Title: Validation of ICA-Based Myogenic Artifact Correction for Scalp and Source-Localized EEG

Abstract: Muscle electrical activity, or "electromyogenic" (EMG) artifact, poses a serious threat to the validity of electroencephalography (EEG) investigations in the frequency-domain. EMG is sensitive to a variety of psychological processes and can mask genuine effects or masquerade as legitimate neurogenic effects across the scalp in frequencies at least as low as the alpha band (8-13Hz). Although several techniques for correcting myogenic activity have been described, most are subjected to only limited validation attempts. Attempts to gauge the impact of EMG correction on intracerebral source models (source "localization" analyses) are rarer still. Accordingly, we assessed the sensitivity and specificity of one prominent correction tool, independent component analysis (ICA), on the scalp and in the source-space using high-density EEG. Data were collected from 17 participants while neurogenic and myogenic activity was independently varied. Several protocols for classifying and discarding components classified as myogenic and non-myogenic artifact (e.g., ocular) were systematically assessed, leading to the exclusion of one-third to as much as three-quarters of the variance in the EEG. Some, but not all, of these protocols showed adequate performance on the scalp. Indeed, performance was superior to previously validated regression-based techniques. Nevertheless, ICA-based EMG correction exhibited low validity in the intracerebral source-space, likely owing to incomplete separation of neurogenic from myogenic sources. Taken with prior work, this indicates that EMG artifact can substantially distort estimates of intracerebral spectral activity. Neither regression- nor ICA-based EMG correction techniques provide complete safeguards against such distortions. In light of these results, several practical suggestions and recommendations are made for intelligently using ICA to minimize EMG and other common artifacts.
Dear Dr. Fletcher,

Please find attached the revised manuscript “Validation of ICA-Based Myogenic Artifact Correction for Scalp and Source-Localized EEG” (B.W. McMenamin et al.).

We were excited to read that the reviewers found this to be “a very interesting, solid paper” (Reviewer #1), that Reviewer #1 “like[d] the general design of the validation study, and the focus on sensitivity and specificity,” and that Reviewer #2 thought that it was “an exceptionally well written paper on a topic of considerable and underappreciated importance…[that] has several virtues…including] a thoughtful and systematic comparison of 2 aspects of EMG artifact correction methodology in a 9-cell factorial design, clear explanations of the issues addressed and the methods undertaken, and useful advice for EEG researchers, especially those interested in source analysis. Repeatedly I was impressed with the care with which the authors identified difficult procedural issues in the course of the work reported.”

In the revision, we have addressed each of the residual concerns outlined by the reviewers. These revisions are detailed in the attachment. By and large, we found the concerns to be thoughtful and constructive. Consequently, we believe that the revised manuscript is markedly stronger and clearer than the original submission. As such, we believe that it will be of considerable interest to both practitioners and consumers of EEG research.

We hope that you share our enthusiasm.

Sincerely,

Richard J. Davidson
Detailed Response to Reviewers

We generally found the reviewers’ concerns to be constructive and helpful. We appreciate the time each of them has invested in this process.

Our response is organized into two sections. In the first section, we describe how we addressed a general issue that cropped up across reviewers. In the second section, we respond to each of the more specific concerns in the order in which they appeared in the three reviews.

General Issue: Detailed Results vs. Summary

A general issue was the tension between reporting the detailed results of our comprehensive analyses—at the potential expense of overwhelming the reader and eroding the SNR of our message— versus providing a streamlined, clear summary—at the potential expense of the archival record. The strategy adopted in the original submission for dealing with this dichotomy, which relied on moving some results to the Supplementary Results file while increasing the amount of summary material in the main report, did not appear to be entirely successful at solving this conundrum.

Concerns about the amount of detail remained. As Reviewer #1 noted, “many analyses are reported, and I was sometimes lost in details that appeared less relevant. Please carefully reconsider whether it is really informative to report all these contrasts and results.” This was echoed by Reviewer #2, “Pages 18-19 - These pages consist mostly of numerous, very short paragraphs that narrate information in figures and tables in more detail than is needed. This material could be largely eliminated, with the reader relying on the ”Summary” section now at the bottom of page 19. (In contrast, the somewhat parallel narrative that follows on pages 20-22 is warranted, in part because of the greater novelty and lesser familiarity of source-analysis implications of the present work.).”

Conversely, Reviewer #2 found the summary material to be redundant, given the amount of detail: “Page 23-24 - Paragraphs 2 through 6 of the Discussion can be virtually eliminated, as they do little more than repeat some of the results.”

In the revision, we addressed these concerns in several ways.

