MethodsX 11 (2023) 102378

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/methodsx

Automated preprocessing of 64 channel electroenchephalograms recorded by biosemi instruments

Ádám Kiss [a], Olívia Mária Huszár [a], Balázs Bodosi [a], Gabriella Eördegh [b], Kálmán Tót [a], Attila Nagy [a,*], András Kelemen [c,d]

[a] Department of Physiology, Faculty of Medicine, University of Szeged, Dóm Tér 10, Szeged 6720, Hungary
[b] Faculty of Health Sciences and Social Studies, University of Szeged, Szeged, Hungary
[c] Department of Applied Informatics, University of Szeged, Szeged, Hungary
[d] Department of Technical Informatics, Faculty of Science and Informatics, University of Szeged, Szeged, Hungary

A R T I C L E I N F O

Method name:
ADC-sided rereferencing

Keywords:
EEG
Rereference
MNE
Common-mode

A B S T R A C T

Preprocessing is a mandatory step in electroencephalogram (EEG) signal analysis. Overcoming challenges posed by high noise levels and substantial amplitude artifacts, such as blink-induced electrooculogram (EOG) and muscle-related electromyogram (EMG) interference, is imperative. The signal-to-noise ratio significantly influences the reliability and statistical significance of subsequent analyses. Existing referencing approaches employed in multi-card systems, like using a single electrode or averaging across multiple electrodes, fall short in this respect. In this article, we introduce an innovative referencing method tailored to multi-card instruments, enhancing signal fidelity and analysis outcomes. Our proposed signal processing loop not only mitigates blink-related artifacts but also accurately identifies muscle activity. This work contributes to advancing EEG analysis by providing a robust solution for artifact removal and enhancing data integrity.

- Removes blink
- Marks muscle activity
- *Re*-references with design specific enhancements

Specifications table

Subject area:	Neuroscience
More specific subject area:	*EEG signal processing*
Name of your method:	*ADC-sided rereferencing*
Name and reference of original method:	*https://doi.org/10.1007/s10548–019–00707-x*
Resource availability:	*https://doi.org/10.5281/zenodo.8272584*

Method details

The biosemi64 system

The EEG system developed by Biosemi utilizes 32-channel Analog-to-Digital Converter (ADC) cards. Each card is positioned within the AD box [1], and each extension is associated with its distinct cable bundle. However, the presence of separate ADC cards and

* Corresponding author.
E-mail address: nagy.attila.1@med.u-szeged.hu (A. Nagy).

https://doi.org/10.1016/j.mex.2023.102378
Received 22 August 2023; Accepted 12 September 2023

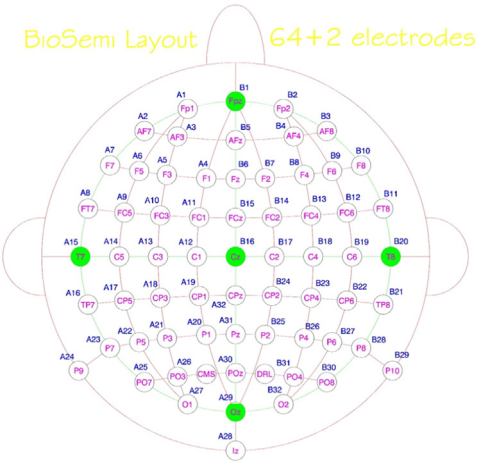

Fig. 1. Image of the recording setup and an electrode map [3]. Data channels marked with 'A' are sampled by one ADC card, while those marked with 'B' from the other.

the utilization of dual-sided cabling gives rise to the emergence of a lateralized common mode component. To address this, a driven right leg approach [2] is implemented in the system's design. By virtue of its controlled output current and its accommodation of the capacitance inherent in the human body, the driven right leg mechanism ensures that the two common-mode components remain within the acceptable input range of the ADCs, albeit without complete elimination (Fig. 1).

In this setup, the EEG signals are sampled at 2048 Hz.

Muscle activity

Individually wired motor units generate a composite signal resulting from the superposition of multiple muscle-fiber action potentials. This electromyographic (EMG) signal often exhibits comparable or even greater amplitude than the EEG component in the recording. In accordance with the Central Limit Theorem, the summation of these underlying elementary EMG signals should theoretically yield a distribution that approximates a Gaussian curve.

