Summary: A study was performed to investigate and compare the relative performance of blind signal separation (BSS) algorithms at separating common types of contamination from EEG. The study develops a novel framework for investigating and comparing the relative performance of BSS algorithms that incorporates a realistic EEG simulation with a known mixture of known signals and an objective performance metric. The key finding is that although BSS is an effective and powerful tool for separating and removing contamination from EEG, the quality of the separation is highly dependant on the type of contamination, the degree of contamination, and the choice of BSS algorithm. BSS appears to be most effective at separating muscle and blink contamination and less effective at saccadic and tracking contamination. For all types of contamination, principal components analysis is a strong performer when the contamination is greater in amplitude than the brain signal whereas other algorithms such as second-order blind inference and Infomax are generally better for specific types of contamination of lower amplitude.

(Clin Neurophysiol 2007;24: 1–1)

Cerebral electrical activity measured by EEG scalp sensors is highly susceptible to contamination by both noncortical biologic artifacts (e.g., eye movements and blinks, muscle activity, cardiac pulse) and environmental noise (e.g., line noise, radio and electrical interference). A significant challenge for EEG research is the isolation and removal of artifacts and noise that contaminate the cortical signal of interest.

The impact of environmental noise is traditionally minimized through a combination of shielded recording environments and common-mode rejection (CMR). These techniques are successful at reducing environmental noise contamination but carry costs for experimental design. Shielded rooms are expensive and restrict experimentation to the laboratory environment, effectively precluding field measurements. The high quality of CMR in modern amplifiers has reduced the need for shielded environments and liberated EEG research from the laboratory; however, line noise such as that generated by mains power can still be a significant problem. Spectral filters are often used to eliminate persistent line noise, but this is problematic because the spectral profile of the line noise often overlaps that of the cortical signal of interest.

Noncortical biologic artifacts are the principal source of contamination of EEG recordings. The main generators of these artifacts are movements and blinks of the eyes, muscle activity primarily of the face (especially the jaw) and neck, and cardiac pulse. EEG experimental design is generally constrained by the desire to minimize the effect of these artifacts. Common techniques include the use of fixation targets to minimize eye movement, and instructing subjects not to blink, talk or move during data acquisition (Gratton, 1998). Although these methods can be effective they are inherently problematic because they preclude tasks that require eye movement, they are ineffective for involuntary movements and low-compliance groups, and they introduce additional task complexity for the subject. Even when using these methods contamination of EEG still occurs, and as such it is common practice to manually reject contaminated data before analysis. The rejection process typically involves manually reviewing the data, identifying contaminated segments, and then subsequently rejecting those segments. This process is laborious, especially on large datasets, and can result in unacceptable data loss when there is a high degree of contamination.

The alternative to rejecting contaminated data is to correct or remove the artifact/noise. Blind source separation (BSS) describes a class of algorithms that have been prominent in the recent artifact/noise correction literature. BSS is the process of recovering the source signals from a linear mixture of measured signals. Identification/classification of each source signal as being either brain or artifact/noise allows the removal of artifact signals without distortion of the desired brain signals. There are three distinct steps required for removal of artifact/noise from EEG using BSS: 1) separate (unmix) the measured EEG into sources using a BSS algorithm, 2) identify and discard artifact/noise sources and retain brain sources, and 3) project the retained brain sources back into sensor space resulting in artifact/noise-free EEG. The success of BSS-based artifact/noise removal is critically dependant on the first two processes, specifically good separation of brain from artifact/noise sources and correct classification of artifact/noise sources. This article focuses specifically on the first process by assessing the quality of source separation.
The quality of the source separation is determined primarily by the choice of BSS algorithm. The key difference between the various BSS algorithms is the measure (or statistic) that they use to assess degree of separation (or independence) of the signals. Several BSS algorithms have been specifically applied to artifact removal from EEG including (but not limited to) principal components analysis (PCA) (Lagerlund et al., 1997; Ille et al., 2002), second-order blind inference (SOBI) (Joyce et al., 2004), and independent components analysis (ICA) (Vigario, 1997; Jung et al., 2000).

PCA uses second-order statistics to recover uncorrelated (geometrically orthogonal) signals from a linear mixture. It could be argued that PCA is not a valid BSS algorithm because orthogonality is an unrealistic constraint for real-world sources. Nonetheless PCA blindly recovers signals using a second-order statistical heuristic, and even if these signals are not analogous to true sources the technique has proven useful in many domains for separating signals. PCA has been used in the construction of spatial filters for separation of artifact and brain signals (Berg and Scherg, 1991b; Lagerlund et al., 1997; Ille et al., 2002). Berg and Scherg (1991b) report that filters based on PCA are more effective at ocular artifact decontamination than other non-BSS methodologies including regression (see Gratton, 1998, for review) and source localization (Berg and Scherg, 1991a). However, it has been demonstrated that PCA is unable to separate some artifact signals from brain signals when they have similar amplitudes (Lagerlund et al., 1997).