We condensed the results of the scalp validation tests to a single interim summary,

“Validity of ICA-Based EMG Correction

As summarized in Table 1, only four protocols showed questionable or better performance across all tests of sensitivity (Figures 2, 4, and 5), specificity (Figures 4, 5, and 8), and correction-induced artifact (Figures 2 and 8): the Minimal-EMG protocol paired with Minimal- or Intermediate-NNNM filtering and the Maximal-EMG protocol paired with Minimal- or Maximal-NNNM filtering. For detailed results, see Supplementary Tables 1-3. Moreover, inspection of Table 1 indicates that among these four, the Minimal-EMG/Intermediate-NNNM protocol invariably equaled or exceeded the performance of the Minimal-EMG/Minimal-NNNM protocol; likewise, the Maximal-EMG/Maximal–NNNM protocol always outperformed the Maximal-EMG/Minimal-NNNM protocol. Accordingly, these
two combinations of EMG correction and NNNM filtering were subjected to additional testing in the intracerebral source-space using LORETA.”

We also slightly condensed the results relevant to understanding the impact of uncorrected EMG on the scalp to minimize redundancy with the Discussion.

Consistent with Reviewer #2’s suggestion, only slight changes were made to the results of the source-space validation tests.

The first portion of the Discussion (original paragraphs 2-6) was re-organized and condensed to summarize and contextualize two key sets of results:

1. The marked impact of uncorrected EMG artifact on the scalp and in the source-space (revised paragraphs 2 and 3, respectively).
2. The variable (across protocols and individual participants) and striking impact of the different ICA protocols (revised paragraph 4).

We do not think it unreasonable to devote this amount of space (~1 manuscript page) to these findings, particularly given that they are (i) rarely discussed in prior EMG correction studies, (ii) likely to be of substantial interest to EEG researchers—including those with no interest in using ICA for EMG correction, and (iii) likely to help make a quite complex set of findings intelligible to the reader. As noted above, in keeping with the spirit of Reviewer #2’s comment, we have carefully re-worked the Results and this section of the Discussion to eliminate trivial redundancies.

Specific Concerns

Reviewer #1

R1.1. In my experience, ICA cannot easily disentangle EMG contributions from channel noise from gamma band activity, the three main sources of power in the 30-100 Hz range. ‘True’ EMG-ICs usually express (not only, but substantial) high frequency power, a tangential dipolar pattern AND an extracranial dipole source location. Channel-noise ICs on the other hand usually pick up activity from single channels, thus expressing a radial pattern (high spatial frequency), and gamma-band ICs, if they can be extracted at all, are characterized by a lower spatial frequency and much less of a dipolar spatial pattern. Which ICs did the authors consider muscle-ICs, in terms of dipole localization and orientation, and how did they deal with the other types?

As noted briefly in the Method section and described in detail in the Supplementary file (e.g., Supplementary Figures 2-13), ICs were classified on the basis of each component’s time-series, power spectrum, and scalp topography.

ICs were classified as EMG if they displayed clear-cut myogenic activity in the absence of any identifiable neurogenic activity (Supplementary Figure 9). These ICs were chiefly distinguished based on spectra with broad peaks around either 40Hz or greater than 70Hz. On the scalp, they showed one of two topographies: a moderately broad distribution that mimicked the underlying
scalp musculature and peaked along the edge of the montage, or small cluster(s) of cephalic or extracephalic electrodes. In the time-domain, they exhibited periods of high-frequency activation.

ICs were classified as Neurogenic (e.g., alpha, gamma) if they displayed clear-cut neurogenic activity in the absence of any artifactual activity (Supplementary Figures 7-8). These ICs were characterized by broad, smooth topographies, often with a clearly dipolar pattern, with peak loadings well away from the edge of the montage. In the frequency-domain, they exhibited a clear 1/f pattern, often with a peak in the alpha band (8-13Hz). In the time-domain, they displayed sustained periods of activation with low-frequency oscillations.

By default, components that individually accounted for <0.2% of the variance were categorized as Low-Variance. In cases where the determination was unambiguous, exceptions were made. The remaining ICs were classified as a combination of the two sources (Neuro-Dominant or Myo-Dominant), or artifact (residual Gross or Ocular). Components that met the minimum variance criterion, but proved impossible to unambiguously categorize were classified as Noise (Supplementary Figure 12). Noise components tended to show strongly 1/f–shaped spectra.

Dipole modeling (i.e., dipfit) was not used for classification purposes. This is consistent with nearly all prior methodological reports of ICA-based artifact correction (e.g., Delorme et al., 2007). To our knowledge, dipole-based IC-by-IC classification has only been employed in one prior report (Ohla et al., Brain Topography, 2009) where, unlike the present study, the timing of artifacts (electrogustometry stimulation) were under precise experimental control, hence permitting the computation of time-locked single-subject average waveforms (with reasonable SNR) for dipole source modeling (for a related conference presentation, see also Onton & Makeig, 2009).

R1.2. The conclusions made in the section 'Two factors could plausibly account for the inability of infomax ICA to fully separate myogenic from neurogenic sources... (p 27)' would be more convincing if ICs identified as representing muscle activity would have been analyzed in detail.

Unfortunately, it is not clear to us from this brief statement how a detailed analysis of the EMG components would further our argument.

