

Independent Component Analysis as a Tool to Eliminate Artifacts in EEG: A Quantitative Study

*Jorge Iriarte, *Elena Urrestarazu, *[†]Miguel Valencia, *Manuel Alegre, [†]Armando Malanda, *César Viteri, and *Julio Artieda

*Clinical Neurophysiology Section, Clínica Universitaria, University of Navarra and [†]Department of Electrical and Electronic Engineering, Public University of Navarra, Pamplona, Spain.

Summary: Independent component analysis (ICA) is a novel technique that calculates independent components from mixed signals. A hypothetical clinical application is to remove artifacts in EEG. The goal of this study was to apply ICA to standard EEG recordings to eliminate well-known artifacts, thus quantifying its efficacy in an objective way. Eighty samples of recordings with spikes and evident artifacts of electrocardiogram (EKG), eye movements, 50-Hz interference, muscle, or electrode artifact were studied. ICA components were calculated using the Joint Approximate Diagonalization of Eigen-matrices (JADE) algorithm. The signal was reconstructed excluding those components related to the artifacts. A normalized correlation coefficient was used as a measure of the changes caused by the suppression of these components. ICA produced an evident clearing-up of signals in all the samples. The morphology and the topography of the spike were very similar before and after the removal of the artifacts. The correlation coefficient showed that the rest of the signal did not change significantly. Two examiners independently looked at the samples to identify the changes in the morphology and location of the discharge and the artifacts. In conclusion, ICA proved to be a useful tool to clean artifacts in short EEG samples, without having the disadvantages associated with the digital filters. The distortion of the interictal activity measured by correlation analysis was minimal. **Key Words:** EEG—Artifacts—Independent component analysis.

EEG signals are a complex summation of electrical brain activity that are recorded by distant electrodes according to the solid angle and the volume conduction laws. The origin of every signal may be very complicated, and therefore, its exact meaning is not clear. EEG activity can be severely contaminated by eye movements, blinking, muscle, and heart and line noise, among

others. These artifacts constitute a serious problem for the interpretation and analysis of the EEG signals. Several methods have been used to minimize them but with limited success. Digital filters can eliminate waves of a selected frequency range or band. Their main limitation is that they do not discriminate between artifact and brain waves, thus removing all activity in the frequency selected. As a consequence, they may distort the real morphology of the brain activity. Regression in the time or frequency domain can be performed in EEG and electrooculogram channels to remove eye movement artifact, but it cannot be used to remove muscle or line noise. The use of principal component analysis cannot completely separate artifacts with similar amplitudes than that of the brain signal (Ille et al., 2002; Jung et al., 2001;).

This study was supported by the Government of Navarra, grants for research in Health 12/2003 and 15/2003. Dr. Urrestarazu is a Fellow supported by a grant for research from the Department of Education of the Basque Government.

Address correspondence and reprint requests to Dr. Jorge Iriarte, Clinical Neurophysiology Section, Department of Neurology, Clínica Universitaria, University of Navarra, Avda Pío XII 36, Pamplona 31008, Navarra, Spain; e-mail: jiriarte@unav.es.

Independent component analysis (ICA) is a technique that exploits independence to extract an estimation of the sources from mixtures of signals optimizing some contrast function (Brown et al., 2001; Hyvarinen and Oja, 2000). Independent component analysis neural approaches use a suitable nonlinearity in the learning algorithm (Karhunen et al., 2000). It was originally planned as a procedure to separate different voices speaking at the same time (“cocktail party problem”). Later on, its use was extended to other kinds of signals. In pregnant women, it has been used to separate the electrocardiogram (EKG) of the mother and the child (De Lathauwer et al., 2000). It has also been used for MRI, positron emission tomograms, audiograms, and others (Arfanakis et al., 2000; Calhoun et al., 2001; Hoyer and Hyvarinen, 2000; Lee et al., 2001; Suzuki et al., 2002). Independent component analysis was recently applied to EEG and magnetoencephalography for research purposes with encouraging results (Ikeda and Toyama, 2000; Maeda et al., 2001; Moritz et al., 2000; Sato et al., 2001; Tong et al., 2001; Vigario et al., 2000). Also, ICA has been used in combination with a variant of MUSIC (Multiple Signal Classification) as a new approach to the dipole source estimation (Kobayashi et al., 2002a, b). Most of the published studies have been focused on the elimination of ocular artifacts, because they are easier to identify (Jung et al., 2000a, b).

The aim of our study was to objectively test the capability of ICA to eliminate different artifacts in routine EEG recordings without altering the rest of the signal, thus comparing artifact-free and ICA-cleaned fragments of interictal EEG activity from monitored epileptic patients.

METHODS

Two series of samples of interictal EEG (between 4 and 10 seconds each) were selected from two different patients monitored in our video-EEG unit. Patient 1 had symptomatic frontal epilepsy. Patient 2 had focal epilepsy caused by unilateral mesiotemporal sclerosis. Both patients had frequent spikes with a constant topographical distribution in artifact-free recordings. Forty EEG fragments, including at least one spike and one or several combined typical artifacts (eye movements, EKG, line noise, muscle, electrode artifact), were selected from each subject. Also, 10 artifact-free fragments, including one or several spikes, were selected from each patient. Only fragments with easy-to-identify artifacts were selected. The samples were selected visually in the original recordings by one of the authors (J.I.). The EKG artifact was a rhythmic spike around 1 Hz. Eye movement

artifact was noted on the frontopolar electrodes as a large delta wave. The 50-Hz artifact was corroborated by frequency analysis. The muscle artifact was seen as bilateral fast spikes lasting for several seconds. The electrode artifact was identified as a problem in a single electrode and had a very different signal of unequivocal noncerebral origin.