Joyce et al. (2004) describe a method for removal of eye movement and blink artifact from EEG using SOBI. SOBI is a second-order algorithm that separates signals by ensuring that they are uncorrelated at different time lags, which means that SOBI can isolate temporally correlated sources. The data presented by these authors demonstrate the success of this technique for removing ocular artifact. The authors justify their choice of SOBI by referring to unpublished work that compares SOBI to other higher-order techniques, namely Extended Infomax, FastICA, and Jade. Presumably the main advantage of SOBI over these higher-order algorithms is that it takes advantage of the temporal domain.

ICA refers to a group of BSS algorithms that recover statistically independent signals from a linear mixture using higher-order statistics as the measure of independence. Jung et al. (2000) have demonstrated that an ICA algorithm, called Extended Infomax, can effectively separate both artifact and noise from cortical sources in the EEG data. Infomax ICA is a neural network algorithm that recovers statistically independent sources by maximizing the joint entropy of the network output. Extended Infomax is an extension of Infomax that permits separation of both subgaussian and supergaussian distributed signals. Jung et al. (1998a; 1998b; 2000) report that extended Infomax recovers more brain signal than both regression and PCA.

There is a rapidly expanding literature that demonstrates the capacity of BSS algorithms to recover brain signal from contaminated EEG; however, there has been no empirical investigation of the relative performance of these various algorithms for artifact/noise reduction other than anecdotal accounts of unpublished work. The current article introduces a framework to directly compare the performance of various BSS algorithms and uses this method to assess many of the popular algorithms at separating the common types of EEG artifact. Source identification/classification is often performed by visual inspection and is therefore susceptible to experimenter judgment/bias. This bias makes it difficult compare the relative performance of different BSS algorithms. A key feature of this article is that we introduce an artificially mixed dataset with a known mix of brain and artifact/noise signals. This dataset makes it possible to automate and standardize the source identification/classification process and to eliminate experimenter bias. Furthermore, this presents the opportunity to quantify the performance of each algorithm objectively so that they can be directly compared. Finally, the robustness of the algorithms with varying degree of contamination is assessed.

METHOD

EEG Acquisition

Two EEG records were acquired and used in the construction of the artificially mixed dataset. The first record was a 40-channel EEG (NuAmps; Compumedics; Melbourne; 1,000 samples per second, 200 Hz low-pass filter) acquired from a single male participant while performing a series of computer-directed tasks. Thirty-two electrodes were positioned according to the international 10 to 20 montage (these will be referred to as scalp channels) and the remaining eight electrodes were positioned at sites highly susceptible to artifact contamination (these will be referred to as artifact channels). The artifact channels were located on the left and right cheeks, left and right neck, left and right outer canthus (horizontal EOG), and 1.5 cm below the left and right eyes (vertical EOG). During acquisition, the participant performed seven computer-directed tasks designed to elicit typical EEG artifact, specifically horizontal and vertical saccadic eye movements, horizontal and vertical tracking eye movements, jaw and neck muscle movements, and voluntary blinking. This EEG record was used to provide contaminant signals.

The second EEG record was a 128-channel EEG record (Synamps, Compumedics; 5,000 samples per second, 1,250 Hz low-pass filter) acquired from a single male participant (different participant from the first EEG record) while performing mental serial subtraction. These data were acquired as part of a larger study. A neuromuscular blocking agent was administered after obtaining written informed consent to completely paralyze the participant and eliminate muscle artifact from all measurements. The 32 channels that corresponded to the scalp channels in the first EEG record were extracted and used in the current study. Both EEG records were rereferenced to a computed average reference, de-meaned, and the sample rates appropriately resampled to 1,000 samples per second with a 200 Hz low-pass filter. This second EEG record was used to provide brain signals.

Artificially Mixed Dataset

The key feature of the dataset used here is that it was constructed with a known mixture of known brain and con-
taminating signals (artifact and noise). Real contaminated EEG data are inadequate because the original signal mixture is unknown, and without this knowledge it is difficult to assess the quality of unmixing. Realistic head models are a potential solution, but there is still the problem of determining appropriate source locations and time courses. Jung et al. (1998b) used electrocorticographic recordings as brain sources and projected them through a three-shell spherical head model to create simulated scalp EEG. However, this approach does not account for artifact sources and their corresponding time course. Therefore, we constructed an artificially mixed dataset that consisted of a known mixture of known brain and contaminant signals while retaining, as closely as possible, the character of EEG.

