Jefferson.

Thomas Jefferson University
Jefferson Digital Commons

Department of Neurology Faculty Papers | Department of Neurology

4-4-2017

AR2, a novel automatic muscle artifact reduction software method for ictal EEG interpretation: Validation and comparison of performance with commercially available software.

Shennan Aibel Weiss
University of California Los Angeles

Ali Akbar Asadi-Pooya
Thomas Jefferson University, Aliakbar.Asadi-Pooya@jefferson.edu

Sitaram Vangala
University of California Los Angeles

Stephanie Moy
University of California Los Angeles

Dale H. Wyeth
Thomas Jefferson University, dale.wyeth@jefferson.edu

See next page for additional authors

Let us know how access to this document benefits you

Follow this and additional works at: https://jdc.jefferson.edu/neurologyfp

Part of the Neurology Commons

Recommended Citation

Weiss, Shennan Aibel; Asadi-Pooya, Ali Akbar; Vangala, Sitaram; Moy, Stephanie; Wyeth, Dale H.; Orosz, Iren; Gibbs, Michael; Schrader, Lara; Lerner, Jason; Cheng, Christopher K; Chang, Edward; Rajaraman, Rajsekar; Keselman, Inna; Churchman, Perdro; Bower-Baca, Christine; Numis, Adam L; Ho, Michael G; Rao, Lekha; Bhat, Annapoorna; Suski, Joanna; Asadollahi, Marjan; Ambrose, Timothy; Fernandez, Andres; Nei, Maromi; Skidmore, Christopher T.; Mintzer, Scott; Eliashiv, Dawn S.; Mathern, Gary W; Nuwer, Marc R; Sperling, Michael R.; Engel, Jerome; and Stern, John M, "AR2, a novel automatic muscle artifact reduction software method for ictal EEG interpretation: Validation and comparison of performance with commercially available software." (2017).

Authors
Shennan Aibel Weiss, Ali Akbar Asadi-Pooya, Sitaram Vangala, Stephanie Moy, Dale H. Wyeth, Iren Orosz, Michael Gibbs, Lara Schrader, Jason Lerner, Christopher K Cheng, Edward Chang, Rajsekar Rajaraman, Inna Keselman, Perdro Churchman, Christine Bower-Baca, Adam L Numis, Michael G Ho, Lekha Rao, Annapoorna Bhat, Joanna Suski, Marjan Asadollahi, Timothy Ambrose, Andres Fernandez, Maromi Nei, Christopher T. Skidmore, Scott Mintzer, Dawn S. Eliashiv, Gary W Mathern, Marc R Nuwer, Michael R. Sperling, Jerome Engel, and John M Stern

This article is available at Jefferson Digital Commons: https://jdc.jefferson.edu/neurologyfp/130

Check for updates

SOFTWARE TOOL ARTICLE

REVISED AR2, a novel automatic muscle artifact reduction software method for ictal EEG interpretation: Validation and comparison of performance with commercially available software [version 2; referees: 2 approved]

Previously titled: AR2, a novel automatic artifact reduction software method for ictal EEG interpretation: Validation and comparison of performance with commercially available software

Shennan Aibel Weiss [1,2], Ali A Asadi-Pooya[2], Sitaram Vangala[3], Stephanie Moy[1], Dale H Wyeth[2], Iren Orosz[4], Michael Gibbs[1], Lara Schrader[1], Jason Lerner[1], Christopher K Cheng[1], Edward Chang[1], Rajsekar Rajaraman[1], Inna Keselman[1], Perdro Churchman[1], Christine Bower-Baca[1], Adam L Numis[1], Michael G Ho[1], Lekha Rao[1], Annapoorna Bhat[2], Joanna Suski[2], Marjan Asadollahi[2], Timothy Ambrose[2], Andres Fernandez[2], Maromi Nei[2], Christopher Skidmore[2], Scott Mintzer[2], Dawn S Eliashiv[2], Gary W Mathern[5], Marc R Nuwer[1], Michael Sperling[2], Jerome Engel Jr[1], John M Stern[1]

[1]Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, USA
[2]Department of Neurology, Jefferson Comprehensive Epilepsy Center, Thomas Jefferson University, Philadelphia, USA
[3]Department of Medicine, Statistics Core, University of California Los Angeles, Los Angeles, USA
[4]Department of Radiology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, USA
[5]Departments of Neurosurgery, Psychiatry, and Biobehavioral Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, USA

v2 **First published:** 10 Jan 2017, **6**:30 (doi: 10.12688/f1000research.10569.1)
Latest published: 04 Apr 2017, **6**:30 (doi: 10.12688/f1000research.10569.2)

Abstract

Objective: To develop a novel software method (AR2) for reducing muscle contamination of ictal scalp electroencephalogram (EEG), and validate this method on the basis of its performance in comparison to a commercially available software method (AR1) to accurately depict seizure-onset location.
Methods: A blinded investigation used 23 EEG recordings of seizures from 8 patients. Each recording was uninterpretable with digital filtering because of muscle artifact and processed using AR1 and AR2 and reviewed by 26 EEG specialists. EEG readers assessed seizure-onset time, lateralization, and region, and specified confidence for each determination. The two methods were validated on the basis of the number of readers able to render assignments, confidence, the intra-class correlation (ICC), and agreement with other clinical findings.
Results: Among the 23 seizures, two-thirds of the readers were able to delineate seizure-onset time in 10 of 23 using AR1, and 15 of 23 using AR2

Open Peer Review

Referee Status: ✓ ✓

	Invited Referees 1	2
REVISED version 2 published 04 Apr 2017	✓ report	✓ report
version 1 published 10 Jan 2017	? report	? report

(p<0.01). Fewer readers could lateralize seizure-onset (p<0.05). The confidence measures of the assignments were low (probable-unlikely), but increased using AR2 (p<0.05). The ICC for identifying the time of seizure-onset was 0.15 (95% confidence interval (CI), 0.11-0.18) using AR1 and 0.26 (95% CI 0.21-0.30) using AR2. The EEG interpretations were often consistent with behavioral, neurophysiological, and neuro-radiological findings, with left sided assignments correct in 95.9% (CI 85.7-98.9%, n=4) of cases using AR2, and 91.9% (77.0-97.5%) (n=4) of cases using AR1.
Conclusions: EEG artifact reduction methods for localizing seizure-onset does not result in high rates of interpretability, reader confidence, and inter-reader agreement. However, the assignments by groups of readers are often congruent with other clinical data. Utilization of the AR2 software method may improve the validity of ictal EEG artifact reduction.

1 **Patrícia Figueiredo**, Instituto Superior Técnico University of Lisbon Portugal, **Rodolfo Abreu**, Instituto Superior Técnico University of Lisbon Portugal
Rodolfo Abreu, Instituto Superior Técnico University of Lisbon Portugal

2 **David M. Groppe**, The Krembil Neuroscience Centre Canada

Discuss this article

Comments (0)

This article is included in the Neuroinformatics channel.

Corresponding author: Shennan Aibel Weiss (shennan.weiss@jefferson.edu)

How to cite this article: Weiss SA, Asadi-Pooya AA, Vangala S *et al.* **AR2, a novel automatic muscle artifact reduction software method for ictal EEG interpretation: Validation and comparison of performance with commercially available software [version 2; referees: 2 approved]** *F1000Research* 2017, **6**:30 (doi: 10.12688/f1000research.10569.2)

Copyright: © 2017 Weiss SA *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Grant information: Dr. Weiss was supported by an Epilepsy Foundation Award Research and Training Fellowship for Clinicians.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: No competing interests were declared.

First published: 10 Jan 2017, **6**:30 (doi: 10.12688/f1000research.10569.1)

REVISED **Amendments from Version 1**

We have substantially revised the manuscript in order to address the concerns of the three reviewers. In an effort to more transparently convey effect size, we have revised our statistical approach by performing the student's paired t-test and providing the reader with the t-values. We also correct for multiple comparisons using the Holm-Bonferroni method.

In the methods section we provide greater detail regarding the AR2 methodology and also indicate that the experimental dataset was used to derive parameters which could have overestimated the efficacy of the approach. In the introduction and discussion, we offer improved explanations of the approach and results derived from an expanded body of literature.

We have slightly modified Figure 1, Figure 2, Figure 4, and Figure 5.

See referee reports

Introduction

The scalp electroencephalogram (EEG) is a critical diagnostic tool in the evaluation of seizures, but artifact from muscle contraction often limits its use because of the obscuring of the cerebrally generated potentials. This problem is present in 11% of ictal EEGs overall and up to 70% of frontal lobe seizures[1–3]. The inability, or lack of precision, to discern the seizure-onset zone from scalp EEG often necessitates additional testing, including (positron emission tomography) PET, magnetoencephalography, ictal Single-photon emission computed tomography (SPECT), and intracranial EEG[4]. Each of these tests adds undesired time and cost to the evaluation.

Digital filters are the common approach to maximizing the likelihood of identifying a seizure-onset zone from EEG with muscle artifact. This filtering reduces muscle artifact by attenuating all frequencies beyond a selected value[5], but it may impair the integrity of the EEG recording since brain-generated potentials may be in the same frequency band[6,7]. Recently, new technologies to reduce muscle artifact based on independent component analysis (ICA)[8–10] have become available. ICA derives spatial features that can remove artifacts that have static scalp topographies and time courses of activity that are distinct from that of EEG sources. ICA artifact correction is necessarily imperfect and will remove some neurogenic components of the EEG as well. However, the degree of EEG distortion may be negligible and ICA has proven effective at removing EMG and ocular artifacts from EEG data recorded from normal individuals in laboratory settings[11–20]. Prior studies have demonstrated that ICA-based methods improve the interpretation of artifact-laden ictal EEG recordings; in these studies researchers manually performed the ICA analysis prior to performing the EEG interpretation[15,16]. Automatic artifact reduction using ICA[17–19] has become commercially available and is included in the latest versions of popular EEG viewer software[20]. Ictal scalp EEG recordings present extraordinary challenges to ICA artifact reduction algorithms because the number of EMG artifact sources increases.

Despite the utilization of these software products by neurologists around the globe, the clinical benefit has not been established. It is also unknown if the new approaches introduce confounding artifacts that may lead to erroneous interpretations.

The goal of this study was to assess the validity of a commercially available EEG artifact reduction tool (AR1) that uses different montages and within electrode analysis to identify artefactual independent components[20], and compare its validity to a novel automatic artifact reduction tool (AR2), which was developed at the University of California Los Angeles on the basis of inter-reader agreement, confidence, and congruence with other clinical findings.

Methods

Implementation

The custom software algorithm involved importing EEG scalp recordings as European Data Format (EDF) files in Matlab 8.4 (Mathworks, Natick, MA). Prior to performing ICA to remove muscle artifact, the algorithm first identified epochs of the scalp EEG record contaminated by muscle artifact and determined the electrodes that were suspected of having high recording impedance during that epoch. The purpose of these calculations was to exclude these electrodes from the ICA calculations.

The imported EEG was band pass filtered (16–70 Hz) using a 500th order finite impulse response filter, i.e. FIR1 in referential montage. We then calculated the normalized instantaneous amplitude of the band-pass filtered signal using a Hilbert transform. This signal was smoothed using moving averaging, and the algorithm identified the longest epoch in which the time series remained greater than one standard deviation. We next calculated the normalized mutual information (MI)[21] adjacency matrix across all scalp electrode contacts during the (16–70 Hz) band-pass filtered artifact epoch of greatest duration and assigned each scalp EEG electrode a single MI value derived from the maximum pairwise MI values in the adjacency matrix. We then determined if this maximum mutual information value exceeded a threshold value defined by visual inspection of the scalp EEG used in the experimental dataset, and if that electrode should be included in subsequent artifact reduction processing. If the recording lacked an artifact epoch, or all channels were excluded, artifact reduction was applied to the referential recordings from all recording electrodes.

The high pass filtered (>16 Hz) scalp EEG was then separated into consecutive 120-second trials (24,000 data points) and each trial was processed using CUDAICA[22,23]. A 120 second trial length was chosen to optimize processing time. The purpose of the ICA was to separate the (>16 Hz) seizure activity, from the (>16 Hz) muscle artifact. The 16 Hz cut-off for the filter was chosen to isolate the vast majority of the muscle artifact. Independent components that explained an amount of variance above a particular threshold were excluded from the signal. The threshold was selected on the basis of the values of the raw and normalized mixing matrix (i.e. inverse weight matrix) calculated in each of the ICA iterations. We assumed that the last myogenic component and first neurogenic component can be differentiated on the basis of the inverse weight matrix, which provides the spatial distribution of each component, and identifying the independent component that account for the most variance with a focal spatial topography[17] defined on the basis of exceeding a normalized threshold of two standard deviations in at least one electrode of the inverse weight matrix. This threshold was chosen on the basis of visual inspection of the EEG in the experimental dataset and resulting independent components.