The EEG recordings were acquired using the standard video-EEG equipment in our epilepsy unit (64-channel digital EEG with LaMont amplifiers (LaMont Medical, Madison, WI, U.S.A) and Harmonie software (Stellate, Montreal, Quebec, Canada)). Electrodes were placed according to the 10 to 20 system, with additional anterotemporal electrodes in T1/T2 and with both mastoids as reference. The samples were analyzed in the original monopolar montage with the average of both mastoids as the common reference. All the off-line analysis was carried out on a personal computer by means of specific software developed in the MATLAB (TheMathWorks, Inc., Natick, MA, U.S.A.) environment.

The ICA components were calculated using the Joint Approximate Diagonalization of Eigen-matrices (JADE) algorithm (Bartlett et al., 1995; Cardoso, 1998). It is based on the diagonalization of cumulant matrices. The use of cumulants of order 2 and 4 demonstrate good statistical performance, and a fast optimization is obtained by joint diagonalization. Joint Approximate Diagonalization of Eigen-matrices algorithm has been successfully applied to the processing of real data sets, such as mobile telephony and airport radar, as well as in biomedical signals. The strongest point of JADE for elemental applications of ICA is that it works with any parameter-tuning requirement. The shortcoming of the current implementation is that the number of independent components, but not of sensors, is limited in practice by the available memory, depending on the computer it is run on. Because of this, the number of possible components was fixed at 20, as a first trial. If the separation was not good enough, it was repeated with the maximum number of possible components (the same as original recording electrodes).

For EEG independent component analysis, the rows of the \mathbf{x} input matrix are the EEG signals recorded at different electrodes; the rows of the output data matrix $\mathbf{u} = \mathbf{W}\mathbf{x}$ are the time courses of activation of the ICA components (sources estimations), and the columns of the inverse matrix \mathbf{W}^{-1} give the projection strengths of the respective components onto the scalps sensors. “Filtered” EEG can be derived as $\mathbf{x}' = \mathbf{W}^{-1}\mathbf{u}'$, where \mathbf{u}' is the matrix of activation waveforms, \mathbf{u} , with rows