The artificially mixed dataset was composed of two different signal types: brain and contaminant. The brain signal was actual scalp EEG from the second EEG record, which was acquired during paralysis and as such was completely free of any muscle, ocular, or movement contamination. Furthermore, this EEG was acquired in a shielded room to minimize noise contamination.

The contaminant signals were derived from the first EEG record, which was acquired during the seven computer-directed tasks. This involved identifying the artifact channels most susceptible to contamination for each artifact type and then regressing those channels against all other EEG channels to determine the distribution of each artifact across the scalp. The selected electrodes were: left and right horizontal EOG for horizontal saccadic and tracking artifact; left and right lower vertical EOG for vertical saccadic and tracking and blink artifact; left and right cheek for jaw muscle; left and right neck for neck muscle. These electrodes were selected because they are sufficiently distant from the brain and so are likely to contain little brain signal relative to the artifact. The regression coefficients were then used as weights to calculate the projection of these (predominantly) artifact channels to all scalp sites. Finally, a small amount of noise was added to each set (−60 dB power relative to the brain signals) to simulate measurement noise that would occur with real EEG. The result was seven sets of simulated artifact signals with a realistic scalp distribution. The artifact channels were stripped from each of the contaminant sets. Gaussian white noise was also used as an additional set of contaminant signals.

Given the assumption that scalp EEG is the linear sum of source signals propagated to the scalp, signal mixing in the artificially mixed dataset was achieved by linearly adding each of the contaminant signals to the brain signal. The advantage of this approach is that knowledge of source activity is not required because each of the signal sets (brain and contaminant) is a scalp estimate of multiple sources. The power ratio between brain and contaminant signals was varied between −40 and +40 dB in 2 dB increments. This manipulation permitted the investigation of the robustness of BSS performance with varying degree of contamination.

**BSS Algorithms**

Five popular BSS algorithms were used in the current experiment. The algorithms came from a variety of sources but all were available as open source code and were applied using default parameters recommended by their respective authors. Two were second order (PCA and SOBI) and the remaining were variations of ICA (Infomax, JadeR, and FastICA). Infomax and JadeR were obtained from the EEGLAB toolbox (Delorme and Makeig, 2004), SOBI was obtained from the ICALAB toolbox (http://www. bsp. brain. riken. go. jp/ICALAB/), and FastICA was obtained from the FastICA package for Matlab (http://www. cis. hut. fi/projects/ica/fastica/). PCA was calculated using inbuilt Matlab functions. All analyses were performed in Matlab on a Linux workstation.

**Source Classification**

As mentioned earlier, the correct classification of sources is critical to ensure good-quality artifact/noise reduction. This process is often performed by visual inspection and is susceptible to bias. To eliminate such biases, it is important to automate and standardize the process. This was possible with the artificially mixed dataset due to the known mixture of signals, a constraint that does not hold with real EEG.

The method involves taking the artificially mixed dataset (BC), which is the sum of the brain signals (B) and a set of contaminant (C) signals, and using a BSS algorithm to calculate the mixing matrix (M) such that:

\[ BC = M \cdot S_{bc} \]

where \( S_{bc} \) is the unmixed sources of BC. The mixing matrix (M) is a set of weights that detail the contribution of each separated source in source space to each EEG channel in sensor space. The sources of the artificially mixed dataset (\( S_{bc} \)) are calculated by the product of \( M^{-1} \) and BC:

\[ S_{bc} = M^{-1} \cdot BC \]

To determine whether a source is brain or contaminant we calculate the contributions of brain and contaminant to the source, they are \( S_{b} = M^{-1} \cdot B \) and \( S_{c} = M^{-1} \cdot C \) respectively. The covariance of these contributions with the source (i.e., \( S_{b} \cdot S_{bc} \) and \( S_{c} \cdot S_{bc} \)) gives estimates of the relative strength of the contributions. The log ratio of these covariances can then be used to classify the source. A positive ratio indicates that the source is predominantly brain, whereas a negative ratio indicates that the source is predominantly contaminant.