Blinks

Experienced EEG specialists often rely on their intuition to select Independent Components (ICs) [4]originating from blinks, a process strongly influenced by subjectivity. In this method, we propose an objective approach for blink artifact removal that is both autonomous and reproducible. This involves applying simple image processing steps to the topoplots. The core concept is to identify ICA components primarily situated in the frontal lobe, characterized by the composite contours resulting from the superimposition of elemental waves from the levator palpebrae superioris muscle. Utilizing line-fitting algorithms, we can discern the distinct linearity inherent in these blink-related wavefronts.

Fig. 2 illustrates the Independent Component Analysis (ICA) weights of the components. In the heatmap, the red or 'hot' end represents the highest in-phase amplitudes, while the blue or 'cold' end signifies the highest counter-phase amplitudes. Given the diverse algorithms underlying a wide range of ICA implementations, the in-phase and counter-phase weights may interchange. To mitigate this phenomenon and ensure a consistent component order for reproducible outputs, we introduce a fixed and constant random seed. This approach eliminates potential variability in the results. The highlighted components exhibit frontal dominance; interestingly, despite this, human judgment categorizes them as non-blink-related.

The recorded signal on a channel

The superposition of various physical phenomena can be observed on each registered channel. Predominantly, this encompasses the vector summation of electrotonus potentials originating from cortical activity. The component perpendicular to the scalp's surface is directly projected onto the corresponding electrodes. Adjacent electrodes also capture this signal, albeit with an amplitude decrease proportional to the sine of the angle of incidence. Consequently, the adjacent channels lack complete independence. The cable bundle accumulates additional environmental noise referred to as the common-mode component. Distinct path variations result in differing common-mode components for each cable bundle. Each bundle terminates in its dedicated amplifier card, further distinguishing between common-mode noises and the bilaterally situated nature. The superposition of muscle activities and blink-related electrophysiological signals is also observable on the recorded EEG channels. Additionally, each recorded data channel is subject to noise originating from the active electrode, the amplifier, the ADC itself, and possibly other contributing factors.

$$\begin{aligned}\text{Recorded signal} = {} & \text{EEG signal of the recording electrode} + \text{crosstalk signal of the adjacent electrodes} + \text{common mode noise} \\ & + \text{own noise} + \text{signals of muscle activity} + \text{blinking artifacts}\end{aligned}$$