The pruned EEG calculated for each 120 second trial of EEG (i.e iteration of CUDAICA) was concatenated, and subsequently the entire raw ictal EEG was low pass filtered (<16 Hz) using a 500th order symmetric digital FIR filter, and the resulting low pass filtered EEG was reconstituted by addition of the waveforms with the high pass (>16 Hz) filtered EEG, following the exclusion of the independent components suspected to represent muscle artifact. The reconstituted and modified ictal EEG was exported from Matlab format to EDF for subsequent visual analysis.

Operation

All computations were carried out using compiled Matlab 8.4 custom scripts on a cluster of HP SL230s Gen 8 ES-2670 nodes with dual-eight-core 2.6 GHz Intel ES-2670 central processing units, 4 GB of memory per core, and NVIDIA Tesla graphics processing units. Minimal system requirements for operating AR2 include Matlab v8.4 or above, an Intel Xeon CPU, 2 GB of memory, a NVIDIA GPU, which is CUDA compatible, and CUDAICA. For scalp EEG files exported from Neuroworkbench (Nihon-Kohden, Irvine, CA, USA), executing the AR2 software method requires only inputting the file name of the EDF file of interest at the command line. For EDF files exported from other equipment manufacturers, full automation of the AR2 software method can be easily accomplished with simple modifications of the input parameters.

Patients and sample selection

We tested AR2 retrospectively using 23 seizures from eight adult patients with suspected focal-onset seizures undergoing evaluation at the UCLA Seizure Disorder Center. The patients and seizures were selected by S.A.W, whom was not a reviewer, from a review of consecutive clinical neurophysiology case conference presentations between January 1, 2014 and December 1, 2015 and based on case conference consensus that the ictal EEG records were uninterpretable due to muscle artifact contamination when reviewed with conventional digital filtering. For each of these patients, between 1–4 uninterpretable seizures were selected for inclusion in the study on the basis of a lack of ECG, electrode, and salt bridge artifact by S.A.W. Clinical data for each patient included seizure semiology, inter-ictal epileptiform abnormality, unobscured findings and radiological reports from MRI, PET, SPECT scans. The EEG and clinical records were deidentified and research informed consent was not required. This study was approved under UCLA IRB#15-001481. The video EEGs were acquired using a EEG-1200 amplifier (Nihon-Kohden, Irvine, CA) at a sampling rate of 200 Hz, low frequency cut-off 0.08 Hz. Electrodes were placed according to the 10–20 international system with the additional anterotemporal electrodes at T1/T2. The duration of the exported EEG recording included the entire seizure and a 2–3 minute peri-ictal epoch.

Muscle artifact removal

AR1 was the commercially available Persyst v12 artifact reduction software[20] (Persyst Development, San Diego, CA). The methods are proprietary. AR2 was developed by S.A.W and involved a two-step procedure consisting of a custom algorithm. EEG processed by AR2 was also interpreted using the Persyst v12 artifact reduction software.

Performance measures of AR1 and AR2

The AR1 and AR2 processed data were reviewed in Persyst v12 without video by 26 neurologists with a specialization in EEG, 20 of whom were board certified. The readers were blinded to which records received AR1 or AR2, and each reader reviewed the 46 seizures in random. Following review of each ictal record, the reader completed a multiple choice questionnaire (Supplementary File 1), which assessed ability to visualize seizure-onset (Y,N) lateralize seizure-onset (L,R,N), locate the region of ictal onset (anterior temporal, anterior frontal, mid-temporal, temporal-parietal-occipital, occipital, none), and self-identify confidence of interpretation on a 5 point scale [(5) entirely confident (4) somewhat sure (3) probable (2) not confident (1) unlikely i.e. slight probability] for each measure. When time of onset, laterality, or the seizure onset region was not assigned the confidence was taken as (0). Readers were not provided with a definition of seizure-onset.

EEG analysis

During the interpretation of the ictal EEG processed by AR1 or AR2, no restrictions were placed on the use of Persyst v12 built in EEG filters (low-pass, high-pass, band-pass), or changes to montage. A comment in each recording was used to demarcate the time prior to the clinical seizure but not the EEG onset. The assessment was not time limited.

Statistical analysis

Differences in EEG interpretation utilizing AR1 and AR2 were assessed using the paired student's t-test and the McNemar test on paired nominal data. The Bonferroni-Holm method was used to correct for multiple comparisons. Agreement across readers (Y,N,L,R), using either AR1 or AR2, was calculated using the inter-class correlation coefficient (ICC). For these outcomes, missing values were imputed to be in between non-missing values, and were analyzed using cumulative logit mixed effects models, which capture this ordering in the values and accounts for the clustering of readings into patients, and seizures within patients. Agreement across readers for onset region was calculated using a Fleiss kappa and treating the missing values as a category of response. Errors are given as standard error of the mean (s.e.m), unless otherwise specified.

Results

Implementation of the AR2 method

We applied the AR2 method developed at UCLA to the 23 seizures in the dataset. The method was automatic and unsupervised and separated the high-pass filtered (> 16 Hz) scalp EEG recordings into putative neurogenic and myogenic components (Figure 1). After pruning the putative myogenic components, the putative neurogenic components were reconstituted with the low-pass filtered (< 16 Hz) scalp EEG (Figure 2). The AR2 and AR1 processed scalp EEG recordings were subsequently inspected by the 26 specialists (Figure 3).

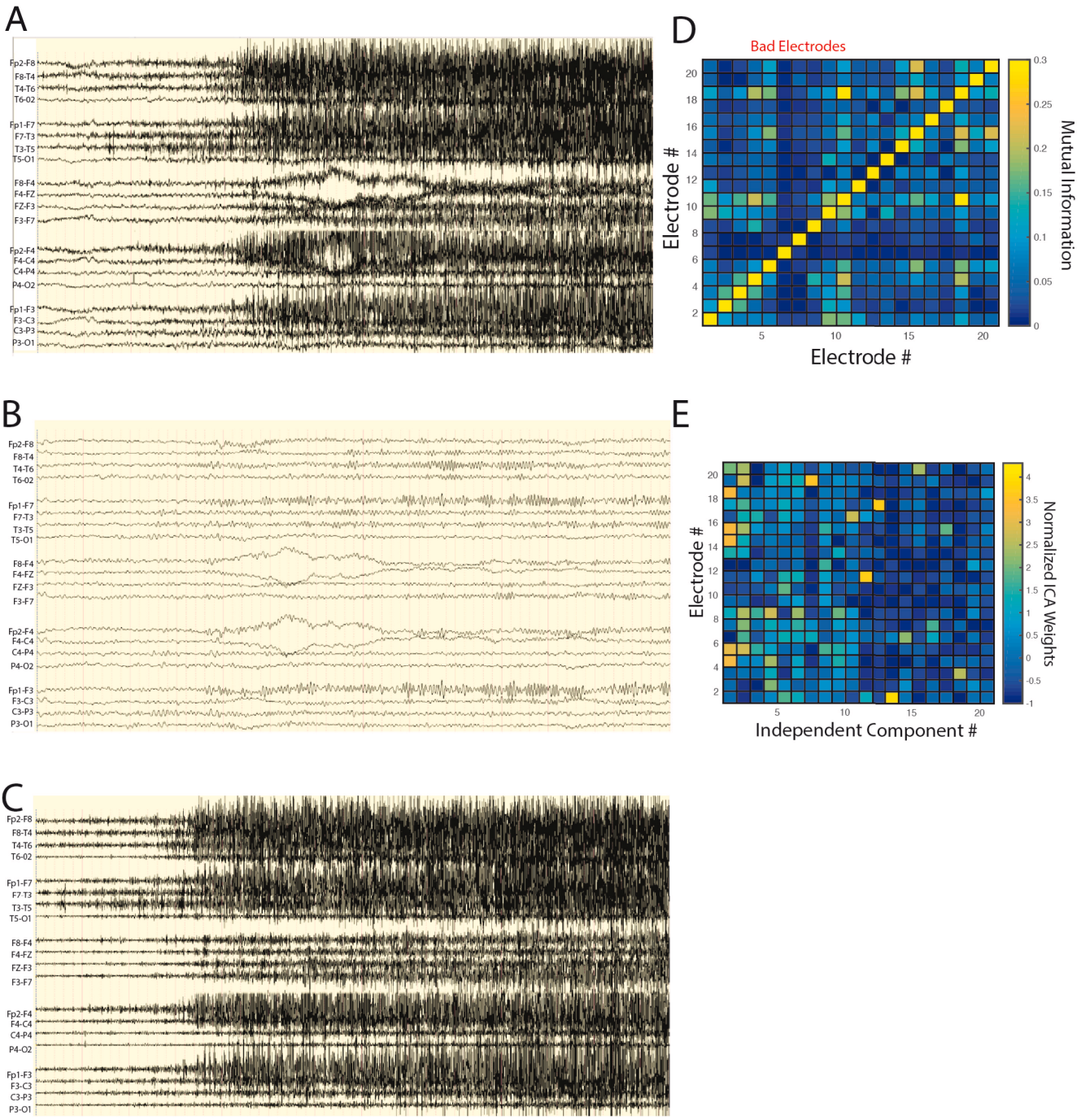

Figure 1. The AR2 method automatically separates independent components containing myogenic from neurogenic potentials. The AR2 method automatically separates independent components containing myogenic from neurogenic potentials in the beta and gamma band on the basis of spatial topography and explained variance. **A.** Unprocessed scalp ictal EEG recording that was deemed uninterpretable. **B.** The same epoch after applying a low pass (<16 Hz) filter demonstrating a lack of a convincing ictal rhythm. **C.** The ictal epoch after applying a high pass (> 16 Hz) filter demonstrating dense muscle artifact. **D.** An example of a mutual information adjacency matrix calculated during an epoch of artifact in the high pass (> 16 Hz) filtered scalp EEG recording. Three scalp electrode recordings exhibited relatively low mutual information with all other electrodes and were designated poor quality and excluded from further processing to optimize INFO-MAX ICA based artifact reduction. **E.** The normalized inverse weight matrix of all independent components across scalp electrode recordings for the seizure in panel A. Independent components 1-13 exhibited strong focality and were designated as containing myogenic potentials, while independent components 14 and above were designated neurogenic.