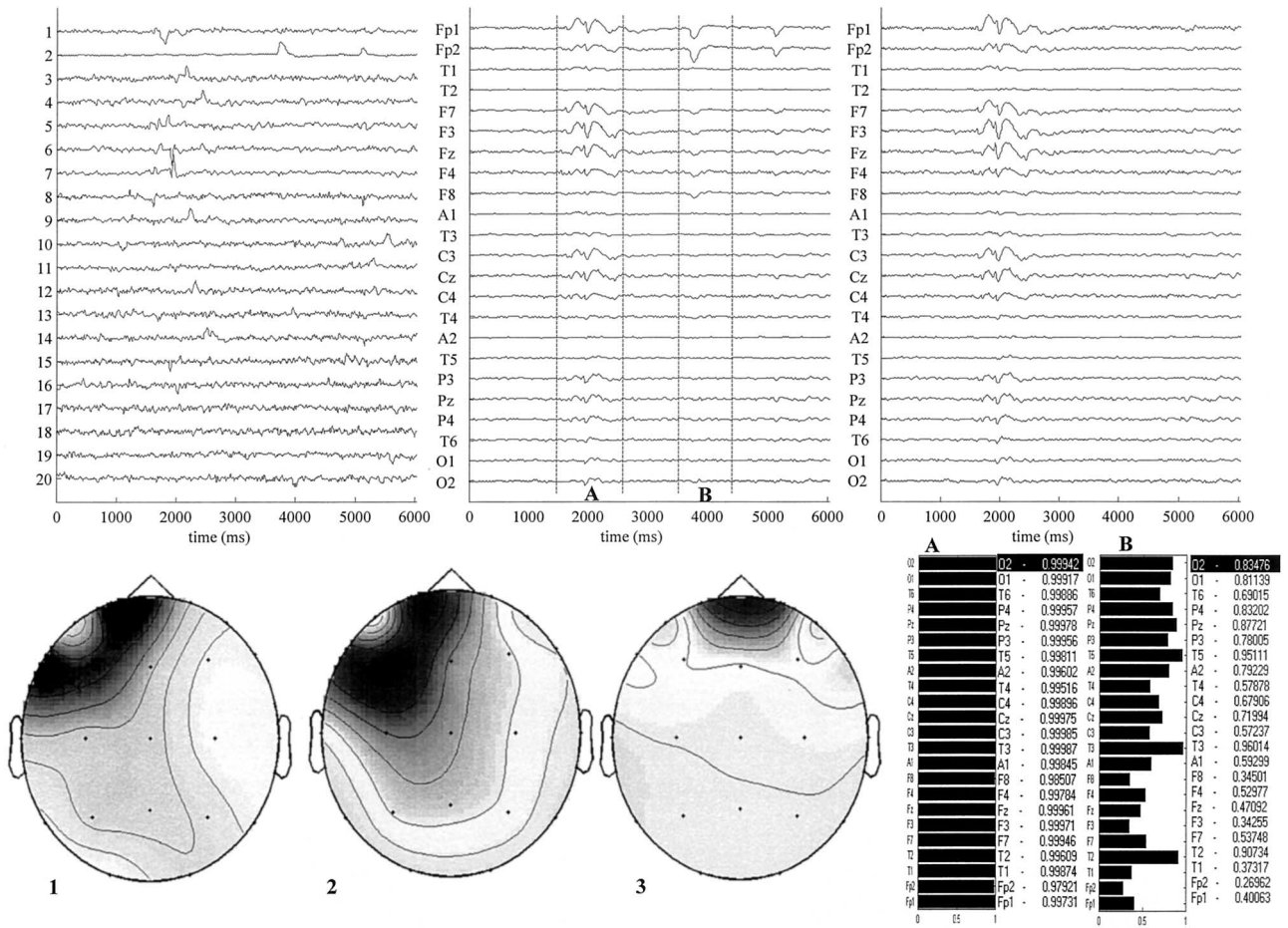


FIG. 1. In the upper left corner we show the components extracted from the original EEG (center). On the right, EEG after the removal of the eye movements, in the same referential montage. On the bottom, maps reconstructed from the original EEG at the time of the spike (1), from the component (component 4) of the left fronto-central spike-and-wave (2), and from the component (component 2) of the eye (3). To the right of the maps, correlation between the original and the cleaned EEG is shown: (A) segment with the spike, (B) segment with the eye movement (columns represent the numeric value).

representing artifactual sources set to 0. The scalp topographies of the components can be taken as an indicator of the biologic origin of the sources.

Once the JADE components were calculated, the components that might account for the artifact were identified, considering their typical morphology, frequency spectrum, topography, and timing. After that, the signal was reconstructed, excluding the component or components supposedly related to the artifacts. The topography of the spikes (isopotential maps) after the reconstruction was compared with the topography of the spikes in the artifact-free samples from the same subject. Finally, a normalized correlation coefficient between the original and the cleaned signal for each selected segment was calculated. This coefficient informed about the changes in each channel after remov-

ing the artifacts. Also, in those sweeps that had mains artifact (50 Hz in Europe), the frequency spectra were computed before and after the elimination of the corresponding component. The off-line analysis was done blind by one of the authors (J.I.). Each sample was saved with a numeric identification, because it was impossible to visually identify the source (patient and artifact). He saved the figures for the examination in PowerPoint files and identified the originals and the cleaned EEG with the number exclusively. Two independent examiners (J.I., E.U.) reviewed the initial samples and the EEGs after the elimination of the artifact to also visually identify the presence of the same morphology and location of the spike and the elimination of the artifact. They had to identify the spike (i.e., the patient) and the kind of artifact.

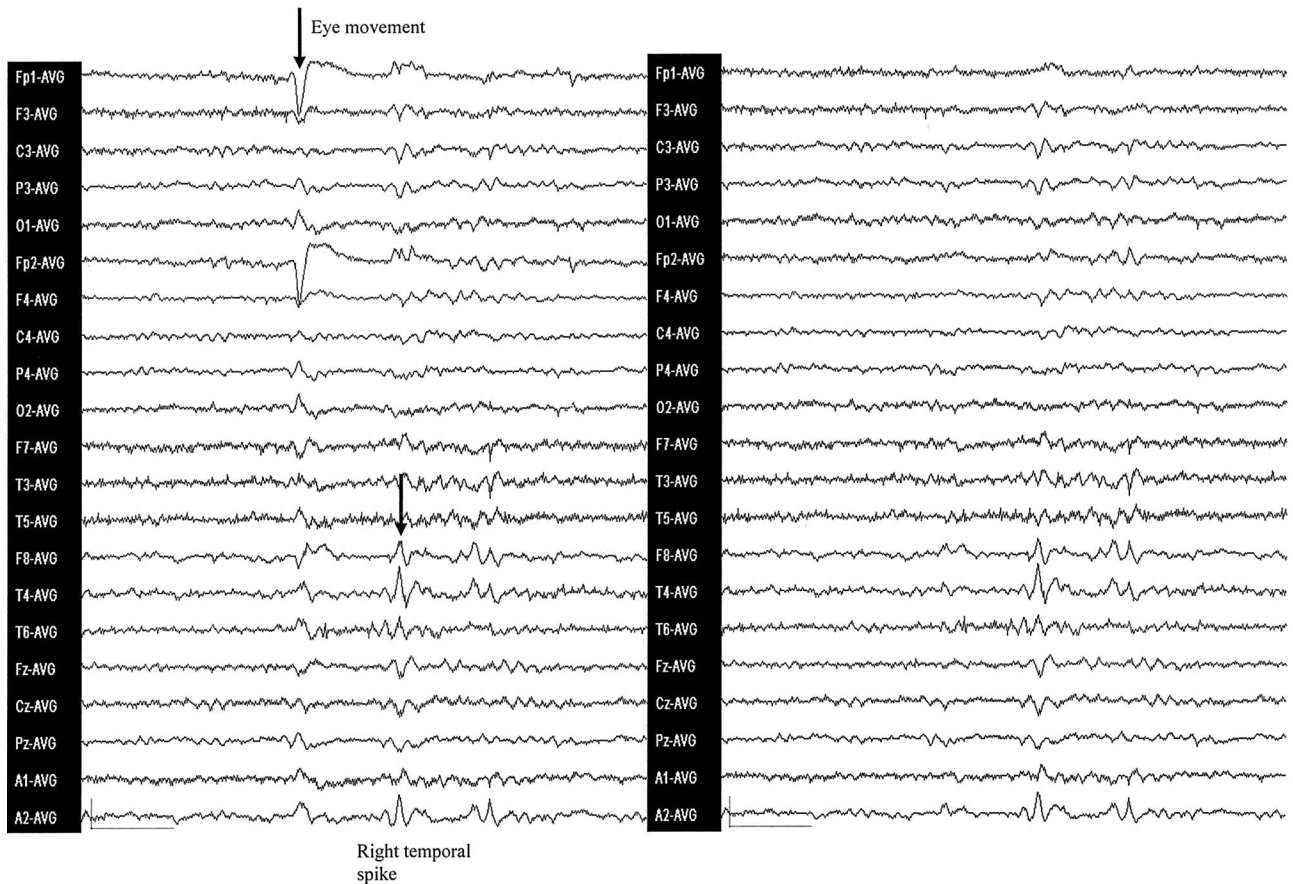


FIG. 2. Original EEG in Harmonie format. Removal of the eye blinking. Scale: 100 μ V, 1 s. The eye artifact is removed while the other waves remain unchanged.