We argued that the inadequate performance of ICA in the source-space likely reflects two factors. First, the fact that source modeling (e.g., LORETA) uses information from all electrodes in the array, not just those in the scalp ROIs, which means that inadequate correction outside of the ROIs (e.g., along the edge of the array) can exert a deleterious effect on source-space analyses. Second, we described 3 observations that suggest that infomax ICA failed to adequately separate EMG from EEG—an assertion in accord with this reviewer’s experience (“In my experience, ICA cannot easily disentangle EMG contributions from channel noise from gamma band activity”).

Taken together, we argued that our results indicate that, depending on the choice of which ICs to discard, ICA-based EMG correction under-corrects (inadequate sensitivity) or over-corrects (inadequate specificity) the data across the entire electrode array: “In particular, it [Maximal-
EMG correction paired with Maximal-NNNM filtering failed to adequately remove EMG when neural activity was fixed and overcorrected neurogenic activity when EMG was absent. Poor sensitivity was also evident for the other procedures examined in the source-space (i.e., Minimal-EMG/Intermediate-NNNM).”

Related to this reviewer’s comment, we can envision 4 more detailed analyses that could be performed on the (discarded) myogenic ICs: (1) mean topographies, (2) mean PSDs, and (3) distributed source modeling (LORETA solution), and (4) dipole models (dipfit/BESA solution; see e.g., Ohla et al., Brain Topog, 2009).

But it is not at all clear to us how any of these would strengthen (or weaken) our conclusion that ICA fails to adequately separate EEG from EMG (i.e., particular ICs contain a mixture of the two ‘sources’).

**R1.3.** It may also be worth keeping in mind that muscle ICs could provide important information on human behaviour and cognition (Makeig et al., 2009, Int J Psychophysiol).

We strongly agree (see also Lee, Shackman, Jackson & Davidson, 2009; Onton & Makeig, 2009; Shackman et al., 2006, 2009).

As noted in the Introduction, “EMG…is sensitive to numerous cognitive and affective processes, including cognitive load (Cohen, Davidson, Seulis, Saron & Weisman, 1992; Waterink & Van Boxtel, 1994), facial mimicry (Dimberg, Thunberg & Elmehed, 2000), and induced emotional states (Borden, Petersom & Jackson, 1991; Coan & Allen, 2003; Bradley, Codispoti, Cuthbert & Lang 2001)… Moreover, because EMG covaries with cognitive and affective processes of interest, rejecting data laden with EMG artifact would likely entail discarding some of the most interesting, discriminative periods of neural activity (Davidson, Ekman, Saron, Senulis & Friesen, 1990).”

**R1.4.** I disagree with statements suggesting that a major feature of EMG would be that it is detectable across the whole scalp (e.g., page 3). … The problem is rather that several spatially distinct (and largely independent!) groups of muscles can heavily distort the EMG, and, if strong enough, make EMG artifacts evident in many EEG channels (and expressing different topographies). The argument about less stereotypy is correct, but I would like the authors to consider that different muscle groups can be voluntarily (and spontaneously) activated, thus contributing different maps to the EEG, not one spatially distributed topography.

We largely agree with these statements about extent and stereotypy (see also Goncharova et al., McMenamin et al., 2009; Shackman et al., 2009). Apparent differences in opinion seem to largely reflect poor wording choices on our part, for which we apologize.

To address the issue of extent, we have modified the Introduction to read, “EMG can often be detected across the entire scalp (Goncharova, McFarland, Vaughan & Wolpaw, 2003) due to volume conduction of myogenic activity independently generated by muscles across the head, face and neck. Anterior electrodes are sensitive to facial muscles, such as the corrugator supercilii and frontalis; lateral electrodes are sensitive to the muscles of
mastication, masseter and temporalis; and posterior electrodes are sensitive to muscles at the intersection of the cranium, spine, and torso, such as occipitalis (Supplementary Figure 1).”

To address both issues, we have also revised it to read, “EMG also exhibits less stereotypy than other biological artifacts... EMG, however, arises from the activity of spatially distributed, functionally independent muscle groups, with distinct topographic and spectral signatures (Goncharova et al. 2003). For instance, frontalis activity peaks around 25Hz, whereas temporalis generates a low peak around 20 Hz and broad plateau centered around 40-80 Hz (Goncharova et al., 2003). The spectral composition of myogenic activity also varies as a function of contraction intensity (Goncharova et al., 2003) and fatigue (Chung, Kim & McCall, 2002). This is compounded by the fact that the relative contributions of each muscle group to the cranial EMG vary substantially across elicitors and individuals (Tassinary, Cacioppo & Vanman, 2007) and may differ somewhat between spontaneous and voluntary contractions (Davidson, Shackman & Maxwell, 2004; Morecraft & Tanji, 2009).”