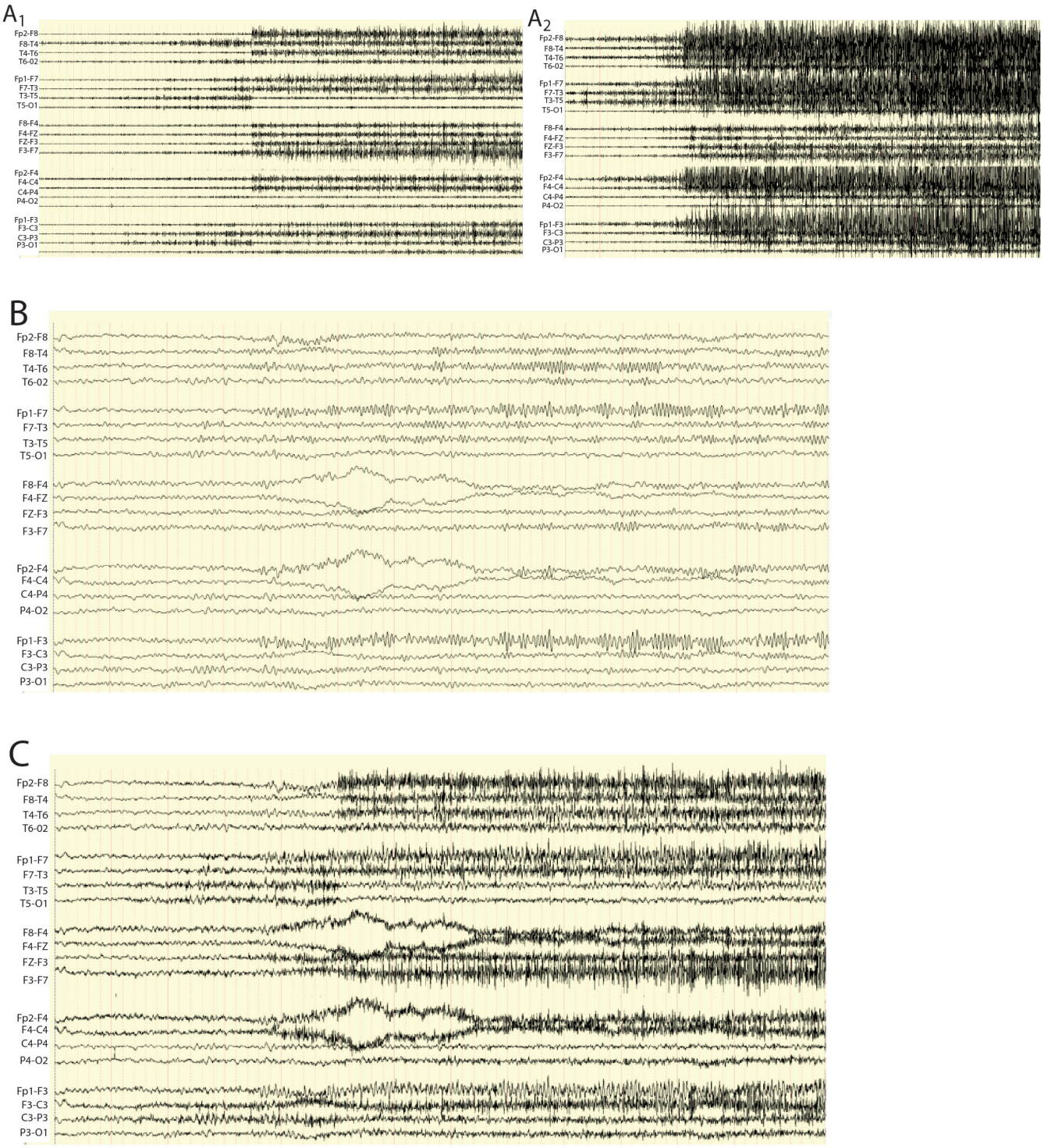

Figure 2. Ictal onset is revealed with reconstitution of the low pass (<16 Hz) ictal scalp EEG with the high pass (>16 Hz) neurogenic independent components. Reconstitution of the low pass (<16 Hz) ictal scalp EEG with the high pass (>16 Hz) neurogenic independent components reveals a clear ictal onset. **A.** The tentative neurogenic independent components (**A1**) and myogenic independent components (**A2**) derived from INFOMAX ICA processing of the high pass (> 16 Hz) filtered ictal scalp EEG recording are shown. The largest amplitude activity in the neurogenic components are evident frontally and in the left hemisphere. **B.** The low pass filtered ictal scalp EEG suggests a possible left frontal onset but a convincing ictal rhythm is lacking. **C.** Reconstitution of the low pass EEG with the neurogenic high pass (> 16 Hz) independent components results in an ictal EEG that demonstrates a more convincing left frontal onset consisting of beta-gamma oscillations with some clear phase reversals in F3 and F7.

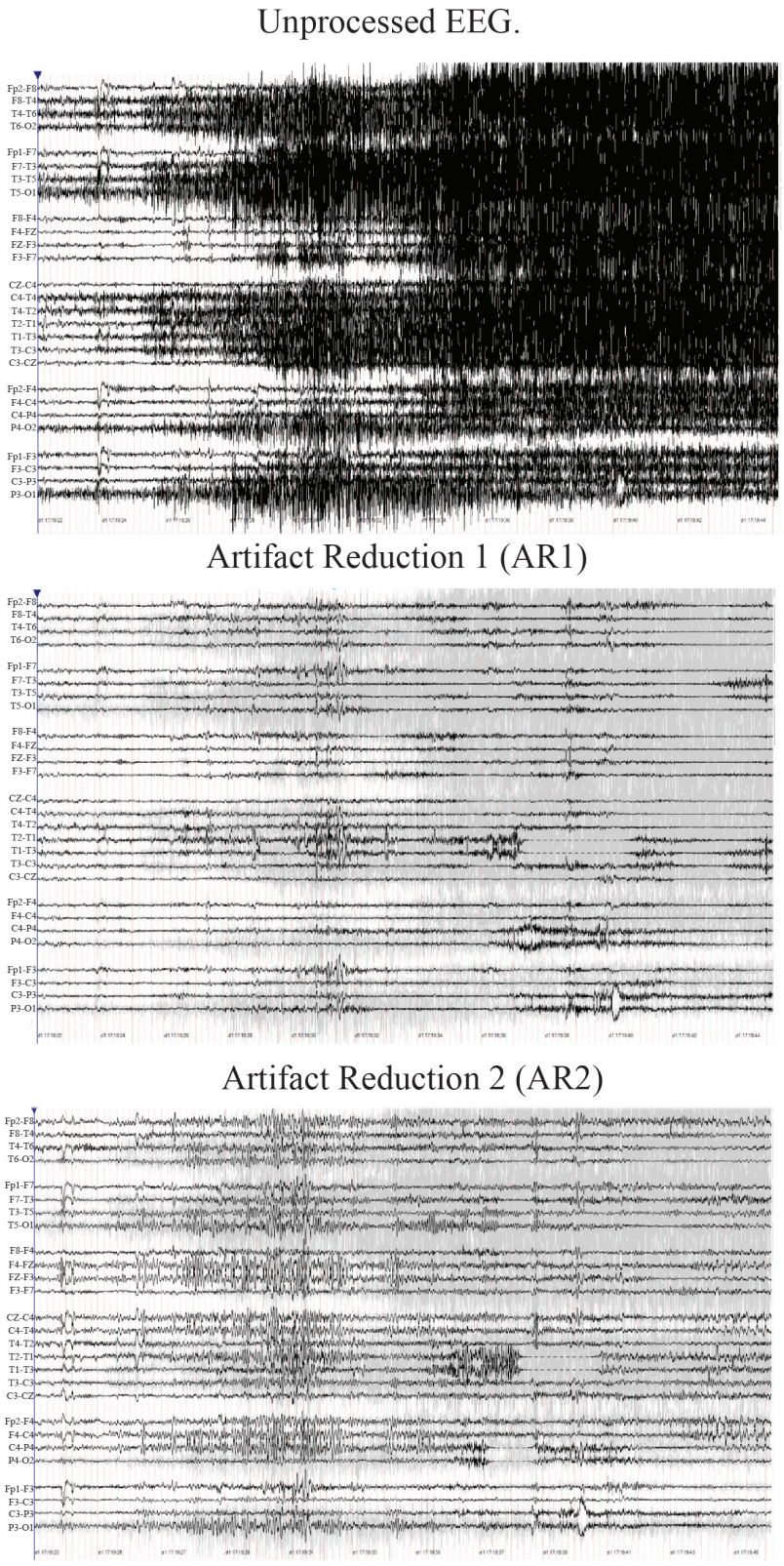

Figure 3. A comparison of the results of artifact reduction methodologies. Ictal scalp EEG recording from seizure 18 prior to artifact reduction processing (top), after processing with artifact reduction methodology 1 (AR1, middle), and after processing with artifact reduction methology 2 (AR2, bottom). Only processing with AR2 reveals a right hemispheric onset followed by clear spread to right frontal regions.

Identifying time of seizure-onset

Across the 23 seizures considered previously uninterpretable with digital filtering (Table 1) two-thirds of the readers were able to delineate the time of seizure-onset for 10 of the 23 using AR1, and 15 of the 23 using AR2 (Figure 4A, n=23, paired t-test $p<0.01$, $t=3.83$). Across the 23 seizures, the mean confidence measure for the determination of seizure-onset was 2.68 +/- 0.19 (probable-not confident) when AR2 was utilized and 2.19 +/- 0.18 (not confident) with AR1 (Figure 5A, d.f.=22, paired t-test, $p<0.01$, $t=4.33$). The inter-class coefficient (ICC) was 0.26 (95% Confidence Interval (CI) 0.21-0.30) with AR2, and 0.15 (95% CI 0.11-0.18) with AR1 (cumulative logit mixed effects models, $p=0.333$).

Lateralizing and localizing seizure-onset

Compared with identifying the time of seizure-onset, fewer readers could lateralize seizure-onset after either AR1 or AR2 (Figure 4B, d.f.=22, paired t-test, $p<0.01$, $t=8.08$ AR1, $t=8.56$ AR2). However, more readers were able to lateralize seizure-onset using AR2 compared to AR1 (Figure 4B, d.f.=22, paired t-test, $p<0.01$ $t=3.30$) and readers were more confident with AR2, although both methods did not produce high levels of confidence. The mean confidence measure for seizure-onset lateralization was 1.87+/- 0.198 (not confident-unlikely) for AR2 and 1.54+/- 0.176 (not confident-unlikely) for AR1 (Figure 5B, d.f.=22, paired t-test, $p<0.01$, $t=2.85$). The ICC was equivalent (cumulative logit mixed effects models, $p=0.501$) for AR1 (ICC=0.33 95% CI 0.30-0.37) and AR2 (ICC=0.28 95% CI 0.25-0.31). For localizing the region of seizure-onset reader confidence (Figure 5C), and agreement was very low (Figure 6, AR1 Fleiss' kappa = 0.1199, 95% CI = 0.116-0.124, AR2 Fleiss' kappa = 0.121, 95% CI =0.118-0.125). For one of the seizures, the laterality assignments were different when AR2 was used as compared to AR1 (Figure 4B, McNemar $p<0.05$).

Comparison of seizure-onset lateralization assignments with other clinical findings

We identified the patients with at least two consistent clinical findings that lateralized the suspected seizure-onset zone (SOZ). Compared to AR1, more readers were able to render seizure-onset laterality assignments using AR2, and these assignments were more often congruent with other clinical data (Table 2). These clinical findings included seizure semiology, onset of seizures without EEG obscuration, structural MRI, PET, or SPECT findings. If any of the clinical findings were contradictory with respects to the laterality of the suspected SOZ, the SOZ was designated unknown. Overall, 4 patients (#1,4,5,6) had clinical findings that supported a left-hemispheric SOZ, and 1 patient (#7) had clinical findings that supported a right-hemispheric SOZ (Table S1). Among the 5 patients with clinical seizure onset lateralization based on independent data, if the reader lateralized the seizure-onset to the left using AR2 they were correct in 95.9% (95% CI 85.7-98.9%) of cases, but using AR1 they were correct in 91.9% (95% CI 77.0-97.5%) of cases (Table 3, $p<0.0607$).

Discussion

In this study, we present a new artifact reduction software, AR2, and its application compared with a commercially available tool, AR1. 26 neurologists used the two methods to interpret 23 ictal EEG recordings that were uninterpretable due to muscle artifact when reviewed with conventional filtering. The major findings from this study include: 1) the utilization of artifact reduction software results in non-uniform interpretation of ictal EEG, with many readers not able to render assignments; 2) when readers did render seizure-onset laterality assignments it often agreed with other clinical findings; 3) although the study size was small, the AR2 software method increased the number of readers that rendered assignments, and reader confidence suggesting it aids in diagnosis.

Both AR1 and AR2 are digital signal processing software tools[8,15,20] that may confound accurate ictal EEG interpretation by altering the appearance of the EEG. Digital filtering also can mislead[5]. One concern about AR1 and AR2 relates to the uncertainty that myogenic activity was fully removed, and neurogenic components were unaffected during waveform alteration. Specifically, the readers were not confident in their interpretations, and the determination of seizure lateralization sometimes differed between the AR1 or AR2 methods. As such, the artifact reduction methods may introduce false positive findings. This demonstrates the limits of EEG artifact reduction approaches and puts the advantages into perspective.

The reliability of localization by ictal scalp EEG in the absence of artifact is between 65–75% for lateralization[24]. Neurologists disagree more on the interpretation of ictal EEG processed with artifact reduction software, however the seizure-onset laterality assignments rendered by a quorum are often correct. Further refinement of this technology may successfully improve the efficiency of video-EEG monitoring and the utilization of epilepsy surgery; however, correlation with epilepsy resective surgery outcomes will be required for further validation.

With regard to AR2, the novel software method developed for this study, the slight improvement seen in ictal EEG interpretability after applying the method suggests that the algorithm can (1) sometimes produce signals that are, exclusively or mainly, EEG or EMG, and (2) identify which signals are of brain origin and which are contaminant. The effectiveness of AR2 could possibly be improved by utilizing autocorrelations to identify the myogenic independent components[17]

One explanation for AR2's ability to isolate myogenic from neurogenic activity may be related to the respective dipole generators of each. ICA produces independent components that may resemble single equivalent dipoles[14]. Presumably, networks of myocytes exhibit shorter distance connectivity then networks of neurons that produce beta and gamma oscillations, and thus the two generators can be distinguished on the basis of the focality[17] of the independent components topography.