RESULTS

The computation of the JADE components in each sample was fast (about 4s in a 800 mol/LHz PC). Artifact sources were identified easily; in some cases, the process of selecting and eliminating components had to be repeated several times until a clear suppression of each artifact was obtained. The removal of the component of the artifact produced an evident clearing-up of the signals in all the samples. The two examiners were in complete agreement in the identification and location of the spike, before and after the removal of the artifact, and in the nature of the artifact. The results obtained with the different artifacts will be discussed separately.

Eye movements

The component related to the eye movement artifact was easily identified in all cases (component 2, in Fig. 1 in the left corner). The topographical projection

of this component showed a maximum in the frontopolar region, as seen in Fig. 1 on the bottom left. However, the component responsible for the spike (component 3) had a topographical projection with a negativity on the left fronto-central area that was very similar to the topography of the spike seen on the clean EEG.

The morphology and the topography of the spike were similar before and after removing the artifact. When the component 2 was removed, the EEG changed significantly, and the eye movement waves from Fp1 and Fp2 disappeared (Fig. 1). The correlation for the segment of the spike or spike-and-wave (A) preelimination and postelimination of the eye artifact was very high ($r > 0.95$) (Fig. 1, bottom, A). The correlation for the segment of the eye movement (B) was very low in the frontal electrodes ($r < 0.40$) (Fig. 1, bottom, B). In Fig. 2, we show an example using the original Harmonie file from patient 2. The morphology and the distribution of the spike-and-

F2

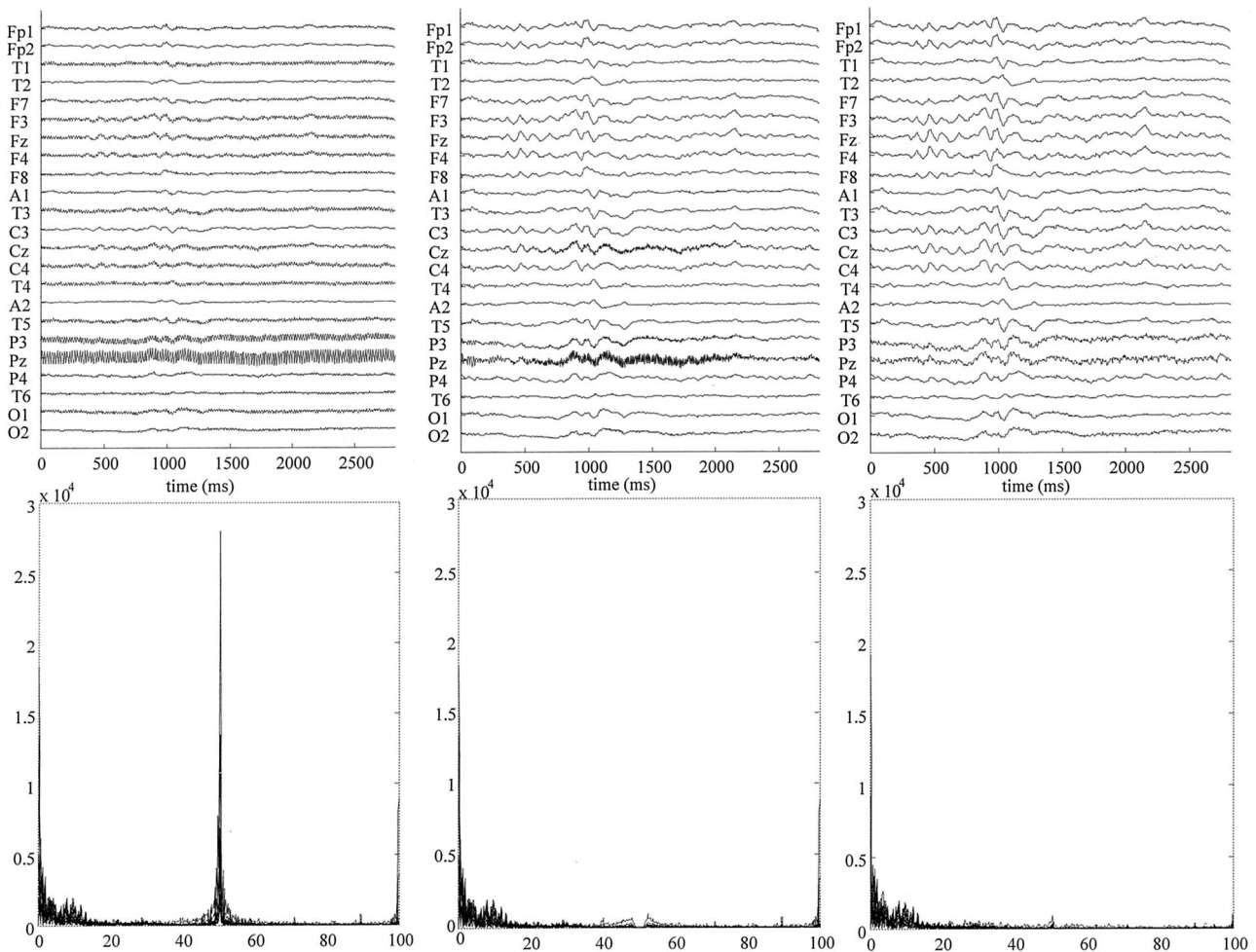


FIG. 3. Removal of the 50-Hz artifact from several EEG channels. Original EEG (left), EEG cleaned a digital filter of 50 Hz (center), and EEG cleaned using ICA (right). On the bottom, the power spectrum of the corresponding EEGs is shown. The 50-Hz filter completely eliminated the 50-Hz peak without changing the rest of the spectrum. The removal of the 50-Hz component using ICA produced a big but not total elimination of the 50 Hz activity, and also it eliminated the 100-Hz artifact (cleaning most of the artifact in the electrode Pz).

wave is similar to the spike-and-wave of Fig. 1, but the eye movement was completely cleaned.

