysis of video records was able to confine the EEG analysis to trials that filled the criterion of task compliance, some artifacts remained in the EEG signals (in particular, low-frequency voltage drifts). We implemented and tested two different filters to remove these artifacts (causal high-pass filtering and detrending), but given the potential problems associated with the use of such filters (Kappenman & Luck, 2010; Rousselet, 2012), we emphasize that they should be used with caution and their effects tested to make informed decisions about whether to reject trials with low-frequency drift altogether or filter this component from the EEG analysis. Finally, we have discussed and demonstrated that approaches that rely on group-level analysis of the data by means of bootstrapping are useful in visualizing average ERPs and

may, in some cases, offer a viable alternative for statistical comparison of ERPs between experimental conditions.

Acknowledgments We gratefully acknowledge the efforts of the infants and parents who participated in the studies. The authors thank Charles Nelson and Alissa Westerlund for their help in creating the simulated data. This research was supported by grants from the Academy of Finland (#218284) and the European Research Council (# 283763). The authors declare no conflict of interest.

References

- Allen, E., Erhardt, E., & Calhoun, V. (2012). Data visualization in the neurosciences: Overcoming the curse of dimensionality. *Neuron*, 74(4), 603–608.
- Atkinson, J., & Braddick, O. (1976). Steroscopic discrimination in infants. *Perception*, 5, 29–38.
- Craston, P., Wyble, B., Chennu, S., & Bowman, H. (2009). The attentional blink reveals serial working memory encoding: Evidence from virtual and human event-related potentials. *Journal of Cognitive Neuroscience*, 21(3), 550–566. doi:10.1162/jocn.2009.21036
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. doi:10.1016/j.jneumeth.2003.10.009
- Dimberg, U., & Thunberg, M. (1998). Rapid facial reaction to emotional facial expressions. *Scandinavian Journal of Psychology*, 39, 39–45.
- Fujioka, T., Mourad, N., He, C., & Trainor, L. J. (2011). Comparison of artifact correction methods for infant EEG applied to extraction of event-related potential signals. *Clinical Neurophysiology*, 122(1), 43–51. doi:10.1016/j.clinph.2010.04.036
- Halit, H., Csibra, G., Volein, Á., & Johnson, M. H. (2004). Face-sensitive cortical processing in early infancy. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 45, 1228–1234.
- Hoehl, S., & Wahl, S. (2012). Recording infant ERP data for cognitive research. *Developmental Neuropsychology*, 37(3), 187–209. doi:10.1080/87565641.2011.627958
- Isler, J. R., Tarullo, A. R., Grieve, P. G., Housman, E., Kaku, M., Stark, R. I., & Fifer, W. P. (2012). Toward an electrocortical biomarker of cognition for newborn infants. *Developmental Science*, 15(2), 260–271. doi:10.1111/j.1467-7687.2011.01122.x
- Junghöfer, M., Elbert, T., Tucker, D. M., & Rockstroh, B. (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, 37(4), 523–532. doi:10.1111/1469-8986.3740523
- Kappenman, E. S., & Luck, S. J. (2010). The effects of electrode impedance on data quality and statistical significance in ERP recordings. *Psychophysiology*, 47(5), 888–904. doi:10.1111/j.1469-8986.2010.01009.x
- Martens, S., Korucuoglu, O., Smid, H. G., & Nieuwenstein, M. R. (2010). Quick minds slowed down: Effects of rotation and stimulus category on the attentional blink. *PLoS One*, 5(10), e13509.
- Nelson, C. A., & McCleery, J. P. (2008). Use of event-related potentials in the study of typical and atypical development. *Journal of the American Academy of Child & Adolescent Psychiatry*, 47(11), 1252–1261. doi:10.1097/CHI.0b013e318185a6d8
- Nyström, P. (2008). The infant mirror neuron system studied with high density EEG. *Social Neuroscience*, 3(3–4), 334–347. doi:10.1080/17470910701563665
- Palmu, K., Kirjavainen, T., Stjerna, S., Salokivi, T., & Vanhatalo, S. (2013). Sleep wake cycling in early preterm infants: Comparison of polysomnographic recordings with a novel EEG-based index.

- Clinical Neurophysiology*, 124, 1807–1814. doi:10.1016/j.clinph.2013.03.010
- Peltola, M. J., Hietanen, J. K., Forssman, L., & Leppänen, J. M. (2013). The emergence and stability of the attentional bias to fearful faces in infancy. *Infancy*, 18, 905–926. doi:10.1111/infa.12013
- Righi, G. R., Westerlund, A., Congdon, E. L., Troller-Renfree, S., & Nelson, C. A. (in press). Infants' experience-dependent processing of male and female faces: Insights from eye tracking and event-related potentials. *Developmental Cognitive Neuroscience*.
- Rousselet, G. A. (2012). Does filtering preclude us from studying ERP time-courses? *Frontiers in Psychology*, 3(131) doi:10.3389/fpsyg.2012.00131
- Rousselet, G. A., Husk, J. S., Bennett, P. J., & Sekuler, A. B. (2008). Time course and robustness of ERP object and face differences. *Journal of Vision*, 8(12) doi:10.1167/8.12.3
- Sinisalo, L., Mäki, J., Stjerna, S., & Vanhatalo, S. (2012). SurePrep, an easy alternative for skin preparation in neonatal EEG monitoring. *Acta Paediatrica*, 101(8), e378–e381. doi:10.1111/j.1651-2227.2012.02724.x
- Sorce, F. F., Emde, R. N., Campos, J. J., & Klinnert, M. D. (1985). Maternal emotional signaling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology*, 21, 195–200.
- Stjerna, S., Alatalo, P., Mäki, J., & Vanhatalo, S. (2010). Evaluation of an easy, standardized and clinically practical method (SurePrep) for the preparation of electrode–skin contact in neurophysiological recordings. *Physiological Measurement*, 31(7), 889.
- Vanhatalo, S., Voipio, J., Dewaraja, A., Holmes, M. D., & Miller, J. W. (2003). Topography and elimination of slow EEG responses related to tongue movements. *NeuroImage*, 20(2), 1419–1423. doi:10.1016/S1053-8119(03)00392-6
- Wörmann, V., Holodynski, M., Kärtner, J., & Keller, H. (2012). A cross-cultural comparison of the development of the social smile: A longitudinal study of maternal and infant imitation in 6- and 12-week-old infants. *Infant Behavior and Development*, 35(3), 335–347. doi:10.1016/j.infbeh.2012.03.002