Data and software availability

All software code for the new AR2 software developed by S.A.W. is openly and permanently available at https://github.com/shennanw/AR2.

Table 1. Clinical description of patients. Clinical description of patients and ictal EEG laterality and focus assignments with AR1 and AR2. Abbreviations (L:left, R:right), PET findings refer to hypometabolism, SPECT findings to hyperperfusion. The focus was determined on a majority basis across all the assignments made by the readers for a subject's seizure(s).

PatientAge Gender	Aura/Semiology	IEDs	Un-Obscured Seizure Onset Laterality	sMRI	PET/SPECT	Seizure Onset or Spread Laterality (AR2)	Seizure Onset or Spread Laterality (AR1)	AR2 focus	AR1 focus
#1 46M	Somato-sensory (warmth)/arousal from sleep, hyperkinetic,	none	left frontal ictal rhythm	nonlesional	normal PET, SPECT left insula	1. 14/21 L	1. 17/19 L	ant/mid temporal	ant temporal
#2 32M	Somato-sensory (discomfort)/right facial grimacing, right leg elevation, breath holding	none	none	nonlesional	PET right temporal, SPECT bilateral frontal lobes	2. 6/10 L 3. 6/8 L 4. 13/16 L	2. 6/7 R 3. 6/6 R 4. 7/7 R	ant. Frontal	mid. Temporal
#3 23M	Tachycardia/arousal from sleep, hyperkinetic, b/l dystonic posturing	none	none	nonlesional	normal PET	5. 8/10 R 6. 14/16R 7. 7/11 R	5. 6/11 L 6. 6/6 L 7. 6/6 R	frontal/mid temporal	mid temporal
#4 53M	Visual disturbance/ behavioral arrest, cursing, right arm dystonic posturing	L temporal	L anterior temporal	L MTS, L parietal encephalomalacia	PET L parietal	8. 18/19 L 9. 16/21L 10. 17/21 L	8. 14/17 L 9. 10/13 L 10. 21/23 L	ant/mid temporal	ant temporal
#5 20M	Vague/right head and eye version, right arm clonic movements,	L temporal	L temporal	L frontal polymicrogyria	normal PET	11. 11/18 R 12. 7/11 L 13. 9/10 L	11. 9/18 R 12. 10/13 L 13. 7/11 R	ant/mid temporal	ant/mid temporal
#6 27M	None/arousal from sleep, dyscognitive, right head and body version.	L frontal	L frontal	normal	PET L inferior frontal	14. 20/22 L 15. 21/24 L 16. 22/24 L	14. 25/25 L 15. 24/24 L 16. 24/25 L	ant frontal/ ant temporal	ant frontal/ ant temporal
#7 26F	None/nocturnal arousal or daytime events, hyperkinetic, right or left dystonic posturing	L and R temporal	None	Right middle cranial fossa arachnoid cyst	PET R parietal lobe	17. 21/23 R 18. 12/16 R 19. 21/23 R 20. 20/21 R	17. 20/23 R 18. 18/21 R 19. 14/16 R 20. 6/11 R	ant/mid temporal	ant/mid temporal
#8 19M	Lightheaded/loss of consciousness, right > left arm clonic movements, and posturing	L and R temporal	None	L mesial temporal CD, R>L gyrus rectus encephalomalacia	PET L>R temporal lobe	21. 12/16 R 22. 12/23 R 23. 23/26 L	21. 14/16 L 22. 23/24 L 23. 22/22 L	ant/mid temporal	ant/mid temporal

F1000Research 2017, 6:30 Last updated: 22 MAY 2017

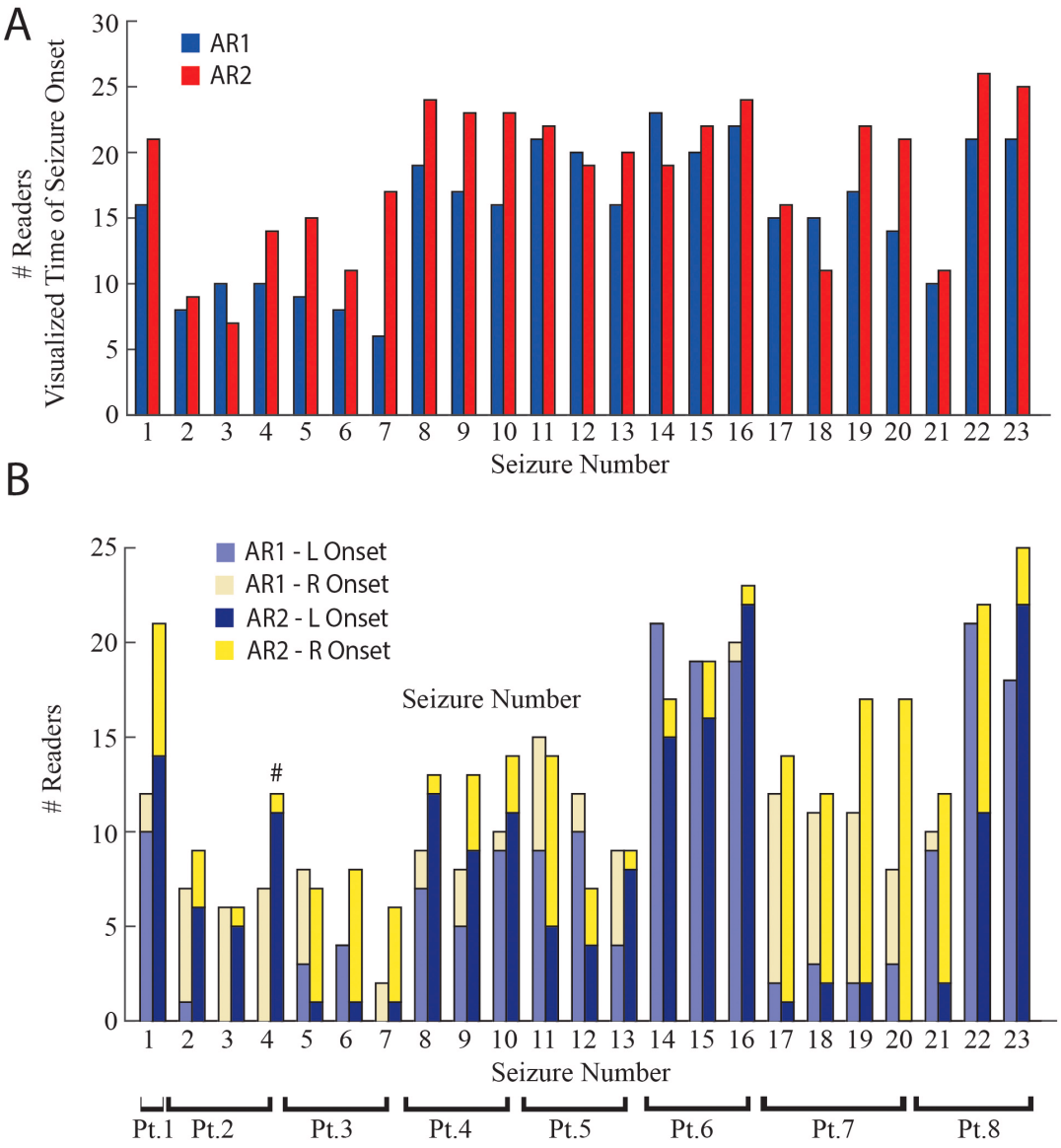

Figure 4. More readers could lateralize seizure onset utilizing AR2 as compared to AR1. More readers could visualize the time of seizure onset, and assign laterality to seizure onset utilizing AR2 as compared to AR1, and the assigned laterality of seizure onset sometimes differed between the two methods. **A**. Bar plot of the number of readers whom visualized the time of onset for each seizure utilizing AR1 (blue) or AR2 (red). Across seizures more readers visualized seizure onset utilizing AR2 compared with AR1 ($p<0.01$). Asterisks indicate statistically significant differences between the two methods in individual seizures (McNemar, $p<0.05$, Bonferroni-Holm corrected). **B**. Stacked bar plot of the number of readers selecting a left- or right-sided seizure onset utilizing AR1 (light blue, left; light yellow, right) or AR2 (dark blue, left; yellow, right). Across seizures more readers lateralized seizure onset utilizing AR2 compared with AR1 ($p<0.01$). Asterisks indicate statistically significant differences in individual seizures (McNemar, $p<0.05$, Bonferroni-Holm corrected), number sign indicates a significant change in the determination of laterality utilizing AR2 compared to AR1 (McNemar, $p<0.05$, Bonferroni-Holm corrected).

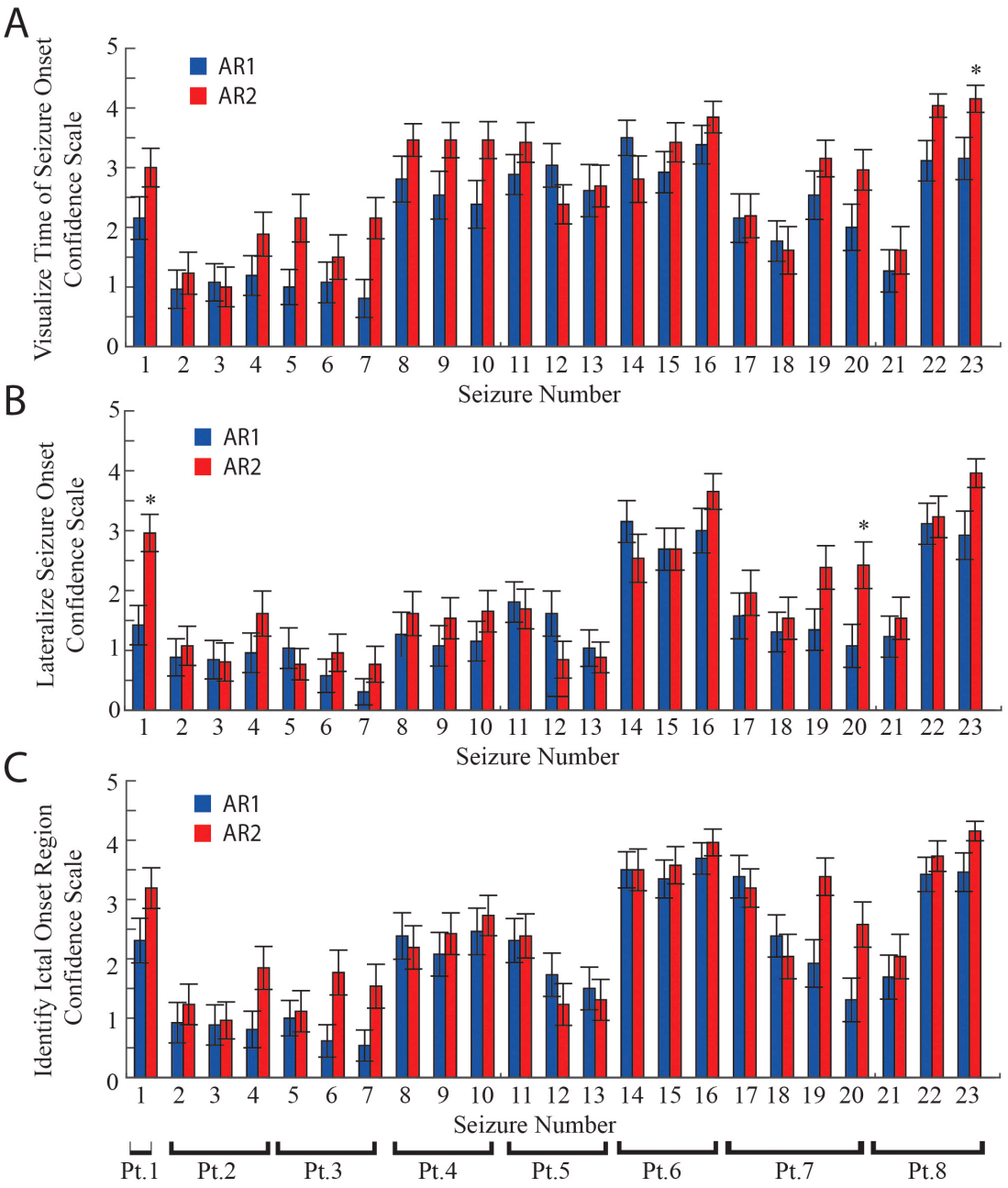

Figure 5. Confidence in the interpretation of ictal EEG onset improves with utilization of AR2 as compared to AR1. A. Bar plot of the mean confidence scale values for visualizing the time of seizure onset for the 23 seizures interpreted utilizing AR1 (blue), and AR2 (red). Across seizures, confidence scale values were greater when AR2 was utilized as compared with AR1 ($p<0.01$). Asterisks indicate differences in confidence values in individual seizures ($p<0.05$, Bonferroni-Holm corrected). Error bars are calculated as s.e.m. **B.** The respective mean confidence scale values for seizure onset lateralization. **C.** The respective mean confidence scale values for seizure focus localization. Across seizures, confidence scale values for lateralizing seizure onset, and identifying the seizure focus were greater when AR2 was utilized as compared with AR1 ($p<0.05$).