50-Hertz artifact

The mains (50 Hz) artifact was present in most of the samples, in a variable degree. One or two components accounted for this artifact and were very easily identified by means of the frequency spectrum. When removed, a cleaning of all the channels was noted. The morphology of the other waves under the artifact did not change significantly.

Figure 3 shows an example. The EEG resulting from the removal by ICA of the 50 Hz mains component was cleaner than the EEG after the use of a 50 Hz digital filter (above, center and right). The spectrum of the original

EEG signal showed a large peak at 50 Hz (bottom, left). After the removal of the 50-Hz component, the spectrum showed a specific reduction of the 50-Hz peak (bottom, center) and also showed the elimination of the harmonic at 100 Hz. The 50-Hz digital filter eliminated all of the 50 Hz power (bottom, right), which created an artificial gap in the spectrum.

Electrocardiogram

The EKG artifact was also clearly identified within the JADE components from the original EEG. An example is shown in Fig. 4, with a prominent EKG artifact in electrode C3 (left). The calculated components are shown in the second column. Component 4 corresponds to the EKG. After its removal (third column), the EEG

F4

F3

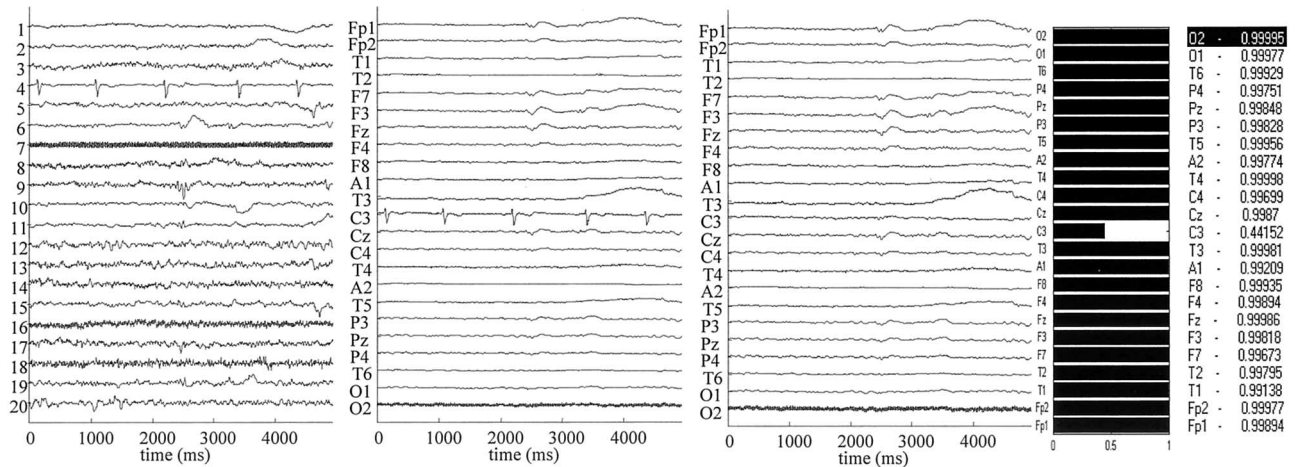


FIG. 4. Original EEG and the EEG after the removal of the EKG from a channel (C3). On the left are the separated components and the original EEG. Component 4 accounts for the EKG. On the two columns on the right the final EEG and the correlation between the two EEGs is shown. The removal of component 4 only modifies the channel C3 ($r = 0.44$, while on the other channels it is higher than 0.99).

was clean from the EKG artifact. The correlation coefficient showed a very high correlation (higher than 0.99) in all the electrodes except in the channel with the EKG artifact (0.44 for C3). The morphology and the topography of the spike present in the recording remained unchanged. The results were similar in the rest of the samples tested.

Electrode artifact

The component responsible for electrode artifacts was identified in all cases, although with more difficulty than with the previous ones. The removal of this component produced a great change in the electrode with the artifact and kept a high correlation with the other channels. Figure 5, above, shows a sample of EEG with artifact in O1 and O2. After removing the components for these artifacts, the other channels did not change. Correlation values were very high for all the electrodes except O1 and O2 (higher than 0.99, and 0.18 and 0.33, respectively). The left frontal spike remained unchanged except in O1, where it became visible after the suppression of the artifact (as it was in artifact-free samples).

Muscle artifact

Muscle artifact was more difficult to suppress than the rest of the artifacts. Usually, more than one component accounted for it, and the correlation obtained after its suppression was slightly lower in nonartifactual channels. Figure 5, below, shows an example of the removal of muscle artifact from several channels (T4, T3, F7, F8, C3, C4) with small changes in the others. The correlation found was low for the electrodes F7, F8, T3, T4 and T2 (0.45 to

0.53). The morphology and topography of the spike present in the recording remained unchanged instead of the small changes observed in the channels where it was present.