We also revised the Introduction to read, “Peri-cranial muscle or myogenic activity is distinguished by its relatively high amplitude, broad spectral and often anatomical distributions, and exquisite sensitivity to a variety of psychologically interesting processes.”

R1.5. It would help to move some sections of the introduction into the discussion section (e.g., page 6, discussion of simulation and scripted approaches).

We agree. Accordingly, the Introduction was revised to read, “Although ICA shows much promise as a tool for correcting EMG and other kinds of biological artifact (e.g., Jung, Makeig, Humphries, Lee, McKeown, Iragui & Sejnowski, 2000), attempts to assess its validity have been limited. Many validation studies have relied on small samples of ad hoc data (Jung et al., 2000; Wallstrom, Kass, Miller, Cohn & Fox et al., 2004; Flexer, Bauer, Pripfl, & Dorffner, 2005; Ting, Fung, Chang, & Chan, 2006; Frank & Frishkoff, 2006). While others have used simulations (e.g., Crespo-Garcia, Atienza & Cantero, 2008; De Clercq, Vergult, Vanrumste, Van Hees, Palmini, Van Paesschen & Van Huffel, 2005; Delorme et al., 2007; Fitzgibbon, Powers, Pope & Clark, 2007; Frank & Frishkoff, 2006; Romero, Mananas & Barbanoj, 2008). In simulations, real or artificial EMG activity is mathematically “injected” into otherwise artifact-free EEG. The potential problem with this strategy is that the assumptions underlying injection (e.g., the degree of temporal and spatial correlation with neurogenic signals) may not characterize real EMG contamination, potentially limiting external validity and biasing the results in favor of correction techniques founded on similar assumptions (Grouiller, Vercueil, Krainik, Segebarth, Kahane & David, 2007; Hoffmann & Falkenstein, 2009).

Accordingly, the major aim of the present study was to quantitatively assess the quality of EMG artifact correction afforded by ICA. Ideally, validation would quantitatively establish that a technique possesses a high degree of sensitivity (i.e., attenuates myogenic artifact) and specificity (i.e., preserves neurogenic signals) in a reasonably large and varied dataset. This requires data in which the presence or absence of EMG (“ground truth”) is definitive or can be reasonably assumed. To this end, the dataset previously employed by McMenamin et al. (2009) for testing the validity of GLM-based correction techniques was reanalyzed using ICA. This had the advantage of facilitating direct comparisons across correction techniques...”
R1.6. I like the general design of the validation study, and the focus on sensitivity and specificity, but some aspects are less convincing. For instance, [1a] the neurogenic contrast was performed with a focus on the alpha band only. Also, [2a] the use of a distributed source model is not without risk, making quantifications for further analyses less straightforward than dipole models (although I agree that LORETA makes more sense for the analysis of alpha activity).

It might be the case that EMG-correction with ICA is better (or worse!) [1b] at other frequency ranges, and [2b] it might be the case that other source modeling approaches (linear beamforming) are more robust against poor un-mixing of signal and artifact.

Accordingly, the 'Recommendations and Conclusions' section should be phrased a little bit more carefully, and communicate the limitations of the study.

We agree. In order to address the concern about spectral specificity, we revised the Discussion to read, “Second, our conclusions derive from analyses of the alpha band (8-13Hz) during extended blocks of myogenic activity. While this represents a reasonable analog to studies using blocked manipulations of emotion (e.g., threat of shock, emotional films), the degree to which these conclusions generalize to event-related designs or other frequency bands is unclear. Still, it seems likely that the quality of performance for the neighboring theta band (4-8Hz) would be similar to that observed for alpha. In contrast, we anticipate that the quality of correction would be lower for bands, such as gamma (>30Hz), that lie closer to peak EMG activity. Indeed, exploratory analyses indicated poor performance for ICA in the 70-80Hz range (Footnote 5). Finally, the degree to which the present conclusions generalize to stronger or weaker EMG contamination is unknown (Fitzgibbon et al., 2007).”

To address the concern about specificity to LORETA, we revised the Discussion to read, “Consideration of these observations and the extant literature yields four recommendations. First, if widespread EMG artifacts are suspected, Maximal-EMG correction protocol with Maximal-NNNM filtering should be employed. This method exhibited the best combination of sensitivity and specificity across all tests (Table 1). It also outperformed GLM-based EMG correction, which never demonstrated excellent performance in any of our validation tests (Supplementary Tables 7-9). Second, given its merely adequate sensitivity and inconsistent specificity, we no longer recommend the use of GLM-based EMG correction techniques (cf. McMenamin et al., 2009) for studies characterized by widespread EMG artifact. Third, for investigations where specificity is a smaller concern than sensitivity, any of the Intermediate- or Maximal-EMG correction ICA-based protocols represent reasonable choices. Fourth, the use of distributed modeling techniques, such as LORETA, to estimate the intracerebral sources of spectral EEG in studies with prominent myogenic activity is not recommended. It remains to be seen whether it is reasonable to do so using other approaches, such as dipoles or beamformers (Michel, Murray, Lantz, Gonzalez, Spinelli & Grave de Peralta, 2004; (Michel, Murray, Lantz, Gonzalez, Spinelli & Grave de Peralta, 2004; Nazarpour, Wongsawat, Sanei, Chambers & Oraintara, 2008).”