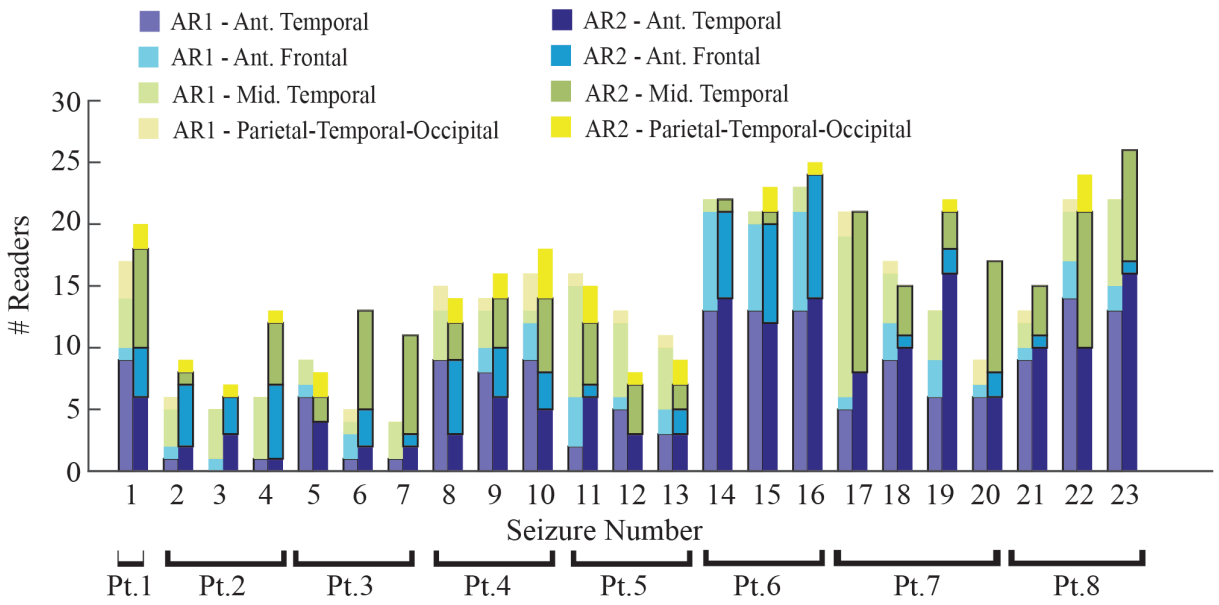

Figure 6. Differences in ictal onset region assignments using AR1 or AR2. Stacked bar plot of the ictal onset region assignments using either AR1 (lighter colors) or AR2 (darker colors) for all 23 seizures. Overall, across seizures, more readers were able to render an assignment using AR2 as compared to AR1 ($p<0.05$). Inter-reader agreement using for assigning the ictal onset region was marginal using either AR1 or AR2.

Table 2. Contingency table of agreement between assigned seizure onset laterality and other clinical findings. Contingency table of the agreement between seizure-onset laterality using AR1 (left), and AR2 (right) and the laterality of seizure-onset assigned on the basis of other clinical data for all the study patients and seizures. Note that clinical seizure-onset lateralization was not available for all patients, and when readers rendered a laterality decision that matched the laterality based on other clinical data, the assignments "agreed".

		AR1		
		EEG seizure-onset lateralization		
		Y	N	
Clinical seizure-onset lateralization	Y	Agree 145	Disagree 32	187
	N	83		151

		AR2		
		EEG seizure-onset lateralization		
		Y	N	
Clinical seizure-onset lateralization	Y	Agree 171	Disagree 39	154
	N	107		127

Archived source code as at time of publication: doi, 10.5281/zenodo.229893[21]

License: GNU Public License 3.

The raw scalp ictal EEG files that were analyzed in this study using AR2, as well as the scalp ictal EEG files following processing using AR2 are available from Zenodo[25]: **Dataset 1. Validity of two automatic artifact reduction software methods in ictal EEG interpretation.** Doi, 10.5281/zenodo.221095[22] (https://www.zenodo.org/record/221095#.WF63m7YrLdR)

The raw data used for the comparative assessments are available from Zenodo[26]: **Dataset 2. Validity of two automatic**

Table 3. Agreement between seizure-onset laterality and other clinical findings. Agreement between seizure-onset laterality assignments using either AR1 or AR2 and the suspected laterality of the SOZ assigned on the basis of other clinical data. Parentheses indicate the 95% confidence interval. "n" refers to the number of subjects.

Artifact Reduction Method	Reader Assignment of Seizure-Onset Laterality	Percentage of reader assignments in concordance with SOZ laterality defined by other clinical criteria.
AR1	Right	59.3 (28.5-84.2) (n=1)
	Unknown	66.8 (38.1-86.9) (n=3)
	Left	91.9 (77.0-97.5) (n=4)
AR2	Right	61.8 (31.3-85.1) (n=1)
	Unknown	71.4 (42.8-89.3) (n=3)
	Left	95.9 (85.7-98.9) (n=4)

artifact reduction software methods in ictal EEG interpretation. Doi. 10.5281/zenodo.223329 (https://zenodo.org/record/223329#.WHN-HLYrLdQ)

Author contributions

S.A.W designed the study, analyzed the data, drafted and revised the manuscript, A.A.P analyzed the data, and revised the manuscript, S.V analyzed the data, and revised the manuscript, S.M revised the manuscript, D.H.W revised the manuscript, I.O analyzed the data, and revised the manuscript, M.G analyzed the data, and revised the manuscript, L.S analyzed the data, and revised the manuscript, J.L analyzed the data, and revised the manuscript, C.K.C analyzed the data, and revised the manuscript, E.C analyzed the data, and revised the manuscript, R.R analyzed the data, and revised the manuscript, I.K analyzed the data, and revised the manuscript, P.C analyzed the data, and revised the manuscript, C.B.B analyzed the data, and revised the manuscript, A.L.N analyzed the data, and revised the manuscript, M.G.H analyzed the data, and revised the manuscript, L.R analyzed the data, and revised the manuscript, A.B analyzed the data, and revised the manuscript, J.S analyzed the data, and revised the manuscript, M.A analyzed the data, and revised the manuscript, T.A analyzed the data, and revised the manuscript, A.F analyzed the data, and revised the manuscript, M.N analyzed the data, and revised the manuscript, C.S analyzed the data, and revised the manuscript, S.M analyzed the data, and revised the manuscript, D.S.E analyzed the data, and revised the manuscript', G.W.M analyzed the data, and revised the manuscript, M.R.N analyzed the data, and revised the manuscript, M.S analyzed the data, and revised the manuscript, J.E analyzed the data, and revised the manuscript, J.S designed the study, analyzed the data, revised the manuscript.

Competing interests

No competing interests were declared.

Grant information

Dr. Weiss was supported by an Epilepsy Foundation Award Research and Training Fellowship for Clinicians.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

The authors would like to thank Mrs. Sandra Dewar for her administrative assistance, and Mr. Kirk Shattuck for his technical support.

Supplementary material

Supplementary File 1: Multiple choice questionnaire, completed by the reader after review of the ictal record.

Click here to access the data.

References

1. Williamson PD, Spencer DD, Spencer SS, *et al.*: **Complex partial seizures of frontal lobe origin.** *Ann Neurol.* 1985; **18**(4): 497–504.
PubMed Abstract | Publisher Full Text
2. Laskowitz DT, Sperling MR, French JA, *et al.*: **The syndrome of frontal lobe epilepsy: characteristics and surgical management.** *Neurology.* 1995; **45**(4): 780–7.
PubMed Abstract | Publisher Full Text
3. Foldvary N, Klem G, Hammel J, *et al.*: **The localizing value of ictal EEG in focal epilepsy.** *Neurology.* 2001; **57**(11): 2022–8.
PubMed Abstract | Publisher Full Text
4. Knowlton RC: **The role of FDG-PET, ictal SPECT, and MEG in the epilepsy surgery evaluation.** *Epilepsy Behav.* 2006; **8**(1): 91–101. Review.
PubMed Abstract | Publisher Full Text
5. Gotman J, Ives JR, Gloor P: **Frequency content of EEG and EMG at seizure onset: possibility of removal of EMG artefact by digital filtering.** *Electroencephalogr Clin Neurophysiol.* 1981; **52**(6): 626–39.
PubMed Abstract | Publisher Full Text
6. Bautista RE, Spencer DD, Spencer SS: **EEG findings in frontal lobe epilepsies.** *Neurology.* 1998; **50**(6): 1765–71.
PubMed Abstract | Publisher Full Text
7. Worrell GA, So EL, Kazemi J, *et al.*: **Focal ictal beta discharge on scalp EEG predicts excellent outcome of frontal lobe epilepsy surgery.** *Epilepsia.* 2002; **43**(3): 277–82.
PubMed Abstract | Publisher Full Text
8. Makeig S, Bell AJ, Jung TP, et al.: **Independent component analysis of electroencephalographic data.** *Advances in neural information processing systems.*1996; 145–151.
Reference Source
9. Makeig S, Jung TP, Bell AJ, *et al.*: **Blind separation of auditory event-related brain responses into independent components.** *Proc Natl Acad Sci U S A.* 1997; **94**(20): 10979–84.
PubMed Abstract | Publisher Full Text | Free Full Text
10. Jung TP, Makeig S, Humphries C, *et al.*: **Removing electroencephalographic artifacts by blind source separation.** *Psychophysiology.* 2000; **37**(2): 163–78.
PubMed Abstract | Publisher Full Text
11. 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**(2): 113–24. Review.
PubMed Abstract | Publisher Full Text
12. Delorme A, Sejnowski T, Makeig S: **Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis.** *Neuroimage.* 2007; **34**(4): 1443–9.
PubMed Abstract | Publisher Full Text | Free Full Text
13. Viola FC, Thorne J, Edmonds B, *et al.*: **Semi-automatic identification of independent components representing EEG artifact.** *Clin Neurophysiol.* 2009; **120**(5): 868–77.
PubMed Abstract | Publisher Full Text
14. Delorme A, Palmer J, Onton J, *et al.*: **Independent EEG sources are dipolar.** *PLoS One.* 2012; **7**(2): e30135.
PubMed Abstract | Publisher Full Text | Free Full Text
15. Urrestarazu E, Iriarte J, Alegre M, *et al.*: **Independent component analysis removing artifacts in ictal recordings.** *Epilepsia.* 2004; **45**(9): 1071–8.
PubMed Abstract | Publisher Full Text
16. Vergult A, De Clercq W, Palmini A, *et al.*: **Improving the interpretation of ictal scalp EEG: BSS-CCA algorithm for muscle artifact removal.** *Epilepsia.* 2007; **48**(5): 950–8.
PubMed Abstract | Publisher Full Text
17. Chaumon M, Bishop DV, Busch NA: **A practical guide to the selection of independent components of the electroencephalogram for artifact correction.** *J Neurosci Methods.* 2015; **250**: 47–63.
PubMed Abstract | Publisher Full Text
18. Mognon A, Jovicich J, Bruzzone L, *et al.*: **ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features.** *Psychophysiology.* 2011; **48**(2): 229–40.
PubMed Abstract | Publisher Full Text
19. De Vos M, Deburchgraeve W, Cherian PJ, *et al.*: **Automated artifact removal as preprocessing refines neonatal seizure detection.** *Clin Neurophysiol.* 2011; **122**(12): 2345–54.
PubMed Abstract | Publisher Full Text
20. Nierenberg N, Wilson SB, Scheuer ML: **Method And System For Detecting And Removing EEG Artifacts.** U.S. Patent Application No. 13/684,556. Published; 2013.
Reference Source
21. Strehl A, Ghosh J: **Cluster Ensembles – A Knowledge Reuse Framework for Combining Multiple Partitions.** *J Mach Learn Res.* 2002; **3**: 583–617.
Reference Source
22. Raimondo F, Kamienkowski JE, Sigman M, *et al.*: **CUDAICA: GPU optimization of Infomax-ICA EEG analysis.** *Comput Intell Neurosci.* 2012; **2012**: 206972, 1–8.
PubMed Abstract | Publisher Full Text | Free Full Text
23. Lee TW, Girolami M, Sejnowski TJ: **Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources.** *Neural Comput.* 1999; **11**(2): 417–441.
PubMed Abstract | Publisher Full Text
24. Spencer SS, Williamson PD, Bridgers SL, *et al.*: **Reliability and accuracy of localization by scalp ictal EEG.** *Neurology.* 1985; **35**(11): 1567–75.
PubMed Abstract | Publisher Full Text
25. shennanw: **shennanw/AR2: AR2 [Data set].** *Zenodo.* 2017.
Data Source
26. Weiss S: **Validity of two automatic artifact reduction software methods in ictal EEG interpretation. Dataset 1 [Data set].** *Zenodo.* 2016.
Data Source