Figure 6 compares the removal of the muscle artifact using digital filters (HF 35, LF 0.3, 50-Hz, center) and ICA (right) in a different sample. The suppression of the muscle artifact was better with ICA in all the samples analyzed.

DISCUSSION

Independent component analysis is a novel tool that can extract independent components from recorded signals. Independent component analysis includes a group of algorithms that take advantage of independence in a complex signal to perform blind source separation (Hyvarinen et al., 2001; Stone, 2001). This method has been applied in several areas of physics and medicine. Brain signals recorded by EEG or magnetoencephalography are mixtures of signals and therefore can be analyzed using ICA. In fact, Independent component analysis may be very helpful in identifying different components that correspond to different brain generators.

The results of our study show that ICA can be a useful tool for identifying and eliminating conventional EEG artifacts in routine EEGs without altering the epileptiform activity. High voltage artifacts such as eye movements and EKG were easily removed; ICA was able to identify a single component for them with a typical morphology very similar to the artifact itself. The 50-Hz artifact was also easy to separate and eliminate. This is probably the reason why the first studies using ICA for

INDEPENDENT COMPONENT ANALYSIS

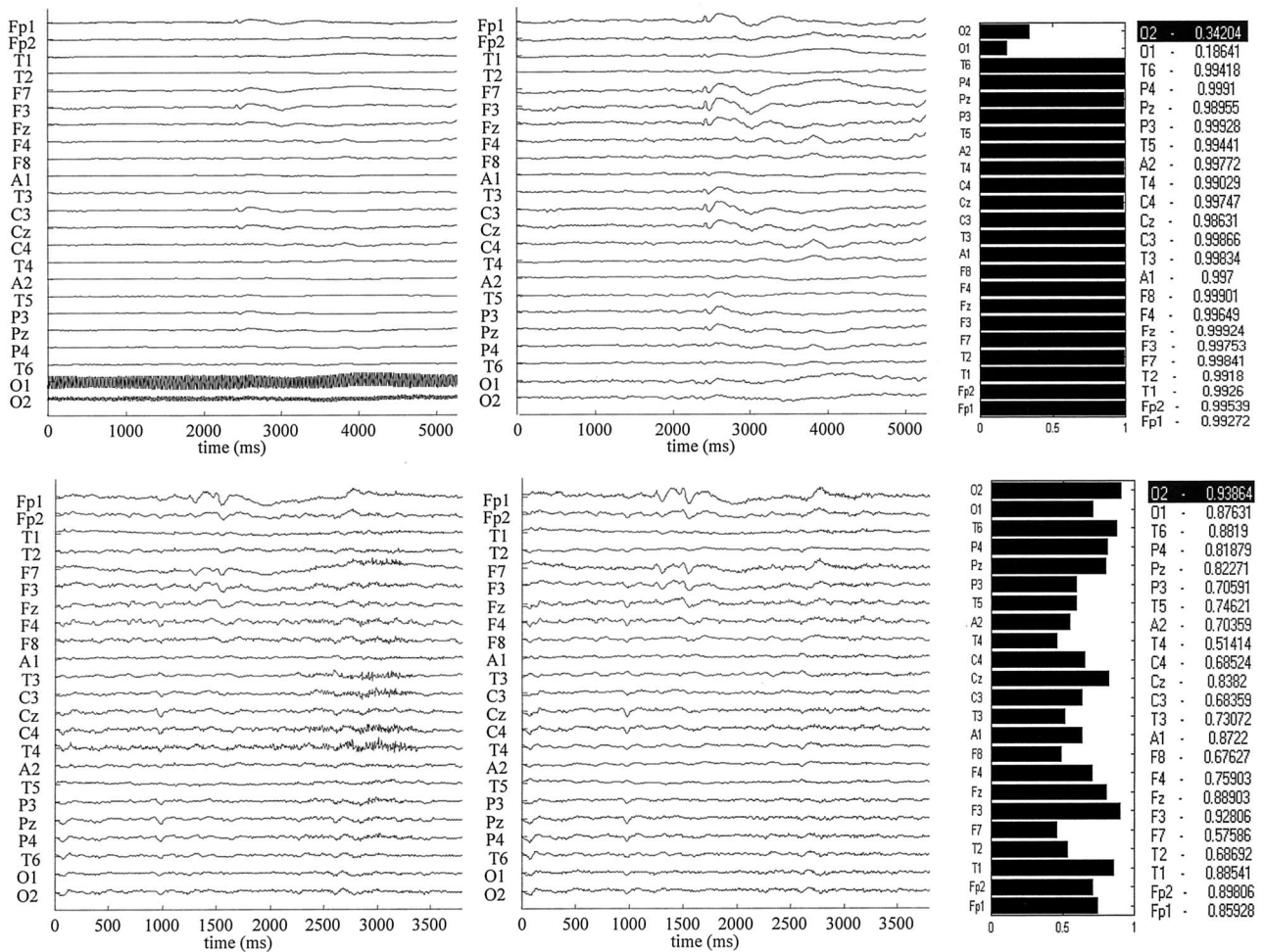


FIG. 5. (Above) Electrode artifact on O1 and O2. After removing its component, the correlation between the EEGs is very high in all channels except in O1 and O2. (Below) Removal of the muscle artifact. The correlation between the two EEGs is still low in T4, T3, F7, F8; it is higher in the electrodes far from the lateral areas.

removing artifacts on EEG were focused on these artifacts (Jung et al., 2000; Vigario, 1997). The chewing or other electrode artifacts were more difficult to remove because the selection of the components involved was not so direct, but in all cases, a significant effect was obtained. We used the original Harmonie files in each sample and the EEG format in Matlab. The first one has a higher quality for visualization, but the results were clear enough in both formats for the two reviewers.