R1.7. The abstract and introduction suggest that ICA is a 'most popular' tool for EMG correction. ... [the] 'most popular' is, of course, simple low pass filtering.
This concern reflects an ambiguous choice of wording on our part. By definition, low-pass filtering does not correct spectral activity in the pass-band (e.g., alpha); i.e., arguably is not really a form of correction at all, in our view.

We have revised the Abstract to read,
“Accordingly, we assessed the sensitivity and specificity of one prominent correction tool, independent component analysis (ICA), on the scalp and in the source-space using high-density EEG.”

We revised the Introduction to read,
“In particular, the sensitivity and specificity of one prominent electromyography (EMG) correction tool, independent component analysis (ICA), remains unclear.”

We revised the Discussion to read,
“In recent years, ICA has rapidly become a popular tool for correcting EMG artifact, despite limited work assessing its validity for this purpose.”

R1.8. the abstract incorrectly suggests that EMG contribution affect only frequencies down to the alpha band.

This was revised to read,
“EMG is sensitive to a variety of psychological processes and can mask genuine effects or masquerade as legitimate neurogenic effects across the scalp in frequencies at least as low as the alpha band (8-13Hz).”

R1.9. Language used in the ms is sometimes at odds with convention and needs revision. For instance, the statement 'despite the fact that ICA requires an enormous ICA investment in human resources' (page 8) is not only misleading, it is wrong. In my lab ICA is used regularly and efficiently, and the same holds in many colleagues working in other labs. ICA is not necessarily a big help in dimensionality reduction and point to the really important bits of EEG data. Maybe this is what the authors want to communicate?

We agree that ICA can be used efficiently and should be used regularly by EEG investigators.

But we respectfully disagree with the crux of this concern. First, to our knowledge there is as yet no well-validated automatic technique for algorithmically classifying EMG ICs derived from high-density montages. Thus, the burden of manual classification cannot yet be eliminated for BSS-based tools. Second, high-density EEG studies in our laboratory and others typically involve 30-300 participants (for a representative study, see e.g., Shackman et al., Psychological Science, in press; http://psyphz.psych.wisc.edu/~shackman/ajshackman_bis_AcceptedPsychSci2009a.pdf). With reduction of 128 channels to 64 PCs, this requires the classification of at least 1920 (30 × 64) components.
As noted in the Supplementary file, “Final classification was by consensus [across two trained raters]. Classification required approximately two hours per participant per rater [~2 min/IC] once the raters were calibrated to the protocol. In the present study, the time invested in calibration approximately equaled that devoted to “final” classifications.” Altogether, more than 100 man-hours were required for the present, rather modestly sized sample.

This could be reduced somewhat in applied studies by employing “pre-trained” personnel or dropping the requirement that each IC be classified by two raters (at the potential expense of reliability) as is typical of the field. (Note that adding a dipole-based criterion would increase this somewhat).

But let us be clear, manual IC classification is clearly a significant investment in time that is not mandated by other EMG correction techniques. For instance, the GLM-based techniques documented in our prior work (McMenamin et al., 2009) require no manual intervention and can be computed for an entire study in a matter of hours on a typical higher-end computer.

Nevertheless, we did not mean to over-emphasize this aspect of ICA.

Accordingly, we have revised the Introduction to read, “Furthermore, despite the fact that ICA requires trained raters to inspect hundreds or even thousands of components for a single high-density EEG study (number of components ≈ channels x participants; see Supplementary Method), the reliability of component classification has only rarely been reported (Viola, Thorne, Edmonds, Schneider, Eichele & Debener, 2009).”

R1.10. p. 4, 'fixed spatial or spectral filters or templates cannot be fruitfully applied...'. Please note that temporal ICA can be regarded as a fixed spatial filter approach! Each component is characterized by a spatial map (inverse weights), and the time course of an independent component is simply given by the multiplication of the weights with the raw data. Accordingly, flexibility exists only in the time domain, and activations express the ‘amount’ each spatial filter contributes to the mixed signal. Please revise accordingly.

We agree. This concern again reflects an unfortunate word choice on our part. By ‘fixed’ we simply meant an unvarying canonical template used across all participants, conditions, etc.