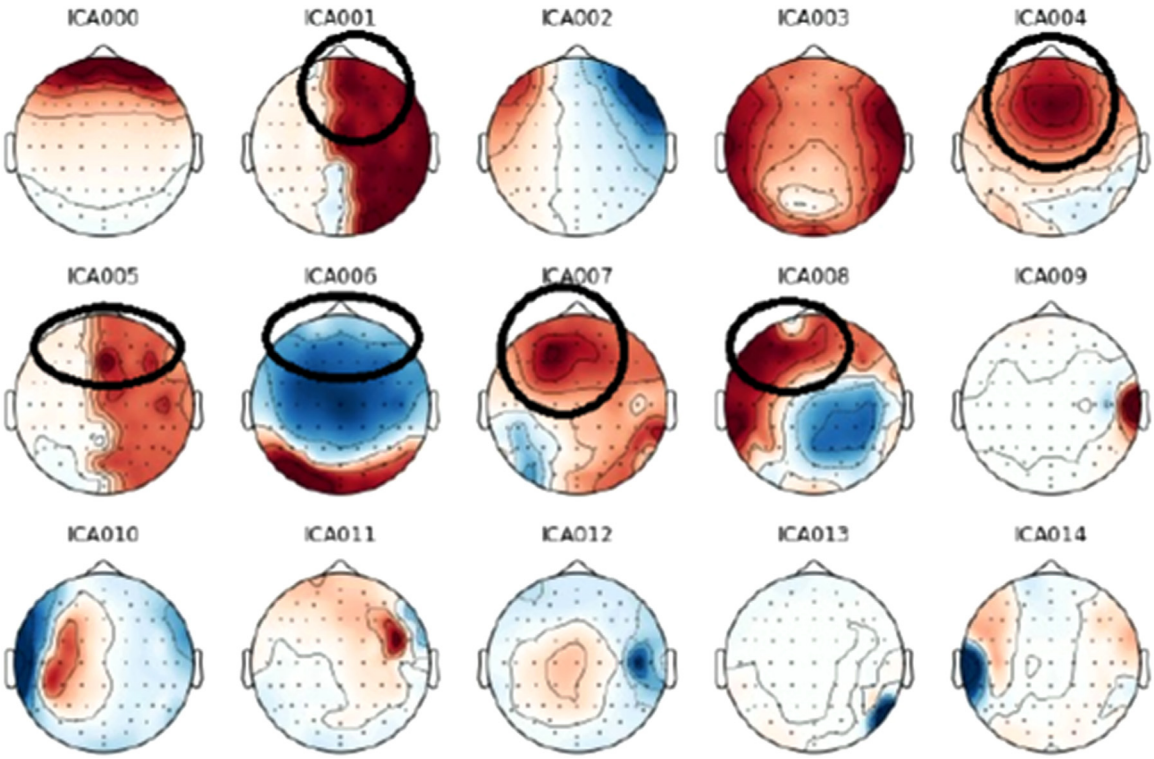

Fig. 2. 15 ICA components of the EEG recording with the highest variance.

The goal of the presented algorithm is to mitigate the impact of common mode noise, muscle activity, and blink-related components. This algorithm, as detailed in this study, offers an automated and highly effective method for eliminating both common mode noise and blink artifacts from the signal. While the muscle component can be automatically detected, it remains challenging to fully compensate for it. In cases where EMG dominance is identified, these segments are appropriately marked for exclusion from subsequent analyses.

Muscle activity detector

In our experience, muscle activities manifest prominent components with high amplitude frequencies ranging between 350 and 650 Hz within their power spectrum. To address this, the raw data is subjected to a filtering process within this frequency band. Subsequently, signal power within this filtered frequency range is computed using a simple squaring operation applied to the filtered samples.

To derive the average signal power, a 0.05 s equiweight finite impulse response filter is employed. Following this, a median filter is implemented using an eleven-bin buffer. The establishment of a baseline level involves computing the average of the previously determined power function, excluding the lowest 0.1 s' worth of elements (based on dataset samples) and the upper half of the sorted samples.

The calculated threshold is obtained by multiplying the baseline level by 15, resulting in a 12 dB event-to-noise ratio, as outlined in reference [5]. The determination of whether a recorded moment primarily represents muscle activity hinges on the number of channels surpassing the threshold. When this count exceeds two, a muscle event is initiated and persists until no channels exceed the threshold. Furthermore, to ensure robustness and accuracy, the identified intervals are symmetrically extended by one second to eliminate ambiguity regarding the boundaries of muscle activity artifacts. Intervals identified through this process, which indicate the presence of muscle activity, are subsequently excluded from the evaluation of the recording.

Blink component selector

The blink components can be identified through the outcome of an Independent Component Analysis (ICA).

1. Generate topo plots depicting the spatial distribution of ICA component weights using default settings from python MNE [6] version 1.4.2. The width of the whole plot should be 7.5 inches, while the height should be 5.7 inches with 100 DPI precision. In this depiction, positive in-phase weights are in red, while negative counter-phase component weights are in blue (Fig. 3).