The correlation coefficient calculated between the initial and the “cleaned” signal allowed us to quantify the changes induced in each of the channels by removing the desired component, thus giving an objective measure of the distortion generated by the suppression of the artifact. Also, the presence of spikes in the recordings was important for supporting the fact that removing the artifact did not distort the rest of the EEG. This result was

confirmed by the correlations and the isopotential maps. Independent component analysis proved to be highly efficient in removing the artifact and keeping a very high correlation for the spikes intermixed in the same channel where the artifact was present. Independent component analysis was able to remove an artifact without changing other elements of the same channel at approximately the same frequency (eye movement and spike-and-wave in Fp1 in Fig. 1, for example). Digital filters may alter the morphology of the EEG significantly, as they eliminate both brain signal and artifact in the frequency range selected. This is the reason why filtered EEG often does not correspond exactly with real brain activity. The correlation coefficient has demonstrated objectively that ICA has this important advantage over digital filters. Previous studies had shown graphically how ICA cleaned some artifacts on

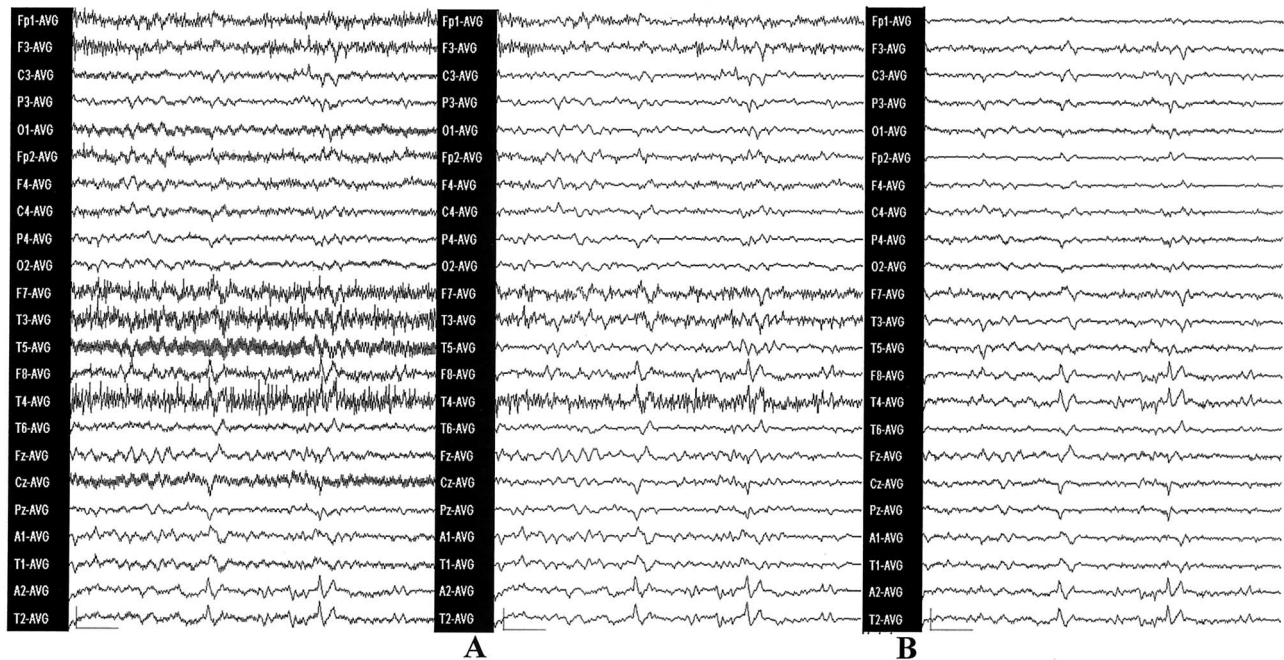


FIG. 6. (Left) Original sample of EEG with muscle artifact in several channels. The EEG in the center corresponds to the EEG using standard digital filters (line 50-Hz, HF: 35 Hz [IIR, order 4], LF: 0.3 Hz [IIR, order 2]). On the right, the EEG after the removal of the components responsible for the muscle artifact is shown (it is presented without any additional digital filter). The spikes are located in T2, A2, F8 and T4. Scale: 100 μ V, 1 s.

the EEG. Our study, on the other hand, has demonstrated the utility of ICA not only visually but also in a quantitative way, using the correlation, the analysis of the frequency spectra, and the isopotential maps of the spikes.

Independent component analysis also has some limitations. The lengths of the sample, the exact number of components, and the selection of the appropriate component (visual for the moment), are problems that may make its use and interpretation difficult. However, the possibility of removing artifacts without modifying the rest of the EEG activity can open new frontiers in electroencephalography. We have applied it to the study of interictal activity. The demonstration of the usefulness in removing artifacts in interictal samples is the first step before trying new goals. In our study, ICA did not change the original reading of the EEG; actually our goal was to identify both the original spike and the artifact to prove that ICA removes some components without modifying the rest of the EEG. This demonstration was necessary because without this fact, we could not be sure of the real action of ICA and, as a consequence, other more complicated analyses or functions would always be unreliable or dubious. Now, we can be sure that ICA can remove real components without modifying others in standard EEGs. Its use in ictal recordings would be even

more interesting. At the beginning of a seizure, the patient frequently presents oral, ocular, or other automatisms that may obscure the EEG changes, making the determination and localization of the EEG ictal onset difficult. The removal of these artifacts will improve the accuracy of the ictal EEG findings. Other important future uses might be the research of the nature of the normal findings in the EEG or the research of epileptiform activity (McKeown et al., 1998, 1999; Makeig et al., 2000). New capabilities of ICA must also be implemented; the use in long recordings and the on-line analyses and filtering would be very helpful and may provoke a clear advance in the EEG technology.