**Performance Metric**

The metric used to assess the performance of a BSS algorithm at separating brain signal from contaminant signal takes advantage of the fact these signals are known in the artificially mixed dataset. Once identified, the brain sources can be projected back into sensor space recovering an estimate (B’) of the original brain signal (B). Similarly, the contaminant sources can be projected back into sensor to recover estimates (C’) of the original contaminant signals (C). The original signals (B and C) were correlated with the recovered signals (B’ and C’) resulting a 2×2 matrix of correlation coefficients:

\[
\begin{bmatrix}
R(B, B') & R(B, C') \\
R(C, B') & R(C, C')
\end{bmatrix}
\]
The diagonals of the correlation matrix are the correlation of the original signals with their recovered counterparts while the off-diagonals are the cross-correlation of the original signals and the recovered signals. Therefore, the ideal correlation matrix is:

\[
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\]

where the original brain correlates perfectly with recovered brain, the original contaminant correlates perfectly with recovered contaminant, and there is no cross-correlation between original brain and recovered artifact or original artifact and recovered brain. In practice, this ideal classification is not actually possible due to a very small cross-correlation between original brain and original contaminant. Nonetheless, the correlation matrix provides important information about the classification performance:

\[
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix} = \text{Complete classification where brain and contaminant have been perfectly separated and correctly classified.}
\]

\[
\begin{bmatrix}
0 & 1 \\
1 & 0
\end{bmatrix} = \text{Complete misclassification where brain and contaminant have been perfectly separated but incorrectly classified.}
\]

\[
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix} \text{ or } \begin{bmatrix}
0 & 1 \\
1 & 0
\end{bmatrix} = \text{Biased classification where all sources (both brain and contaminant) have been classified as a single type.}
\]

\[
\begin{bmatrix}
0.5 & 0.5 \\
0.5 & 0.5
\end{bmatrix} = \text{Poor separation of brain and contaminant.}
\]

To simplify performance comparisons, it is desirable to reduce the correlation matrix to a scalar measure of performance. The determinant of the matrix provides an appropriate measure: complete classification = 1; complete misclassification = -1; biased separation = poor separation = chance separation = 0. This is closely related to informedness (Powers, 2003), the magnitude of which represents the probability that an informed separation and classification (or misclassification) has been made. Optimizing informedness is equivalent to optimizing ROC AUC (receiver operating characteristics area under the curve).

**RESULTS**

Figure 1 presents the results of the performance metric for all five BSS algorithms for each contamination type over a brain-to-contamination ratio (BCR) range of -40 dB to +30 dB. BCR is a power ratio equivalent to SNR, but we distinguish biologic artifact signals and noise signals as types of EEG contamination. Each subsection of this results section examines the best performing algorithms, according to the performance metric, per contaminant type at the BCR where performance is best. For each examination, correlations between original and recovered brain and contaminant signals are presented as well as EEG traces for visual inspection and spectral plots. All spectra were calculated using Welch’s short-time Fourier transform with 1,000-point Hamming windowing. In addition, the improvement in BCR from original to recovered brain signal is presented.

**Horizontal and Vertical Saccade**

At a BCR of 0 dB and below (i.e., when amplitude of contaminant is greater than that of brain), the best BSS algorithm for separating both horizontal and vertical saccadic eye movements from brain was PCA (see Fig 1). In the negative BCR range, the performance of PCA was good with performance increasing with BCR negativity until a plateau at approximately -20 dB. At an extreme BCR of -40 dB (i.e., amplitude of contaminant is 100 times greater than brain), the brain signal recovered by PCA correlated strongly with the original brain signal (r = 0.89 for horizontal saccade, r = 0.80 for vertical saccade), and the recovered contaminant signal correlated strongly with the original contaminant signal (r = 1.00 for horizontal and vertical saccade). Power spectral analysis (Fig. 2) revealed that there was little difference in spectral power between the original and recovered contaminant signals, whereas the recovered brain signal had greater broadband power particularly at higher frequencies suggesting that there was some residual contamination of the recovered brain signal. The new BCRs following removal of the recovered contaminant were 6.77 dB and 4.44 dB for horizontal and vertical saccade contaminants, respectively, which were a substantial improvement on the original BCR of -40 dB (see Table 1).

PCA also performs best in the positive BCR range of 8 to 10 dB, albeit poorly compared with the negative BCR range; however, it failed at 2 to 4 dB (i.e., when the amplitudes of brain and contaminant are similar) and above 10 dB (i.e., when the amplitude of the brain is much larger than the contaminant). At a BCR of 8 dB, the brain signal recovered by PCA correlated well with the original brain signal (r = 0.94 for horizontal saccade, r = 0.93 for vertical saccade); however, the correlation between recovered and original contaminant signals was less strong (r = 0.62 for horizontal saccade, 0.48 for vertical saccade). Power spectral analysis (Fig. 3) revealed that the recovered brain had greater power at higher frequencies than the original brain signal, whereas the recovered contaminant signal had greater power at lower frequencies and less power at higher frequencies than the original contaminant signal. Therefore, at this BCR the brain signal recovered by PCA retains too much high frequency contamination and loses some low frequency signal. The new BCRs following removal of the recovered contaminant were 9.13 dB and 8.06 dB for horizontal and vertical saccade contaminants, respectively, which are only small improvements over the original BCR of 8 dB (see Table 1).