Open Peer Review

Current Referee Status: ✔ ✔

Version 2

Referee Report 12 April 2017

doi:10.5256/f1000research.12199.r21547

✔ **Patrícia Figueiredo** , **Rodolfo Abreu**
Institute for Systems and Robotics and Department of Bioengineering, Instituto Superior Técnico University of Lisbon, Lisbon, Portugal

I appreciate the Authors' efforts to address most of my major concerns/suggestions. I approve this manuscript to pass peer review, however I believe it would still benefit from these minor corrections:

Specific comments

4. "*ICA derives spatial features that can remove artifacts that have static scalp topographies and time courses of activity that are distinct from that of EEG sources."* I believe that what the Authors mean is "ICA derives spatial features that **can be used to identify** artifacts that...". Please confirm.

5. I believe that the Authors forgot to include this relevant reference: *Urigüen, J.A., Garcia-Zapirain, B., 2015. EEG artifact removal—state-of-the-art and guidelines. J. Neural Eng. 12, 31001*[1]. Please confirm.

6. I feel that further details are still needed regarding AR1 because presently, it is not clear the rationale underlying AR1, nor its differences in relation to AR2.

References

1. Urigüen JA, Garcia-Zapirain B: EEG artifact removal-state-of-the-art and guidelines.*J Neural Eng*. 2015; **12** (3): 031001 PubMed Abstract | Publisher Full Text

Competing Interests: No competing interests were disclosed.

Referee Expertise: Neuroimaging, Signal processing

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 07 April 2017

doi:10.5256/f1000research.12199.r21546

David M. Groppe
The Krembil Neuroscience Centre, Toronto, ON, Canada

The authors have provided a thoughtful revision that has address all my major suggestions/concerns. I think the current version of the manuscript is a positive contribution to the literature and approve it.

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Referee Report 16 March 2017

doi:10.5256/f1000research.11390.r20245

David M. Groppe
The Krembil Neuroscience Centre, Toronto, ON, Canada

In this manuscript, Weiss and colleagues present a novel algorithm for removing electromyographic (EMG) artifacts from ictal EEG recordings, called AR2. Moreover, they evaluate the performance of the algorithm on data from 8 patients and compare it to a similar commercial algorithm, AR1 (i.e., Persyst v12's artifact correction software), using readings by 26 neurologists. The data chosen were so corrupted by EMG artifacts that they were not interpretable using conventional frequency-based filtering. Both AR1 and AR2 rely on independent components analysis (ICA) to remove EMG artifacts via spatial filters that are learned from the data. There is strong evidence that ICA is effective at removing EMG (and other EEG-artifacts) from data acquired in controlled, research settings[ref1]-[2]. However, there may be too many EMG sources in highly polluted ictal recordings for ICA to work.

In general, the authors found that both algorithms (1) made around 50% of the seizures interpretable with typically low levels of rater confidence and (2) produced very low-levels of inter-rater agreement. Nonetheless, when compelling seizure-onset lateralization was available from other sources of data (e.g., PET, SPECT), the algorithms led to EEG interpretations that were in concordance in about 80% of seizures (Table 2). Moreover, AR2 tended to slightly outperform AR1. Specifically, neurologists could interpret more seizures and tended to have more confidence in their interpretations following AR2 artifact correction. However, there was no statistically significant difference in inter-rater agreement between algorithms. The authors conclude from this that their AR2 algorithm "may improve the validity of ictal EEG artifact reduction."

In general, I think the authors' work is laudable and that it is a valuable contribution to the literature. AR2 is well motivated given the evidence that ICA is successful at removing EMG (and other EEG artifacts) from data acquired in controlled research settings and the approach they have taken to validate their algorithm is generally sound. Moreover it is impressive that all of the seizures were read by a large number of neurologists, (26; although it is not clear how many were board certified in epilepsy or clinical

neurophysiology) and that they have made all of their code and data public.

However, there are some significant issues with this work that qualify their findings and should be addressed in revisions or future work:

-As the authors note, the data for this study was obtained from a small number of patients (8, only 5 of whom had lateralized seizure foci based on independent data). Thus, it is not clear how robust some of their findings are (e.g., the small differences between AR1 and AR2 performance).
-Although AR2 is a fully automatic algorithm, there are some arbitrary parameters of the algorithm (e.g., the mutual information threshold used to include an electrode in the artifact correction procedure) that must have been set based on exploratory analyses. If the data used to set these parameters are the same data used to validate the algorithm, then the authors are surely over-estimating, to some extent, the automatic performance of the algorithm. The authors need to specify what data were used to fix the parameters of AR2.
-It is important to note that the authors chose extremely contaminated data to evaluate AR1 and AR2 and that these algorithms might be more useful when applied to less contaminated data.
-If I understand the text correctly, AR2 excludes non-artifact contaminated electrodes from its analysis. You should include these electrodes in the ICA decompositions because they will help capture the neurogenic signal you are trying to preserve.
-Since ICA necessarily removes some neurogenic signal along with EEG artifacts, it can help to quantify this by applying your algorithm to non-artifact polluted data [2]. Adding such an analysis to these findings would help us to understand how and how much AR2 might be distorting EEG seizure activity. Electrodes closest to muscles are likely most affected.
-For many statistical hypothesis tests the authors provide only p-values. It would be much more informative if the authors provided test statistics (e.g., t-scores, degrees of freedom), named the type of test (e.g., cumulative logit mixed effect model) and confidence intervals. In particular, confidence intervals will be much better than p-values at communicating how important and robust these effects are [3].
-Figures 4-5 report $p<0.05$ for the results of a large number of statistical tests (23 per subfigure) with no correction for multiple comparisons. You should perform some type of correction (e.g., Bonferroni-Holm or Benjamini & Hochberg's false discovery rate control algorithm).
-To interpret these results, it would greatly help to have inter-reader reliability and reader confidence values for non-artifact contaminated data. Can you get these from the existing literature?
-I think the primary finding of this work is that neither AR1 nor AR2 provide robust artifact correction when applied to such heavily contaminated data and need to be improved. You should discuss what improvements (if any) you think could be made. For example, using higher-density EEG recordings could greatly help. With more electrodes, ICA's performance should improve (given sufficient training data).

In addition to those major points, here are some additional suggestions and points of consideration/clarification:

The abstract should specify the consistency of AR1-derived lateralization with behavioural, neurophysiological, and neuro-radiological findings. Currently, only the consistency with AR2-derived lateralization is reported.
-[pg 3]: Saying "ICA removes artifacts based on source-related features instead of frequencies." is too vague to be informative. You might consider providing more details, such as "ICA derives spatial filters that can remove artifacts that have static scalp topographies and time courses of activity that are distinct from that of EEG sources. ICA artifact correction is necessarily imperfect and will remove some neurogenic components of the EEG as well[4]. However, the degree of EEG distortion may be negligible

and ICA has proven effective at removing EMG and ocular artifacts from EEG data recorded from neuronormal individuals in laboratory settings[2]."

-The introduction should note why ICA might not be able to correct for EMG-ictal artifact, even though it has proven useful for less artifact-polluted research data. Specifically, it may fail because the number of EEG artifact sources may be much greater in ictal data.

-You should include the article by De Vos et al. (2011)[5] in your review of previous algorithms for correcting EEG artifacts in clinical epilepsy data with ICA.

-You say that EEG readings were provided by "26 neurologists with a specialization in EEG." Please specify how many were board certified in epilepsy or clinical neurophysiology.

-It appears that AR2 is applied to epochs that are not contaminated with EMG (pg 3, bottom left). Why try to correct artifacts that aren't there?

-Instead of saying "independent components of greatest order," I think it is more conventional to say "independent components that account for the most variance."

-Please provide the specifications of the analog filter used to acquire the data.

It would help to explicitly report the number of data points per electrode fed to ICA. The reliability of ICA is a function of this[6].

-It might help to clearly state that the AR1 and AR2 processed data were both read using the same graphical user interface (i.e., Persyst's). It took me a little while to figure this out and it's great that you did this.

-It would help to add titles to subfigures (if it is permitted by F1000's formatting guidelines).

-In Figure 1 there is no point to showing both the non-normalized ICA and normalized mixing matrix since the mixing matrix column scale is arbitrary. Just show the normalized mixing matrix. It would also help to view the mixing matrix weights as scalp topographies to see both the quality of the putative neurogenic and EMG ICs.

-[pg 4] You say "Compared to AR1, more readers were able to render seizure-onset laterality assignments using AR2, and these assignments were more often congruent with other clinical data (Table 2)." However in Table 2, 82% of the seizures that were lateralizable with AR1 (i.e., 145/177) agree with clinical findings in contrast to 81% of seizures using AR2 (i.e., 171/210). I think percentage of agreement is more important than the number of seizures in agreement.

-[pg 11] You say "Among the 8 patients, if the reader lateralized the seizure-onset to the left using AR2 they were correct in 95.9%….". Do you mean "Among the 5 patients" with clinical seizure onset lateralization based on independent data?

-I think your statement "With regard to AR2, the novel software method developed for this study, the slight improvement seen in ictal EEG interpretability after applying the method suggests that the algorithm can (1) reliably produce signals that are, exclusively or mainly, EEG or EMG, and (2) identify which signals are of brain origin and which are contaminant." is overly strong. I think "sometimes" is more accurate than "reliably" given the low reader confidence and inter-reader agreement.

-I don't understand your statement "One explanation for AR2's ability to isolate myogenic from neurogenic independent components may be that scalp EEG electrodes record weighted and summated far-field signals from all brain and muscle sources, as well as near-field electrode noise generated at the electrode/skin interface." ICA can separate myogenic from neurogenic activity because they have distinct scalp topographies and largely independent time courses of activity.

-It is fantastic that you have made both AR2's code and your data publicly available. However, there is not enough documentation on your GitHub repo for me to be able to easily understand how to use it (what is scalp_input_matrix.mat?). A little bit more documentation would greatly help.

References

1. Jung TP, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui V, Sejnowski TJ: Removing electroencephalographic artifacts by blind source separation.*Psychophysiology*. 2000; **37** (2): 163-78

PubMed Abstract
2. Mognon A, Jovicich J, Bruzzone L, Buiatti M: ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features.*Psychophysiology*. 2011; **48** (2): 229-40 PubMed Abstract | Publisher Full Text
3. Groppe DM: Combating the scientific decline effect with confidence (intervals).*Psychophysiology*. 2017; **54** (1): 139-145 PubMed Abstract | Publisher Full Text
4. Groppe DM, Makeig S, Kutas M: Independent component analysis of event-related potentials. *Cognitive Science Online*. 2008; **6** (1): 11-44 Reference Source
5. De Vos M, Deburchgraeve W, Cherian PJ, Matic V, Swarte RM, Govaert P, Visser GH, Van Huffel S: Automated artifact removal as preprocessing refines neonatal seizure detection.*Clin Neurophysiol*. 2011; **122** (12): 2345-54 PubMed Abstract | Publisher Full Text
6. Groppe DM, Makeig S, Kutas M: Identifying reliable independent components via split-half comparisons.*Neuroimage*. 2009; **45** (4): 1199-211 PubMed Abstract | Publisher Full Text

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 28 Mar 2017

Shennan Weiss, Thomas Jefferson University

Dear Dr. David Groppe,

We are grateful for your insightful and thoughtful comments and suggestions. Appended below are answers to your inquiries, and changes we have made to the manuscript.