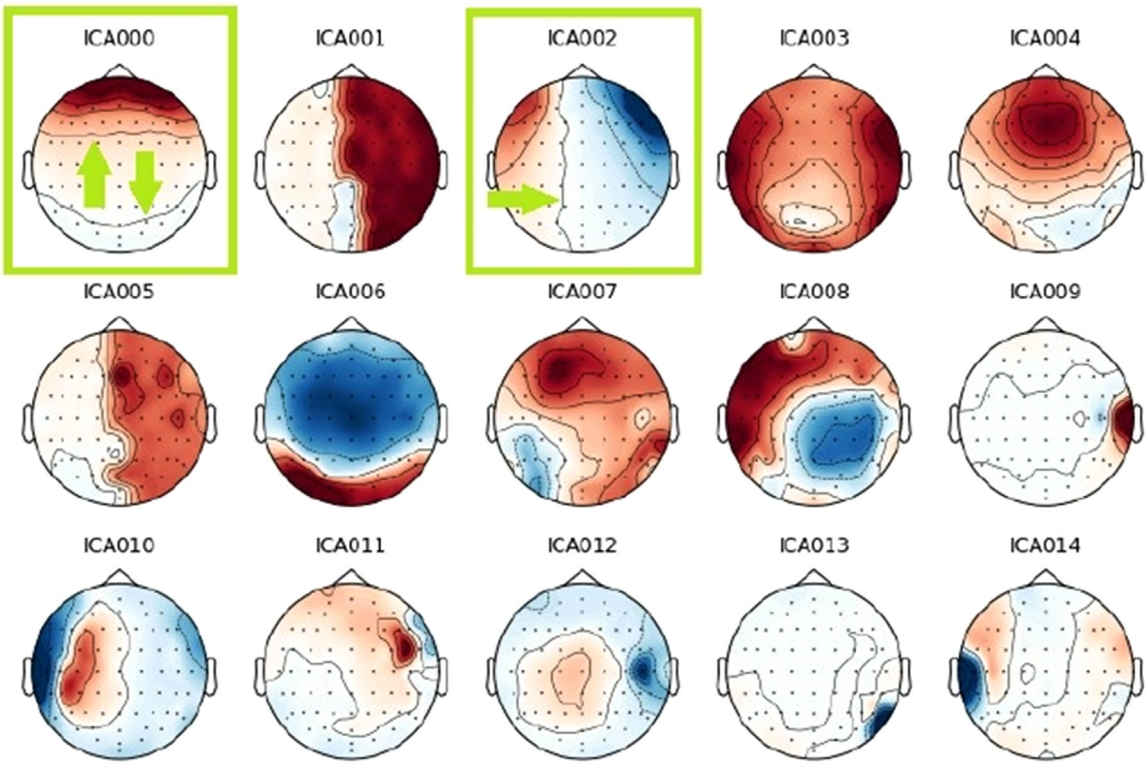

Fig. 3. The first 15 ICA components from the recording, indicating manually identified blink components.

2. Calculate the bounding circle for each ICA component to facilitate independent processing in subsequent steps.
3. In the Hue-Saturation-Value color space, isolate portions of the components based on their hue value. Pixels with hue values above 105 (out of 255) are selected, rendering all other pixels white (Fig. 4).
4. Convert the thresholded image to a black-and-white format and apply a median filter with a 3 × 3 dimension (Fig. 5).
5. Exclude subfigures of the ICA components (as they are not blink-related according to our algorithm) in the following cases:
    1. if the frontal lobe section exhibits more hierarchical components than a specified threshold.
    2. if the subfigures are ranked as having the highest or second-highest count of white pixels.
    3. if in the absence of the frontal lobe, the count of white pixels falls below a threshold of 600. (Fig. 6).
6. Perform iterative erosions and dilations, along with edge-detection, followed by a Hough Line Transformation on the remaining topoplots. If a line is detected with a minimum length of 45 pixels, it can be inferred that the topoplot corresponds to a component originating from a blink (Fig. 7).

Re-*referencing method*

The fundamental idea is that the blink component exhibits an anterior-posterior imbalance, whereas the common mode difference presents a lateral imbalance. Reiterating certain steps enhances the precision of their estimation. The frequency of step repetitions directly influences the accuracy of resultant estimations (whether for blink or common mode components). It's important to note that excessive iterations may potentially accumulate numerical errors.

1. First, apply a band-pass filter to the recording, setting corner frequencies at 1 Hz and 100 Hz.
2. In the initial stage, compute the common mode for each AD card by averaging its posterior channels. Channels unrelated to blinks are included in this summation, rendering blinking effects irrelevant.
    1. Calculate the mean of non-blink-related channels for each amplifier card.
    2. Perform a subtraction to remove the common mode from each channel.
    3. Employ the fast Independent Component Analysis (ICA) algorithm to estimate independent components. Remove blink-related components from a copy of the original recording. Subsequently, work with this modified recording.
3. In the subsequent stage, determine the common mode for each AD card by averaging all of its channels. Note that the initially estimated blink component is already subtracted from the working dataset.
    1. Calculate the mean of every channel for each amplifier card.