In the positive BCR ranges where PCA fails, the best algorithm was a tie among equally unimpressive performances by SOBI, FastICA, and JadeR; however, SOBI had a broader BCR range of performance than the other two. At a BCR of 8 dB, residual contamination in the form of the characteristic discontinuities of saccadic eye movements could be clearly observed in the brain signal recovered by...
SOBI, FastICA, and JadeR (not shown). The naturally occurring BCRs of the saccadic contaminant as measured in this experiment were 0.90 dB and 5.55 dB for horizontal and vertical saccadic movements, respectively (Table 2). At these levels, PCA outperforms the other BSS algorithms at separating horizontal saccadic contamination; however, for separation of vertical saccadic contamination all of the BSS algorithms are equally unimpressive.

**Horizontal and Vertical Tracking**

The best BSS algorithm for separating both horizontal and vertical tracking eye movements from brain was PCA when the BCR was less than 0 dB and SOBI when greater than 0 dB. PCA performed well in the negative BCR range, whereas the performance of SOBI in the positive BCR range was less impressive. The naturally occurring BCRs of the contaminant as measured in this experiment were 4.22 dB for horizontal tracking and 1.18 dB for vertical tracking movements. At these levels there is little difference in the performance metric between PCA and SOBI.

At a BCR of $-40$ dB the brain signal recovered by PCA correlated strongly with the original brain signal ($r = 0.89$ for horizontal tracking, $r = 0.80$ for vertical tracking), and the recovered contaminant signal correlated strongly with the original contaminant signal ($r = 1.00$ for both horizontal and vertical tracking). Similar to saccadic contamination, the power spectra of tracking contamination indicated that there was little difference in spectral power between the original and recovered tracking contaminant signals, while the recovered brain signal had greater broadband power than the original brain signal, particularly at higher frequencies, which indicates a small amount of residual contamination in the recovered brain signal (not shown). Removal of the recovered contaminant signal improved the BCRs to 6.76 dB and 4.34 dB for horizontal and vertical tracking contaminants, respectively (Table 1).

At a BCR of 8 dB, the brain signal recovered by SOBI correlated strongly with the original brain signal ($r = 0.94$ for horizontal tracking, $r = 0.95$ for vertical tracking); however, the recovered contaminant signals had only a weak-to-moderate correlation with the original contaminant signals ($r = 0.48$ for horizontal tracking, $r = 0.28$ for vertical tracking).

Figure 4 shows that the poor correlation between recovered...
and original contaminant signals is due to a failure to recover the slow-wave signal that is characteristic of tracking eye movements. As such this signal is still embedded in the recovered brain signal, although it is difficult to identify visually because the original BCR was 8 dB.

Jaw and Neck Muscle

PCA performs best at separating jaw and neck muscle contaminants from brain signal when the BCR is below 0 dB, but all of the other BSS algorithms perform well when the BCR was above 0 dB. At a BCR of −40 dB, the brain signal recovered by PCA correlated strongly with the original brain signal (r = 0.91 for jaw muscle, r = 0.89 for neck muscle) and the contaminant signal recovered by PCA correlated strongly with the original contaminant signal (r = 1.00 for both neck and jaw muscle). The power spectra (Fig. 5) indicated that there was little difference in spectral power between the original and recovered muscle contaminant signals, whereas the recovered brain signal had greater power at high frequencies than the original brain signal. This is reflective of some residual muscle contamination in the recovered brain signal, which can be visually observed. Nonetheless the BCRs of the recovered brain signal improved to 7.81 dB for

### TABLE 1. Original Brain-to-Contamination Ratio (BCR) and Recovered BCR of the Best-Performing Algorithms for Each Contaminant Type