-As the authors note, the data for this study was obtained from a small number of patients (8, only 5 of whom had lateralized seizure foci based on independent data). Thus, it is not clear how robust some of their findings are (e.g., the small differences between AR1 and AR2 performance).

-- The authors agree that this study is underpowered. Our findings are exploratory at best.

-Although AR2 is a fully automatic algorithm, there are some arbitrary parameters of the algorithm (e.g., the mutual information threshold used to include an electrode in the artifact correction procedure) that must have been set based on exploratory analyses. If the data used to set these parameters are the same data used to validate the algorithm, then the authors are surely over-estimating, to some extent, the automatic performance of the algorithm. The authors need to specify what data were used to fix the parameters of AR2.

-- You are correct that we used the experimental dataset to define the threshold values and thus we are likely over-estimating the performance of the algorithm. We clarify on (pg.3) and (pg.4) that the thresholds were defined using visual inspection of the experimental dataset in the revised manuscript.

-It is important to note that the authors chose extremely contaminated data to evaluate AR1 and AR2 and that these algorithms might be more useful when applied to less contaminated data.

-- On (pg.2) we now specify “Ictal scalp EEG recordings present extraordinary challenges to ICA artifact reduction algorithms because the number of EMG artifact sources increases.”

-If I understand the text correctly, AR2 excludes non-artifact contaminated electrodes from its analysis. You should include these electrodes in the ICA decompositions because they will help capture the neurogenic signal you are trying to preserve.

-- We apologize for the lack of clarity. We only excluded electrodes that had suspected increases in impedance. We specify on (pg.3) “Prior to performing ICA to remove muscle artifact, the algorithm first identified epochs of the scalp EEG record contaminated by muscle artifact and determined the electrodes that were suspected of having high recording impedance during that epoch. The purpose of these calculations was to exclude these electrodes from the ICA calculations.”

-Since ICA necessarily removes some neurogenic signal along with EEG artifacts, it can help to quantify this by applying your algorithm to non-artifact polluted data [2]. Adding such an analysis to these findings would help us to understand how and how much AR2 might be distorting EEG seizure activity. Electrodes closest to muscles are likely most affected.

-- We agree that this analysis would be helpful and should be a focus of future study. Unfortunately, the EEG reviewers who participated in this study are not available to review non-ictal scalp EEG recordings.

-For many statistical hypothesis tests the authors provide only p-values. It would be much more informative if the authors provided test statistics (e.g., t-scores, degrees of freedom), named the type of test (e.g., cumulative logit mixed effect model) and confidence intervals. In particular, confidence intervals will be much better than p-values at communicating how important and robust these effects are [3].

-- As you suggested we now provide t-scores, degrees of freedom, and have named the type of the test in the results. We provide confidence intervals for the cumulative logit mixed effects models results, and the correlation with other clinical data. S.E.M values are provided for the other comparisons in the figures included in the manuscript. The authors are in agreement the confidence intervals are essential to convey effect size[3]

-Figures 4-5 report $p<0.05$ for the results of a large number of statistical tests (23 per subfigure) with no correction for multiple comparisons. You should perform some type of correction (e.g., Bonferroni-Holm or Benjamini & Hochberg’s false discovery rate control algorithm).

-- We have used your Matlab code to perform the Bonferroni-Holm correction on the p values obtained for the individual seizures. The results have been revised accordingly (see methods, statistical analysis).

-To interpret these results, it would greatly help to have inter-reader reliability and reader confidence values for non-artifact contaminated data. Can you get these from the existing literature?

-- We agree and the following sentence has been added to the discussion (pg. 13): The reliability of localization by ictal scalp EEG in the absence of artifact is between 65-75% for lateralization[26].

-I think the primary finding of this work is that neither AR1 nor AR2 provide robust artifact correction when applied to such heavily contaminated data and need to be improved. You should discuss what improvements (if any) you think could be made. For example, using higher-density EEG recordings could greatly help. With more electrodes, ICA's performance should improve (given sufficient training data).

-- Thank you for this helpful suggestion, we have added the following sentence to the discussion (pg. 13): The effectiveness of AR2 could possibly be improved by utilizing autocorrelations to identify the myogenic independent components. We hope that this method can be optimized for 10/20 standard scalp EEG.

In addition to those major points, here are some additional suggestions and points of consideration/clarification:

The abstract should specify the consistency of AR1-derived lateralization with behavioural, neurophysiological, and neuro-radiological findings. Currently, only the consistency with AR2-derived lateralization is reported.

-- We have provided the results for AR1 in the abstract as you suggested.

-[pg 3]: Saying "ICA removes artifacts based on source-related features instead of frequencies." is too vague to be informative. You might consider providing more details, such as "ICA derives spatial filters that can remove artifacts that have static scalp topographies and time courses of activity that are distinct from that of EEG sources. ICA artifact correction is necessarily imperfect and will remove some neurogenic components of the EEG as well[4]. However, the degree of EEG distortion may be negligible and ICA has proven effective at removing EMG and ocular artifacts from EEG data recorded from neuronormal individuals in laboratory settings[2]."

-- Thank you for your suggestion we have made these verbatim changes to the introduction (pg. 2)

-The introduction should note why ICA might not be able to correct for EMG-ictal artifact, even though it has proven useful for less artifact-polluted research data. Specifically, it may fail because the number of EEG artifact sources may be much greater in ictal data.

-- We have addressed this issue as mentioned in a prior comment to you.

-You should include the article by De Vos et al. (2011)[5] in your review of previous algorithms for correcting EEG artifacts in clinical epilepsy data with ICA.

-- done as suggested

-You say that EEG readings were provided by "26 neurologists with a specialization in EEG." Please specify how many were board certified in epilepsy or clinical neurophysiology.

-- 20 of the readers were board certified as now specified on (pg.4.)

-It appears that AR2 is applied to epochs that are not contaminated with EMG (pg 3, bottom left). Why try to correct artifacts that aren't there?

-- As specified in the methods we performed the ICA on 120 second trials irrespective of the beginning and end of the ictal EMG artifact. We used this approach in order to allow the algorithm to function in an automated and unsupervised manner.

-Instead of saying "independent components of greatest order," I think it is more conventional to say "independent components that account for the most variance."

-- We have made this modification as you suggested (pg. 3)

-Please provide the specifications of the analog filter used to acquire the data. It would help to explicitly report the number of data points per electrode fed to ICA. The reliability of ICA is a function of this[6].

-- We now specify 24,000 data points in the methods (pg.4)

-It might help to clearly state that the AR1 and AR2 processed data were both read using the same graphical user interface (i.e., Persyst's). It took me a little while to figure this out and it's great that you did this.

-- We have modified the methods as follows (pg. 5): The AR1 and AR2 processed data were reviewed in Persyst v12 without video by 26 neurologists with a specialization in EEG, 20 of whom were board certified.

-It would help to add titles to subfigures (if it is permitted by F1000's formatting guidelines).

-- As far as I know this is not possible.

-In Figure 1 there is no point to showing both the non-normalized ICA and normalized mixing matrix since the mixing matrix column scale is arbitrary. Just show the normalized mixing matrix. It would also help to view the mixing matrix weights as scalp topographies to see both the quality of the putative neurogenic and EMG ICs.

-- We have changed the figure as you suggested and modified the legend.

-[pg 4] You say "Compared to AR1, more readers were able to render seizure-onset laterality assignments using AR2, and these assignments were more often congruent with other clinical data

(Table 2)." However in Table 2, 82% of the seizures that were lateralizable with AR1 (i.e., 145/177) agree with clinical findings in contrast to 81% of seizures using AR2 (i.e., 171/210). I think percentage of agreement is more important than the number of seizures in agreement.

-- Thank you for this insightful point. The numbers do not refer to the number of seizures in agreement but rather to the number of observations i.e. assignments made that agreed with the laterality defined by other clinical data. Thus, more readers were able to render observations that agreed with other clinical data using AR2 as compared to AR1. However, as you point out the percentage of readers that made rendered a laterality decision that did not agree with the other clinical data using AR2 was comparable to AR1.

-[pg 11] You say "Among the 8 patients, if the reader lateralized the seizure-onset to the left using AR2 they were correct in 95.9%….". Do you mean "Among the 5 patients" with clinical seizure onset lateralization based on independent data?

-- You are correct and we apologize for the lack of clarity. We have modified the results as follows (pg. 12): Among the 5 patients with clinical seizure onset lateralization based on independent data, …

-I think your statement "With regard to AR2, the novel software method developed for this study, the slight improvement seen in ictal EEG interpretability after applying the method suggests that the algorithm can (1) reliably produce signals that are, exclusively or mainly, EEG or EMG, and (2) identify which signals are of brain origin and which are contaminant." is overly strong. I think "sometimes" is more accurate than "reliably" given the low reader confidence and inter-reader agreement.

-- We agree and have modified the sentence as you suggested (pg.13).

-I don't understand your statement "One explanation for AR2's ability to isolate myogenic from neurogenic independent components may be that scalp EEG electrodes record weighted and summated far-field signals from all brain and muscle sources, as well as near-field electrode noise generated at the electrode/skin interface." ICA can separate myogenic from neurogenic activity because they have distinct scalp topographies and largely independent time courses of activity.

-- Thank you for pointing out that this sentence lacks clarity. We have modified this paragraph as follows (pg.13): One explanation for AR2's ability to isolate myogenic from neurogenic activity may be related to the respective dipole generators of each. ICA produces independent components that may resemble single equivalent dipoles[14]. Presumably, networks of myocytes exhibit shorter distance connectivity then networks of neurons that produce beta and gamma oscillations, and thus the two generators can be distinguished on the basis of the focality[17] of the independent components topography.

-It is fantastic that you have made both AR2's code and your data publicly available. However, there is not enough documentation on your GitHub repo for me to be able to easily understand how to use it (what is scalp_input_matrix.mat?). A little bit more documentation would greatly help.

-- We are in the process of improving the documentation of AR2 on GitHub thank you for reviewing the source code.

References

1. Jung TP, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui V, Sejnowski TJ: Removing electroencephalographic artifacts by blind source separation.*Psychophysiology*. 2000; **37** (2): 163-78 PubMed Abstract
2. Mognon A, Jovicich J, Bruzzone L, Buiatti M: ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features.*Psychophysiology*. 2011; **48** (2): 229-40 PubMed Abstract | Publisher Full Text
3. Groppe DM: Combating the scientific decline effect with confidence (intervals). *Psychophysiology*. 2017; **54** (1): 139-145 PubMed Abstract | Publisher Full Text
4. Groppe DM, Makeig S, Kutas M: Independent component analysis of event-related potentials. *Cognitive Science Online*. 2008; **6** (1): 11-44 Reference Source
5. De Vos M, Deburchgraeve W, Cherian PJ, Matic V, Swarte RM, Govaert P, Visser GH, Van Huffel S: Automated artifact removal as preprocessing refines neonatal seizure detection.*Clin Neurophysiol*. 2011; **122** (12): 2345-54 PubMed Abstract | Publisher Full Text
6. Groppe DM, Makeig S, Kutas M: Identifying reliable independent components via split-half comparisons.*Neuroimage*. 2009; **45** (4): 1199-211 PubMed Abstract | Publisher Full Text

Competing Interests: none.

Referee Report 22 February 2017

doi:10.5256/f1000research.11390.r20461

? **Patrícia Figueiredo** , **Rodolfo Abreu**
Institute for Systems and Robotics and Department of Bioengineering, Instituto Superior Técnico University of Lisbon, Lisbon, Portugal

Referee report
The manuscript needs careful revision by a native English speaker within the scientific community. Although I feel that the performance measures used by the authors are adequate, and that a substantial number of EEG specialists quantified them, the overall results are poor, particularly in terms of the specialists' confidence in their assessment. It would be therefore useful to compare the performance with additional methods, for instance, as to understand the behavior of the proposed performance measures across a larger set of methods.

Specific Comments

Title:

1. The title should mention that this novel method specifically addresses EEG artifacts induced by myogenic activity.

Abstract:

2. (Results) The authors should include the consistency value also for AR1.