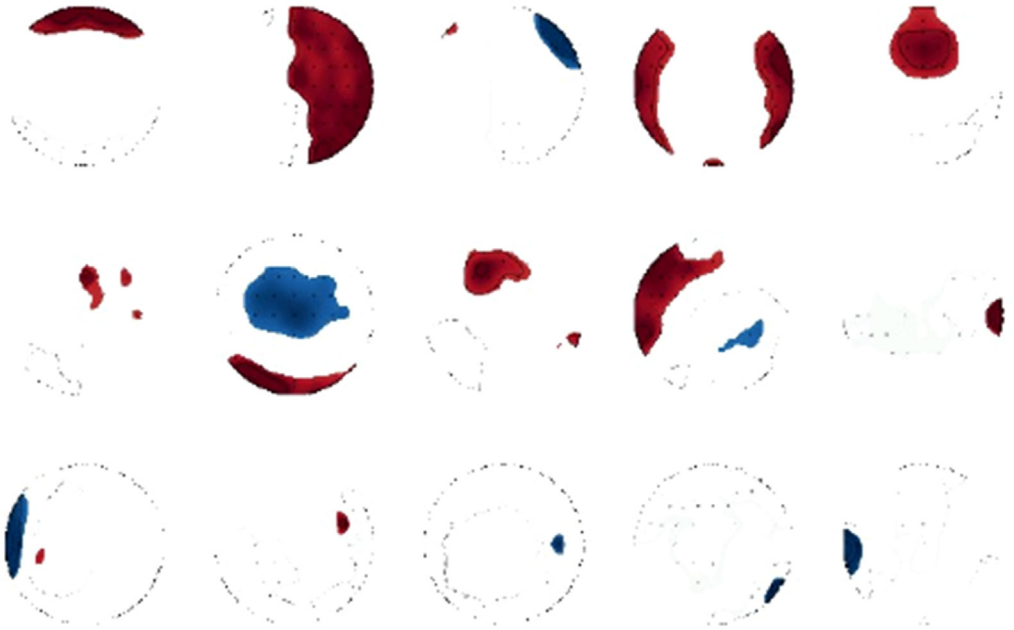

Fig. 4. Selected high weight areas of the ICA components.

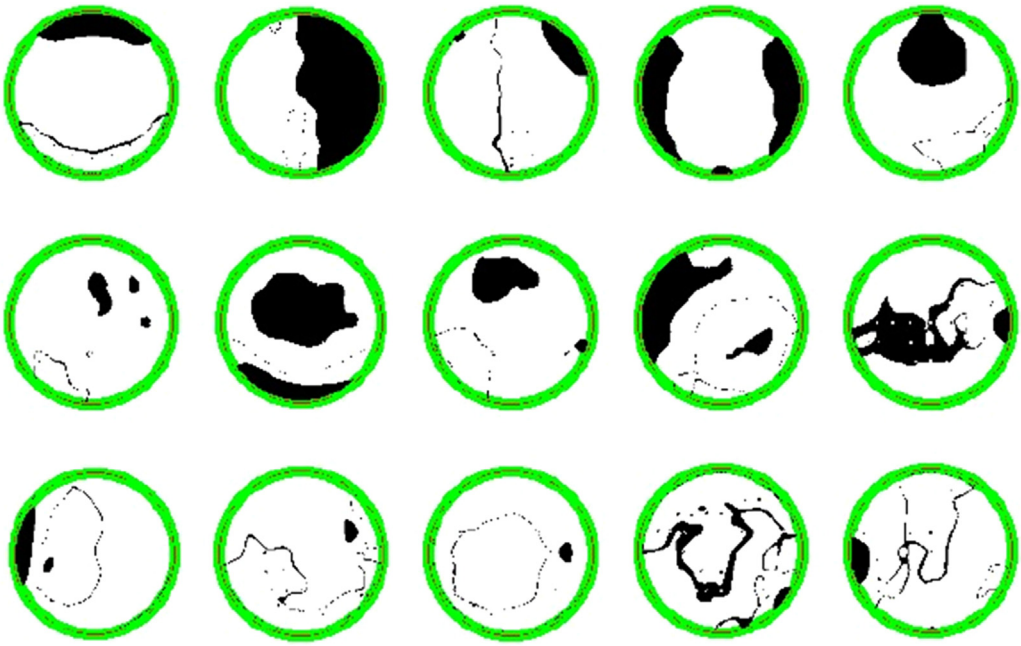

Fig. 5. Median filtered black and white image. The black areas are the region of interest.