Accordingly, we have revised the Introduction to read, “Given marked individual differences in the spectral and anatomical profile of myogenic activity, spatial or spectral filters or templates that are fixed across subjects cannot be fruitfully applied to the correction of EMG artifact”

R1.11. it should be specified whether ICA was used in the 'extended' mode

Yes, it was. This has been corrected in the revision, “This was performed in a single step as part of the ICA using the EEGLAB runica command, implementing the extended Infomax algorithm (Bell & Sejnowski, 1995; Lee, Girolami & Sejnowski, 1999).”
R1.12. it should be specified ...whether the decompositions used for later analysis were reasonably reliable (see Groppe et al., 2009, Neuroimage, for discussion)

In a very recent report, Groppe, Makeig and Kutas (2009) propose a novel, computationally efficient method for quantifying the split-half reliability of ICs. Using this method, they demonstrated across 4 datasets that only a subset of ICs exhibit adequate reliability (on the order of ~30 ICs for de-meaned 64- and 30-channel datasets, broadly consistent with the Model Order estimates described in the Supplemental Results file). This is clearly an important and useful technique for winnowing those (neurogenic) ICs worthy of classification and analysis.

But it seems unlikely that this technique would systematically alter our conclusions concerning ICA-based EMG correction, given that they stemmed from a comprehensive examination of dropping different kinds of EMG and non-EMG/non-EEG ICs.

One of our key conclusions was that ICA does not fully separate EMG from EEG. Whether ICs containing a mixture of the two sources were deemed reliable (and hence retained for manual classification and analysis) or unreliable (and presumably dropped prior to classification) according to the threshold of Groppe would not change the findings supporting this conclusion.

(Note: The URL published in Groppe et al. for the matlab code is broken, so we were not able to examine it in detail [http://www.cogsci.ucsd.edu/~dgroppe/eeglab.html])

Nevertheless, from a purely practical standpoint, the method of Groppe might prove a useful means of objectively reducing the classification burden.

Accordingly, we have revised the Supplementary Methods to read,
“Classification required approximately two hours per participant per rater once the raters were calibrated to the protocol (see below). In the present study, the time invested in calibration approximately equaled that devoted to “final” classifications. Future studies might consider using the method of Groppe, Makeig and Kutas (2009) as a pre-processing step to reduce the number of components requiring classification to the subset evidencing adequate split-half reliability.”

R1.13. p. 25 'Prospects for ICA-Based EMG Correction': I am not sure whether the results of the present study are really more complex than the cited papers on ICA and EEG artifact. This section should simply present a more balanced view on what can, and what cannot be done with ICA. In contrast to what is stated, ICA can indeed be used to reduce the ballistocardiogram artifact - but only in combination with other tools, or at low B0 scanner field strengths (e.g., Debener et al., 2009, Int J Psychophysiol).

We agree that Debener’s research provides a nice parallel to the complexity of results yielded by the present study.

Accordingly, we have revised the Discussion to read.
“As noted in the Introduction, many studies have documented, with varying degrees of rigor, the utility of infomax ICA for attenuating various physical and biological artifacts. Nevertheless, several studies have found ICA to exhibit worse performance than alternative source separation
algorithms for ocular (Wallstrom et al., 2004; Romero et al., 2008) and EMG artifacts (Crespo-Garcia et al., 2008; Fitzgibbon et al., 2007). More recently, Debener et al. (2007) showed that ICA displays low specificity for ballistocardiogram artifacts, evidenced by attenuation of event-related neurogenic activity, under some circumstances (Debener, Mullinger, Niazy & Bowtell, 2008). Paralleling Debener and colleagues’ research, the present findings suggest that ICA is a valid means of correcting EMG in some, but not all, cases. On the scalp, some of the ICA-based protocols we tested displayed adequate sensitivity and specificity, whereas others did not. Furthermore, in the intracerebral source-space, even those protocols that showed the most promising performance on the scalp failed.”

R1.14. Also, ICA can be used reduce electrical artifacts that evoke(!) cortical responses (Sandmann et al., 2009, Brain; Debener et al., 2008, Psychophysiology; Ohla et al., 2009, Brain Topography). This latter finding seems particularly important to me, because it invalidates the common concern (as reflected in the present ms, or in textbooks such as Steve Luck’s ERP book) that ICA fails when signal and artifact covary. At present ICA seems to be the only approach that can at least partly cope with this scenario (for a validation of ICA for eye blink removal, see Hoffmann & Falkenstein, 2008, PloS One)...

None can be considered definitive tests. Sandmann et al. do not provide analyses of sensitivity or specificity. Debener et al. report a case study of 1 individual (compared to normative data). Ohla et al. found that ICA leaves residual artifact, and also do not provide quantitative tests of sensitivity or specificity.

Nevertheless, we agree with this reviewer that, collectively, these reports provide a kind of proof-of-principle. Accordingly, we have revised the Discussion to read,

“Infomax also assumes that sources are mutually temporally independent. To the degree that neurogenic and myogenic activity are too closely coupled in the time domain they would violate this assumption (but cf. Ohla, Hudry & le Coutre, 2009). Future studies could test the degree to which second-order blind source separation algorithms, that do not require strong assumptions, such as Second Order Blind Identification (SOBI) or Algorithm for Multiple Unknown Signals Extraction (AMUSE), produce better separation (Joyce, Gorodnitsky & Kutas, 2004; Romero et al., 2008; Tang, Liu & Sutherland, 2005).”