Introduction:

3. (page 3, first paragraph) "*Each of these tests adds undesired time and cost to the evaluation*". I would say that the necessity of using additional imaging techniques depends on how precise one wants seizure-onset zone delineation to be, as scalp EEG has a poor spatial resolution and localization power. Please elaborate and/or re-phrase the sentence accordingly.

4. (page 3, second paragraph) "*ICA removes artifacts based on source-related features instead of frequencies"*. What do Authors mean with "source-related features"? Actually, there are several studies that use frequency-based criteria for the selection and subsequent removal of artifact-related sources…Please explain.

5. (page 3, second paragraph) Please add more recent reviews/papers on the automatic IC selection for EEG cleaning, such as: *Chaumon, M., Bishop, D.V.M., Busch, N. a., 2015. A Practical Guide to the Selection of Independent Components of the Electroencephalogram for Artifact Correction. J. Neurosci. Methods.* or *Urigüen, J.A., Garcia-Zapirain, B., 2015. EEG artifact removal—state-of-the-art and guidelines. J. Neural Eng. 12, 31001*.[1]

6. (page 3, fourth paragraph) Authors refer to AR1 as a commercially available software, and in fact, detailed information about it is provided in reference [17]. However, the Authors should provide a brief description of the method because: 1) it is the only method which they compare their novel one with; and 2) so that future readers do not need to go through [17] in order to understand the overall rationale of AR1.

Methods:

7. (page 3, *Implementation*, first paragraph) *"(…) a power spectral density algorithm to find extended intervals of elevated high frequency power across electrodes"*. The Authors provide no information about how this algorithm works, nor references; thus, it is presently not possible to reproduce this part of the study.
8. (page 3, *Implementation*, first paragraph) The Authors need to justify their choices in general; particularly, why only compute the adjacency matrix between the epoch of greatest duration across all electrodes? Why compute the adjacency matrix in the first place, and not any other discriminative feature for the presence of muscle artifacts? Why only assign the maximum pairwise MI value in the adjacency matrix to a given electrode and ignore all the rest? How was the MI threshold determined?

9. (page 3, *Implementation*, second paragraph) Again, the Authors need to provide more details overall. Why segment EEG into consecutive epochs of 120 s? How exactly was the variance threshold derived? Also, I did not understand why should be there any order associated with myogenic and neurogenic components ("*We assumed that the last myogenic component and first neurogenic component (…)"*).

10. (page 3, *Implementation*, second paragraph) I understand that one of the expected features of ICs reflecting muscle artifacts is having a focal spatial topography; however, bad channels are also reflected in ICs exhibiting this feature. Thus, I have severe concerns about false positives when using this criterion, as other myogenic-unrelated ICs are probably being selected as well, which may hinder a true assessment of the impact of muscle artifact correction.

11. (page 3, *Implementation*, third paragraph) What does reconstitute mean in this context?

12. (page 4, *Statistical analysis*) Since the performance of AR1 and AR2 is being assessed by 4 different performance measures obtained from 26 EEG specialists, it would be more accurate to use a 2-way repeated measures ANOVA (or its non-parametric equivalent, in the case of the samples not being normally distributed), followed by multiple comparison testing if necessary.

Results:

13. The first three figures have very poor quality. In particular, it is nearly impossible to follow the overall (quite detailed) description of Figure 2 (and it is panel A_1 on the top, left hand-side, and not A_2). Also, the three panels in Figure 3 should be overlaid to facilitate the direct comparison between the two algorithms.

14. Although AR2 outperforms AR1 for most of the performance measures, the results are still poor, making me wonder if either of these methods is suitable for EEG muscle artifact correction.
Discussion
15. (page 12, first paragraph) What do the Authors mean with "*One concern about AR1 and AR2 relates to the lack of understanding of the waveform alteration*"?

16. (page 12, fourth paragraph) "*(…) (1) reliably produce signals that are, exclusively or mainly, EEG or MEG (…)*". Please clarify and elaborate on this claim.

References
1. Chaumon M, Bishop DV, Busch NA: A practical guide to the selection of independent components of the electroencephalogram for artifact correction.*J Neurosci Methods*. 2015; **250**: 47-63 PubMed Abstract | Publisher Full Text

Competing Interests: No competing interests were disclosed.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Author Response 28 Mar 2017

Shennan Weiss, Thomas Jefferson University

Dear Dr. Figueiredo and Dr. Abreu,

Thank you very much for your thoughtful and helpful comments and suggestions. We have substantially revised the manuscript according to your feedback as follows:

1. The title should mention that this novel method specifically addresses EEG artifacts induced by myogenic activity.

-- The title of the paper has been modified to "AR2, a novel automatic muscle artifact reduction software method for ictal EEG interpretation: Validation and comparison of performance with commercially available software"

Abstract:

2. (Results) The authors should include the consistency value also for AR1.

-- As you suggested we have added the consistency values for AR1 to the abstract.

Introduction:

3. (page 3, first paragraph) "*Each of these tests adds undesired time and cost to the evaluation*". I would say that the necessity of using additional imaging techniques depends on how precise one wants seizure-onset zone delineation to be, as scalp EEG has a poor spatial resolution and localization power. Please elaborate and/or re-phrase the sentence accordingly.

-- We have modified the introduction as follows (pg. 3): The inability, or lack of precision, to discern the seizure-onset zone from scalp EEG often necessitates additional testing, …

4. (page 3, second paragraph) "*ICA removes artifacts based on source-related features instead of frequencies"*. What do Authors mean with "source-related features"? Actually, there are several studies that use frequency-based criteria for the selection and subsequent removal of artifact-related sources…Please explain.

-- Thank you for this instructive feedback. We have modified the introduction as follows (pg.3): ICA derives spatial features that can remove artifacts that have static scalp topographies and time courses of activity that are distinct from that of EEG sources.

5. (page 3, second paragraph) Please add more recent reviews/papers on the automatic IC selection for EEG cleaning, such as: *Chaumon, M., Bishop, D.V.M., Busch, N. a., 2015. A Practical Guide to the Selection of Independent Components of the Electroencephalogram for Artifact Correction. J. Neurosci. Methods.* or *Urigüen, J.A., Garcia-Zapirain, B., 2015. EEG artifact removal—state-of-the-art and guidelines. J. Neural Eng. 12, 31001*.[1]

-- Thank you for suggesting the inclusion of this important methods article. We now cite this article in the introduction and discussion.

6. (page 3, fourth paragraph) Authors refer to AR1 as a commercially available software, and in fact, detailed information about it is provided in reference [17]. However, the Authors should provide a brief description of the method because: 1) it is the only method which they compare their novel one with; and 2) so that future readers do not need to go through [17] in order to understand the overall rationale of AR1.

-- Although the complete methods for AR1 are not included in reference 17 we have modified the introduction as follows (pg.3): The goal of this study was to assess the validity of a commercially available EEG artifact reduction tool (AR1) that uses different montages and within electrode analysis to identify artefactual independent components[20], and compare its validity to a novel automatic artifact reduction tool (AR2)…

Methods:

7. (page 3, *Implementation*, first paragraph) *"(…) a power spectral density algorithm to find extended intervals of elevated high frequency power across electrodes"*. The Authors provide no information about how this algorithm works, nor references; thus, it is presently not possible to reproduce this part of the study.

8. (page 3, *Implementation*, first paragraph) The Authors need to justify their choices in general; particularly, why only compute the adjacency matrix between the epoch of greatest duration across all electrodes? Why compute the adjacency matrix in the first place, and not any other discriminative feature for the presence of muscle artifacts? Why only assign the maximum pairwise MI value in the adjacency matrix to a given electrode and ignore all the rest? How was the MI threshold determined?

-- We agree with your comments #7 and #8. We now specify in the methods that the reason we performed this analysis was (pg.3): "Prior to performing ICA to remove muscle artifact, the algorithm first identified epochs of the scalp EEG record contaminated by muscle artifact and determined the electrodes that were suspected of having high recording impedance during that epoch. The purpose of these calculations was to exclude these electrodes from the ICA calculations."
The method used to determine the artifact epoch had actually been modified prior to submission of version 1 of the manuscript. We now better describe this algorithm as "We then calculated the normalized instantaneous amplitude of the band-pass filtered signal using a Hilbert transform. This signal was smoothed using moving averaging, and the algorithm identified the longest epoch in which the time series remained greater than one standard deviation."

9. (page 3, *Implementation*, second paragraph) Again, the Authors need to provide more details overall. Why segment EEG into consecutive epochs of 120 s? How exactly was the variance threshold derived? Also, I did not understand why should be there any order associated with myogenic and neurogenic components ("*We assumed that the last myogenic component and first neurogenic component (…)"*).

-- We agree with your comment and apologize for the lack of clarity. We now specify that (pg.4): A 120 second trial length was chosen to optimize processing time.

In addition, the method have been modified as follows (pg. 4): "We assumed that the last myogenic component and first neurogenic component can be differentiated on the basis of the inverse weight matrix, which provides the spatial distribution of each component, and identifying the independent component that account for the most variance with a focal spatial topography[17] defined on the basis of exceeding a normalized threshold of two standard deviations in at least one electrode of the inverse weight matrix. This threshold was chosen on the basis of visual inspection of the EEG in the experimental dataset and resulting independent components."

10. (page 3, *Implementation*, second paragraph) I understand that one of the expected features of ICs reflecting muscle artifacts is having a focal spatial topography; however, bad channels are also reflected in ICs exhibiting this feature. Thus, I have severe concerns about false positives when using this criterion, as other myogenic-unrelated ICs are probably being selected as well, which

may hinder a true assessment of the impact of muscle artifact correction.

-- We agree with your concerns however in the algorithm we already excluded bad channels using the algorithm described with reference to comments #7 and #8.

11. (page 3, *Implementation*, third paragraph) What does reconstitute mean in this context?

-- We now specify in implementation (pg.4) that: the resulting low pass filtered EEG was reconstituted by addition of the waveforms with the high pass (>16 Hz) filtered EEG

12. (page 4, *Statistical analysis*) Since the performance of AR1 and AR2 is being assessed by 4 different performance measures obtained from 26 EEG specialists, it would be more accurate to use a 2-way repeated measures ANOVA (or its non-parametric equivalent, in the case of the samples not being normally distributed), followed by multiple comparison testing if necessary.

-- We appreciate this helpful feedback. Dr. David Groppe the other reviewer of the manuscript suggested that we use paired t-tests and provide the t-value in order to convey effect size to the reader. We have followed his recommendations. Including both 2-way repeated measures ANOVA and paired t-tests would confuse the reader.

Results:

13. The first three figures have very poor quality. In particular, it is nearly impossible to follow the overall (quite detailed) description of Figure 2 (and it is panel A_1 on the top, left hand-side, and not A_2). Also, the three panels in Figure 3 should be overlaid to facilitate the direct comparison between the two algorithms.

--
a) We have made grammatical changes to figure 2, and corrected the figure A1 vs. A2 labeling. We apologize for this oversight.
b) We attempted to overlay the panels of figure 3 but the result was confusing and not visually appealing. Therefore, we cannot provide this suggested change.

14. Although AR2 outperforms AR1 for most of the performance measures, the results are still poor, making me wonder if either of these methods is suitable for EEG muscle artifact correction.

-- We agree and point out in the discussion that the readers were not confident in their interpretations using either AR1 or AR2 in the discussion (pg. 13).

Discussion
15. (page 12, first paragraph) What do the Authors mean with "*One concern about AR1 and AR2 relates to the lack of understanding of the waveform alteration*"?

-- This sentence has been modified to provide more clarity (pg.13): "One concern about AR1 and AR2 relates to the uncertainty that myogenic activity was fully removed, and neurogenic components were unaffected during waveform alteration."

16. (page 12, fourth paragraph) "*(…) (1) reliably produce signals that are, exclusively or mainly, EEG or MEG (…)"*. Please clarify and elaborate on this claim.

-- We agree with your comment that this sentence is unclear. This paragraph has been modified in the revision and now reads as follows: "One explanation for AR2's ability to isolate myogenic from neurogenic activity may be related to the respective dipole generators of each. ICA produces independent components that may resemble single equivalent dipoles[14]. Presumably, networks of myocytes exhibit shorter distance connectivity then networks of neurons that produce beta and gamma oscillations, and thus the two generators can be distinguished on the basis of the focality[17] of the independent components topography."

References

1. Chaumon M, Bishop DV, Busch NA: A practical guide to the selection of independent components of the electroencephalogram for artifact correction.*J Neurosci Methods*. 2015; **250**: 47-63 PubMed Abstract | Publisher Full Text

Competing Interests: none.