<table>
<thead>
<tr>
<th>Contaminant</th>
<th>Algorithm</th>
<th>Original BCR (dB)</th>
<th>Recovered BCR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal saccade</td>
<td>PCA</td>
<td>−40</td>
<td>6.77</td>
</tr>
<tr>
<td></td>
<td>PCA</td>
<td>8</td>
<td>9.13</td>
</tr>
<tr>
<td>Vertical saccade</td>
<td>PCA</td>
<td>−40</td>
<td>4.44</td>
</tr>
<tr>
<td></td>
<td>PCA</td>
<td>8</td>
<td>8.06</td>
</tr>
<tr>
<td>Horizontal tracking</td>
<td>PCA</td>
<td>−40</td>
<td>6.76</td>
</tr>
<tr>
<td></td>
<td>SOBI</td>
<td>8</td>
<td>9.16</td>
</tr>
<tr>
<td>Vertical tracking</td>
<td>PCA</td>
<td>−40</td>
<td>4.34</td>
</tr>
<tr>
<td></td>
<td>SOBI</td>
<td>8</td>
<td>8.36</td>
</tr>
<tr>
<td>Jaw muscle</td>
<td>PCA</td>
<td>−40</td>
<td>7.81</td>
</tr>
<tr>
<td></td>
<td>Infomax</td>
<td>8</td>
<td>11.70</td>
</tr>
<tr>
<td>Neck muscle</td>
<td>PCA</td>
<td>−40</td>
<td>6.82</td>
</tr>
<tr>
<td></td>
<td>Infomax</td>
<td>8</td>
<td>10.42</td>
</tr>
<tr>
<td>Noise</td>
<td>PCA</td>
<td>0</td>
<td>5.62</td>
</tr>
<tr>
<td></td>
<td>SOBI</td>
<td>0</td>
<td>5.62</td>
</tr>
<tr>
<td>Blinks</td>
<td>PCA</td>
<td>−40</td>
<td>7.34</td>
</tr>
<tr>
<td></td>
<td>Infomax</td>
<td>6</td>
<td>12.60</td>
</tr>
</tbody>
</table>
jaw muscle contamination and 6.82 dB for neck muscle (Table 1), both improvements of greater than 40 dB.

Within the positive BCR range, the best performance by an extremely narrow margin was Infomax at a BCR of 8 dB. At a BCR of 8 dB, the recovered Infomax brain signal correlated strongly with the original brain signal (r = 0.97 for jaw muscle, r = 0.96 for neck muscle) and the recovered contaminant signal correlated moderately strongly with the original contaminant signal (r = 0.76 for jaw muscle, r = 0.65 for neck muscle). The power spectra (Fig 6) indicated that there was little difference in the spectral power between the recovered and original contaminant signals, while the recovered brain signal had greater power at high frequencies than the original brain signal. The BCRs of the recovered brain signal were 11.7 dB for jaw muscle and 10.42 dB for neck muscle (Table 1). The naturally occurring BCR of the muscle contaminant as measured in this experiment was 10.35 dB and 0.08 dB for jaw and neck muscle, respectively. At these levels PCA far outperforms the other algorithms for jaw muscle and is only slightly better than the other algorithms for neck muscle.

**TABLE 2.** RMS and Brain-to-Contaminant Ratio Averaged Across All Electrode Sites for Each Contaminant Type as Measured Naturally

<table>
<thead>
<tr>
<th>Signal Type</th>
<th>RMS (μV)</th>
<th>BCR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain</td>
<td>10.98</td>
<td>0.00</td>
</tr>
<tr>
<td>Horizontal tracking</td>
<td>6.75</td>
<td>4.22</td>
</tr>
<tr>
<td>Vertical tracking</td>
<td>9.59</td>
<td>1.18</td>
</tr>
<tr>
<td>Horizontal saccade</td>
<td>9.90</td>
<td>0.90</td>
</tr>
<tr>
<td>Vertical saccade</td>
<td>5.79</td>
<td>5.55</td>
</tr>
<tr>
<td>Jaw muscle</td>
<td>36.16</td>
<td>-10.35</td>
</tr>
<tr>
<td>Neck muscle</td>
<td>11.09</td>
<td>-0.08</td>
</tr>
<tr>
<td>Blinks</td>
<td>7.57</td>
<td>3.23</td>
</tr>
</tbody>
</table>

**Blanks**

For a negative BCR the best-performing algorithm for separation of blinks from brain signal was PCA, whereas for positive BCR all of the other algorithms performed well. At a negative BCR as extreme as -40 dB, the performance of PCA was good with a strong correlation between recovered and original brain signals (r = 0.90) and between recovered and original blink contaminant signals (r = 1.00). The power spectra (Fig. 7) demonstrated little difference between the
recovered and original artifact signals, and a reduction in low frequency power and an increase in high-frequency power in the recovered brain signal relative to the original. Thus, some low-frequency brain signal has been lost and some high-frequency contaminant retained. Despite this, the BCR of the recovered brain signal improved to 7.34 dB (Table 1), an improvement of over 40 dB. In the positive BCR range the best-performing BSS algorithm by a very narrow margin was Infomax at a BCR of 6 dB at which there was a strong correlation between recovered and original brain signals (r = 0.97) as well as recovered and original contaminant signals (r = 0.88). The power spectra (Fig. 7) indicated that there was an increase in low-frequency power in the recovered contaminant signal relative to the original and an increase in high-frequency power in the recovered brain signal relative to original. Thus, the recovered brain signal retained some high-frequency contaminant and lost some low-frequency signal (not shown). Nonetheless, the BCR improved to 12.60 dB (Table 1). The naturally occurring BCR of blink contaminant as measured in this experiment was 3.23 dB. At this level, all of the algorithms, with the exception of PCA, perform well.