Please note that the Hoffmann and Falkenstein report was cited in the original (and revised) manuscript.

Reviewer #2
R2.1. The language describing the main issues is at times overly dramatic, such as "the grave inferential threat" on page 23.

We have revised this to read,

“Given the substantial inferential threat posed by EMG contamination, there is a pressing need for valid correction tools.”

We revised the Abstract to read,
“This inferential hazard is particularly serious in cases where neurogenic and myogenic activity covary.”

R2.2. Repeated citations of the same papers often do not use "et al." when appropriate.

This was corrected in the revision.

R2.3. Page 4, "Generally, EEG artifacts can be addressed in one of two ways, rejecting contaminated epochs of data or filtering artifact from neurogenic activity." This is a valid and an important point ... [that has] ... already been made, and confirmed empirically, for EOG artifact in ERP research (a paper by Robert Simons in the late 1980s, I believe, probably published in Psychophysiology).

Unfortunately, we were unable to identify the paper. We have, instead, cited a paper by Talsma (2008) that makes a similar point.

R2.4. Page 5, "A second class of EMG correction methods employ..." - should be "employs".

This was corrected.

R2.5. Page 6, the first complete paragraph is a single, overly long sentence.

We agree. This was corrected to read,

“Although ICA shows much promise as a tool for correcting EMG and other kinds of biological artifact (e.g., Jung, Makeig, Humphries, Lee, McKeown, Iragui & Sejnowski, 2000), attempts to assess its validity have been limited. Many validation studies have relied on small samples of ad hoc data (Jung et al., 2000; Wallstrom, Kass, Miller, Cohn & Fox et al., 2004; Flexer, Bauer, Pripfl, & Dorffner, 2005; Ting, Fung, Chang, & Chan, 2006; Frank & Frishkoff, 2006). While others have used simulations that may not adequately represent the true complexity of relations between neuro- and myogenic activity (Crespo-Garcia, Atienza & Cantero, 2008; De Clercq, Vergult, Vanrumste, Van Hees, Palmini, Van Paesschen & Van Huffel, 2005; Delorme et al., 2007; Fitzgibbon, Powers, Pope & Clark, 2007; Frank & Frishkoff, 2006; Romero, Mananas & Barbanoj, 2008).”

R2.6. Pages 8 and 37, Debner should be Debener.

This was corrected.

R2.7. Page 13, "Channels situated..." - should be "Channels that were situated..." for parallel structure and clarity, given the rest of the sentence.

This was corrected.

R2.8. Throughout - The rate at which first-person sentence construction is used is distractingly high.
We have revised the MS to limit first-person constructions.

Reviewer #3

R3.1. The definition of the gross group could be elaborated since it's not clear.

As described in detail in the Supplemental Methods, “Gross Artifact. Several kinds of residual physiological and electromechanical artifacts were collectively classified as Gross. These included reference (Cz) and ground (nasion) sensor artifacts, electrocardiographic (ECG) artifacts, and alternating current (AC) artifacts. Reference and ground artifacts showed widespread, synchronous deflections in the raw time-series corresponding to periods of apparent “activation” in the component time-series, combined with a characteristically uniform topography (Supplementary Figure 2). ECG artifacts showed a characteristic pattern of deflections in the component time-series, typically persisting throughout the recording; a unilateral or, more rarely, bilateral posterior topography; and a low frequency (<3Hz) peak in the frequency-domain (Supplementary Figure 3). In contrast to one prior report (Viola et al., 2009), the raters anecdotally found ECG to be the easiest artifact to classify. AC artifacts were chiefly identified by a 60Hz peak, reflecting residual signal following notch-filtering, and harmonics in the frequency-domain along with sustained activation in the time-domain (Supplementary Figure 4).”

We have revised the Method section in the main report to read, “As detailed in the Supplementary Method and Results (Supplementary Figures 2-13), the remaining components were classified as neurogenic (Neuro), myogenic (Myo), a combination of the two sources (Neuro-Dominant or Myo-Dominant), or artifact (residual Gross or Ocular). Components classified as Gross included reference, ground, electrocardiographic, and alternating current artifacts. Components that met the minimum variance criterion, but proved impossible to unambiguously categorize were classified as Noise.”

R3.2. Why use the word artifact so exclusively when all of the non-neurogenic ICs are essentially artifacts?

We chose these terms to maximize clarity. While it is true that all non-neurogenic ICs can be considered artifact from the perspective of conventional EEG/ERP/ERSP analyses (that seek to maximize neural signal and minimize everything else), in the context of this report we thought it was more clear to discriminate myogenic “signals” from other kinds of artifact—rather than lumping them together. Likewise, given the indeterminate origins of Low-Variance and Noise ICs, it seemed appropriate to conceptually distinguish them from classifiable artifact sources (e.g., ECG, blinks).