REFERENCES

- Arfanakis K, Cordes D, Haughton VM, Moritz CH, Quigley MA, Meyerand ME. Combining independent component analysis and correlation analysis to probe interregional connectivity in fMRI task activation datasets. *Magn Reson Imaging* 2000; 18: 921–30.
- Bartlett MS, Makeig S, Bell AJ, Jung TP, Sejnowski TJ. Independent component analysis of EEG data. *Abstr Soc Neurosci* 1995;21:437.
- Brown GD, Yamada S, Sejnowski TJ. Independent component analysis at the neural cocktail party. *Trends Neurosci* 2001;24:54–63.
- Calhoun VD, Adali T, Pearlson GD, Pekar JJ. Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Hum Brain Mapp* 2001;13: 43–53.

- Cardoso J-F. Blind signal separation: statistical principles. *Proceedings of the IEEE* 1998;9:2009—25.
- De Lathauwer L, De-Moor B, Vandewalle J. Fetal electrocardiogram extraction by blind source subspace separation. *IEEE Trans Biomed Eng* 2000; 47: 567—72.
- Hoyer PO, Hyvarinen A. Independent component analysis applied to feature extraction from colour and stereo images. *Network* 2000; 11:191—210
- Hyvarinen A, Hoyer PO, Inki M. Topographic independent component analysis. *Neural Comput* 2001;13:1527—58.
- Hyvarinen A, Oja E. Independent component analysis: algorithms and applications. *Neural Netw* 2000;13:411—30.
- Ikeda S, Toyama K. Independent component analysis for noisy data—MEG data analysis. *Neural Netw*. 2000;13:1063—74.
- Ille N, Berg P, Scherg M. Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *J Clin Neurophysiol* 2002;19:113—124.
- Jung TP, Humphries C, Lee TW, Makeig S, McKeown MJ, Iragui V, Sejnowski TJ. Removing electroencephalographic artifacts: comparison between ICA and PCA. In: Constantinides T, Kung SY, Niranjani M, Wilson E, eds. *Neural Networks for Signal Processing VIII: Proceedings of the 1998 IEEE Workshop*. New York: IEEE Press, 1998: 63—72.
- Jung TP, Makeig S, Humphries C, Lee TW, McKeown MF, Iragui V, Sejnowski TJ. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 2000a;37:163—78.
- Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol* 2000b;111:1745—58.
- Karhunen J, Malaroui S, Ilmoniemi M. Local linear independent component analysis based on clustering. *Int J Neural Syst* 2000; 10:439—51.
- Kobayashi K, Akiyama T, Nakahori T, Yoshikawa S, Gotman J. Systematic source estimation of spikes by a combination of independent component analysis and RAP-MUSIC. I: Principles and simulation study. *Clin Neurophysiol* 2002;113:713—24.
- Kobayashi K, Akiyama T, Nakahori T, Yoshikawa S, Gotman J. Systematic source estimation of spikes by a combination of independent component analysis and RAP-MUSIC. II: Preliminary clinical application. *Clin Neurophysiol* 2002;113:725—34.
- Lee JS, Lee DS, Ahn JY, Choi YM, Jin DK, Park GS, Choi IS, Sohn MS, Shin EK. Blind separation of cardiac components and extraction of input function from H(2)(15)O dynamic myocardial PET using independent component analysis. *J Nucl Med* 2001;42: 938—43.
- Maeda S, Inagaki S, Kawaguchi H, Song WJ. Separation of signal and noise from in vivo optical recording in Guinea pigs using independent component analysis. *Neurosci Lett* 2001;302:137—40.
- Makeig S, Enghoff S, Jung TP, Sejnowski TJ. A natural basis for efficient brain-actuated control. *IEEE Trans Rehabil Eng* 2000;8: 208—11.
- McKeown MJ, Humphries C, Achermann P, Borbely AA, Sejnowski TJ. A new method for detecting state changes in the EEG: exploratory application to sleep data. *J Sleep Res* 1998;7 (suppl 1): 48—56.
- McKeown MJ, Humphries C, Iragui V, Sejnowski TJ. Spatially fixed patterns account for the spike and wave features in absence seizures. *Brain Topogr* 1999;12:107—16.
- Moritz CH, Haughton VM, Cordes D, Quigley M, Meyerand ME. Whole-brain functional MR imaging activation from a finger-tapping task examined with independent component analysis. *AJNR Am J Neuroradiol* 2000;21:1629—35.
- Sato W, Kochiyama T, Yoshikawa S, Matsumura M. Emotional expression boosts early visual processing of the face: ERP recording and its decomposition by independent component analysis. *Neuroreport* 2001;12:709—14.
- Stone JV. Blind source separation using temporal predictability. *Neural Comput* 2001;13:1559—74.
- Suzuki K, Kiryu T, Nakada T. Fast and precise independent component analysis for high field fMRI time series tailored using prior information on spatiotemporal structure. *Hum Brain Mapp* 2002; 15:54—66.
- Tong S, Bezerianos A, Paul J, Zhu Y, Thakor N. Removal of ECG interference from the EEG recordings in small animals using independent component analysis. *J Neurosci Methods* 2001;108: 11—17.
- Vigario R, Sarela J, Jousmaki V, Hamalainen M, Oja E. Independent component approach to the analysis of EEG and MEG recordings. *IEEE Trans Biomed Eng* 2000;47:589—93.
- Vigario RN. Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalogr Clin Neurophysiol* 1997; 103:395—404.