Noise

All of the BSS algorithms performed best when the BCR for noise contamination was around 0 dB, with performance tapering off for both increasing positive and negative BCR. At a BCR of 0 dB the best-performing algorithms were PCA and SOBI, which were identical in their strong correlation between recovered and original brain signals (r = 0.87) as well as recovered and original contaminant signals (r = 0.86). The power spectra for both algorithms indicated that there was an increase in low-frequency power in the recovered contaminant signal relative to the original and an increase in high-frequency power in the recovered brain signal relative to original. Thus, the recovered brain signal retained some high-frequency contaminant and lost some low-frequency signal (not shown). Nonetheless, the BCR improved to 5.62 dB for both algorithms (Table 1).

DISCUSSION

The present study sought to investigate and compare the relative performance of several popular BSS algorithms at separating common types of EEG artifact. The key finding is that BSS is an effective and powerful tool for separating and removing contamination from EEG; however, the quality of the separation is highly dependant on the type of contamination, the degree of contamination, and the choice of BSS algorithm.
For the biologic contaminants investigated (i.e., saccadic and tracking eye movements, head muscles and blinks), a clear pattern emerged where the PCA algorithm performed best when the amplitude of the contamination was greater than that of the brain whereas one or several of the other algorithms performed best when the amplitude of the contamination was similar or less than that of the brain. The exception to this was saccadic eye movements, where PCA was also the best performer when the amplitude of the contamination was smaller than the brain. For white noise contamination, PCA and SOBI were virtually identical in their performance, which peaked when the amplitudes of brain and contamination were similar and tapered off as the difference in amplitudes increased.

PCA exhibited a very consistent pattern, across all the biologic contaminants, of performing very well when the contamination was much larger in amplitude than the brain (i.e., severe contamination) and less impressively when the contamination was smaller in amplitude than the brain. This was a robust finding that occurred with all of the biologic contaminants, which were independent datasets although they were mixed with the common brain dataset. At a BCR ratio of just above 0 dB (i.e., when the amplitude of contaminant is just smaller than that of brain) PCA performance dips and usually fails completely. The BCR was calculated across all electrode sites, however a biologic contaminant does not typically affect all sites and if the BCR is calculated on only the subset of electrodes that are strongly affected by the contamination it is much closer to 0 dB. Therefore, PCA is failing to separate the contaminant when its amplitude is the same as the amplitude of the brain in the electrodes where the contamination occurs. This finding is not surprising given that PCA is a second-order algorithm and is therefore strongly influenced by amplitude. Furthermore, this phenomenon has been reported elsewhere (Lagerlund et al., 1997).

PCA did not perform as well when the contaminant was smaller in amplitude than the brain, which suggests that PCA is best able to characterize the contaminant signal and struggles when this signal is weak. Nonetheless, the most striking observation is how well PCA performs when the contamination is severe. At a BCR as extreme at just above 0 dB (i.e., amplitude of artifact is 100 times greater than brain), PCA can recover a good estimate of the brain signal and improve BCR by more than 40 dB. The other algorithms tested were unable to successfully separate contamination from brain at such high levels of contamination. This is a particularly
important finding for muscle contamination (especially jaw), which naturally falls into this range where PCA is most effective. However, spectral analyses demonstrated that PCA tended to leave residual contamination at high (>30 Hz) frequencies. Nonetheless, the reduction of contamination was still impressive. PCA is often overlooked in favor of higher-order algorithms; however, the findings of the current study clearly indicate that algorithm selection needs to be guided by a good understanding of the type and degree of contamination.

The other (non-PCA) BSS algorithms also exhibited a characteristic pattern of performance where peak performance was achieved when the amplitude of the contamination was less than that of the brain, with performance decreasing as contamination increased. This pattern was somewhat complementary to PCA, with PCA performing best at severe contamination and one or several of the other algorithms usually best at lower levels of contamination. Specifically, the best-performing algorithms at lower levels of contamination were SOBI for tracking contamination and Infomax for muscle and blink contamination. PCA was still best at low levels of saccadic contamination for a narrow BCR window of 8 to 10 dB, outside of which SOBI, JadeR, and FastICA performed similarly, albeit poor.