Fig. 6. The remaining plots after the selection of point five.

Fig. 7. Outcome of line detection marked in green. The left and middle components are identified as blink-related. White areas stem from ICA components, while purple regions originate from the bounding circle.

2. Implement a gain-compensated subtraction method to remove the common mode from each channel. This method employs a least-means square line-fitting algorithm. Its result should have the intercept point of approximitly zero due to the applied high-pass filtering, and the slope around one, indicating the different common-mode ratio of each channel. The subtraction is weighted by the slope to ensure compensation for the common mode component in each channel.
3. Reapply the fastICA algorithm to estimate independent components and eliminate blink-related ones from the copy of the original recording. Continue working with this modified recording.

4. Repeat this stage twice.
5. In the third stage, calculate the common mode for each AD card by averaging all channels except the surrounding ones. This step minimizes overlap between valuable signals and estimated common mode noise.
   1. Compute the mean of every channel for each amplifier card, excluding surrounding channels.
   2. Use gain-compensated subtraction to eliminate the common mode from each channel.
   3. Reapply the fastICA algorithm to estimate independent components and remove blink-related ones from the copy of the original recording. Continue working with this modified recording.
6. Repeat this stage twice.
7. Eliminate the last estimation of common-mode noise, ensuring both the most accurate blink component and the common-mode noise are removed.
8. Calculate the power spectral density for each channel with 1 Hz resolution, resulting in a Power(channel, frequency) function. Apply the Grubb's statistical test step-by-step from 1 to 20 Hz, with 64 statistical samples for each frequency from the channels. Identify outliers. If a channel is an outlier in more than 15 frequencies of the 20 Hz bandwidth, mark it as bad in the output file, and consider dropping or interpolating it in further evaluation processes.
9. Include metadata such as timestamps of muscle activities and marked bad channels in a description field. Save the results.
10. Conduct manual examinations as needed on the obtained results.

Further improvements

In the future, the algorithm could be expanded to include other types of empty or bad channel detection.

Conclusion

The described algorithm is suitable for routine application in everyday research involving 64-channel, multi-ADC card EEG studies during preprocessing. When executed on a standard desktop computer, the runtime remains below ten minutes for a five-minute recording.

Examples are presented in the supplementary section.

Ethics statements

The study conformed to the Declaration of Helsinki in all respects and was approved by the Regional Research Ethics Committee for Medical Research at the University of Szeged, Hungary (27/2020-SZTE) and the National Center for Public Health and Pharmacy, Hungary (11,818–4/2017EÜIG).

Data availability

Data will be made available on request.

CRediT authorship contribution statement

Ádám Kiss: Software, Writing – original draft. **Olívia Mária Huszár:** Software. **Balázs Bodosi:** Software. **Gabriella Eördegh:** Resources. **Kálmán Tót:** Resources. **Attila Nagy:** Project administration, Writing – review & editing. **András Kelemen:** Supervision, Writing – review & editing.

Acknowledgments

Attila Nagy was supported by University of Szeged [SZTESZAOK-KKA-SZGYA2023/5S479].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mex.2023.102378.

References

[1] Biosemi Activetwo AD-box url: (2023) https://www.biosemi.com/ad-box_activetwo.htm.
[2] B.B. Winter and J.G. Webster, “Driven-right-leg circuit design,” in IEEE Transactions on Biomedical Engineering, vol. BME-30, no. 1, pp. 62–66, Jan. 1983, doi:10.1109/TBME.1983.325168.
[3] Biosemi 64 chennel EEG cap layout url: 2023 https://www.biosemi.com/pics/cap_64_layout_medium.jpg.
[4] A. Hyvarinen, E. Oja, Independent component analysis: algorithms and applications, Neural Netw. 13 (4–5) (2000) 411–430.
[5] Á. Kiss, L. Dudás, A tracking method in fm broadcast-based passive radar systems, AARMS 20 (1) (2021) 67–79 Sep., doi:10.32565/aarms.2021.1.5.
[6] A. Gramfort, MEG and EEG data analysis with MNE-Python, Front. Neurosci. 7 (2013), doi:10.3389/fnins.2013.00267.