In general, the performance of the non-PCA algorithms was poor on saccadic contamination, marginally better on tracking contamination, and best on muscle and blink contamination. The poor performance at separating brain signal from saccadic and tracking contamination was surprising, especially given the good performance on blink contamination, and therefore likely reflects different higher-order statistical properties of these contaminants. Our experience through visual inspection of BSS sources, particularly SOBI, is that tracking and saccadic sources can be clearly observed. However, before this study there had been no measure of the amount of brain signal contained within the sources that visually appear to be contaminant. In the current study, BSS sources that contained the characteristic discontinuities of saccadic contamination and slow-wave of tracking contamination also contained some proportion of brain signal and therefore could not always be safely classified as contaminant. As such, the recovered brain signal retained some of this contamination. Nonetheless, even when performing poorly there is still a small reduction in contamination as a result of
the application of these algorithms. It should be noted that ocular electrodes were deliberately excluded in the current analyses, but may improve separation of these ocular contaminants if they are included in the analysis. The high level of performance on separating brain signal from muscle and blink contamination was observed for all of the non-PCA algorithms, with Infomax outperforming the others by a very narrow margin. The subsequent improvement in BCR after recovery of brain signal using Infomax was small because the original degree of contamination was small; however, spectral analyses revealed that Infomax tends to leave less residual high frequency contamination than PCA.

As can be seen, algorithm selection is essential for good separation of brain and contaminant and it is critically dependent on the type and degree of contamination. For highly contaminated EEG (when the BCR is less than 0 dB), PCA appears to be the best choice of algorithm for saccadic contamination, SOBI for tracking contamination, and Infomax, JadeR, or FastICA for muscle and blink contamination. PCA or SOBI is the best choice of algorithm for white noise contamination at all degrees of contamination although separation fails for all algorithms when the BCR is less than −10 dB. This is not surprising, given that the other algorithms utilize higher-order statistics that are nominally zero for white noise. Hence, these algorithms are likely to overmodel the particular noise signals used in the experiment, to the detriment of their ability to model the brain signal.

Unfortunately, determining the original degree of contamination in real EEG is not a simple task. The use of the artificially mixed dataset in the current study permits the calculation of original BCR as actually measured (Table 2). However, these artifact amplitudes are likely to be inflated because the participant in this study was engaged in tasks specifically designed to elicit contamination. Nonetheless, they may provide some guidance for algorithm selection.

A limitation inherent within the current study is that it uses a simulation that, like all simulations, is simpler than the

FIGURE 7. Example of PCA (I) and Infomax (II) separation of brain from blink contamination at a brain-to-contaminant ratio of −40 dB and 6 dB respectively. The left plots in each major box are brain (B), recovered brain (B’), contaminant (C), and recovered contaminant (C’) EEG traces from a representative electrode (Fz). The upper right plot in each box is the power spectrum (dB relative to 1 V2/Hz) of brain (black) and recovered brain (gray), and the lower right plot in each box is the power spectrum of contaminant (black) and recovered contaminant (gray). In some of the spectral plots it is difficult to visualize the black line because the gray line directly overlays it.
real world and may not generalize well to real data. This has not been investigated and is beyond the scope of this study. It would not be possible to investigate and compare these algorithms without the use of a dataset of known mixture of known signals, which is not possible with real EEG. Furthermore it would be preferable to investigate sources of contamination proximal to the scalp recording sites (i.e., temporalis and frontalis muscle contamination); however, this was not possible with the framework presented, which by design required contamination signals sufficiently distant from the brain such that they contained little, if any, brain signal. Nonetheless, this study developed an artificially mixed dataset that was designed to be as realistic as possible and subsequently provided a novel framework in which BSS algorithms could be compared both empirically and objectively. However, when interpreting the results due consideration should be given to the fact that it is a simulation. An additional limitation is that each of the algorithms was applied using default parameters as supplied by the algorithm authors. Most of these algorithms have many parameters that can be modified to optimize their performance. However, such optimizations are difficult without a reference metric by which to measure the change in performance. This study defines a suitable reference metric that will permit future work to focus on optimizing the algorithms. It is probable that performance improvements on saccadic and tracking contaminants will be achievable through algorithm optimization, but care must be taken to use independent training and validation data.

In summary, the current article develops a novel framework for investigating and comparing the relative performance of BSS algorithms for separating common types of EEG contaminants. The framework incorporates a realistic EEG simulation with a known mixture of known signals that permits the calculation of an empirical performance metric. The key finding is that BSS is an effective and powerful tool for separating and removing contamination from EEG; however, the quality of the separation is highly dependant on the type of contamination, the degree of contamination, and the choice of BSS algorithm.

**ACKNOWLEDGMENT**

The paralyzed EEG data were contributed by the EEG Research Unit, School of Medicine, Flinders University, which was entirely funded by an equipment grant from The Wellcome Trust, London, UK.

**REFERENCES**


AUTHOR PLEASE ANSWER ALL QUERIES

AQ1: AUTHOR—Please spell out each author’s first name.