R3.3. This classification scheme is to some extent subjective and artificial, and one could easily imagine other categories and subcategories that they did not include (e.g., cardiac, respiration, retina-related, etc.).

As noted above, for the most part, these kinds of artifact were classified (e.g., as Gross or Ocular). From our reading of the relevant methodological literature, while our scheme was necessarily subjective it was also reasonably comprehensive.
R3.4. The inter-rater reliability was high, but this is not surprising, and in no way proves any ground truth to this classification scheme. ... The fact that they classified so many ICs as mixed suggests at best that there is plenty of subjective uncertainty about some ICs, and at worst, that ICA is not doing a good job at blind source separation.

This is not quite correct. In fact, as noted in the Discussion section, the combination of detailed classification criteria and high inter-rater reliabilities are consistent with a low degree of “subjective uncertainty.” Such uncertainty, as we understand this reviewer’s term, would tend to inflate disagreements among raters and lower reliability. Taken with our other observations, this result is more consistent with this reviewer’s “worst-case” scenario: ICA failed to wholly separate myo- from neurogenic sources (see also Reviewer #1’s comments).

We do agree that high reliability is a necessary but not sufficient condition for the validity of our classification scheme.

R3.5. One possible problem is that they did not record enough data samples to use ICA without PCA processing. The dimensionality reduction to 64 dimensions could make it virtually impossible to have one IC for each source. The true mixture of sources probably involves hundreds if not thousands of locally coherent brain regions, muscle regions, etc. Because of the limited number of sensors, we can only hope to recover the sources of highest variance. By projecting the data to only 64 dimensions, it is very possible that the resulting ICs are essentially non-physiological (e.g., mixed neurogenic and myogenic). It would have been informative to record at least one longer session with more blocks and to try different PCA dimensions and try to match components across decomposition to see whether some of the ICs “split” into more physiological components once more dimensions are allowed.

First, our decision to pre-process the data using PCA is in accord with suggestions regularly made by the EEGLAB development team to investigators using high-density arrays (see http://sccn.ucsd.edu/pipermail/eeglablist/). That is, this decision was conventional.

Second, as noted in Footnote 3:
“Preliminary inspection of the 128 components extracted from the native electrode array indicated over-fitting, evidenced by fragmentation of artifacts across components (Li, Adali & Calhoun, 2007; Lawrence & Hancock, 1999). By contrast, exploratory analyses (not reported) showed that reduction to 48 or fewer PCs prior to ICA led to under-fitting, evidenced by cross-contamination of EEG, physiological artifacts, and noise (Fava & Velicer, 1996).”

This would seem to satisfy this reviewer’s suggestion that we “try different PCA dimensions...to see whether some of the ICs "split" into more physiological components once more dimensions are allowed.”

Third, as noted in the Supplementary Method and Results, “Just prior to submitting this report, we became aware of a recent ERP study (Mouraux & Iannetti, 2009) exploiting the Bayesian model-order estimation procedure used by the FSL Melodic software package (Beckmann & Smith, 2002, 2004; Rajan & Rayner, 1997). Using
Matlab code kindly provided by the lead author, Andre Mouraux, a post hoc analysis was performed to determine the number of dimensions (“model-order”) characterizing the (128-channel) dataset. This indicated that the median number of dimensions was 39.5 (SD: 6.8) with a range of 23-53. This suggests that the 64-component extraction used in the present report was sufficient to avoid underfitting, but moderately overfitted most participants.” These values are consistent with prior reports (e.g., Naeem, Brunner & Pfurtscheller, 2009; Onton et al., 2006).

Collectively, the latter two lines of evidence suggest that this reviewer’s suggestion that “the true mixture of sources probably involves hundreds if not thousands of locally coherent brain regions, muscle regions, etc.” is wrong, at least in terms of what the measured data are telling us (it is, of course, correct at more cellular levels of analysis). Likewise, the concern that “by projecting the data to only 64 dimensions, it is very possible that the resulting ICs are essentially non-physiological (e.g., mixed neurogenic and myogenic)” (i.e., that the presence of mixed myo/neurogenic ICs reflects an artifact of PCA dimensionality reduction) is also likely to be incorrect. Finally, it is also worth noting that if there are in fact more than 64 true sources, they will be spatially aliased (mixed) in an ICA conducted on any low-density electrode array dataset (Srinivasan, Tucker & Murias, 1998), which is generally what has been used in prior investigations of ICA-based artifact correction.

Note that the possibility that we collected insufficient data (“One possible problem is that they did not record enough data”) is addressed and rejected in Footnote 3. Indeed, we exceeded the \( k \geq 20 \) criterion of Onton et al. (2006) by a factor of 3.

**R3.6.** Another factor possibly contributing to a poor ICA decomposition is that they did not high-pass the data. In my experience and that of many ICA users and developers, ICA does not work well unless you filter out signals below 1 or 2Hz.

It is not entirely clear to us why this would be. It is also at odds with common conventions in the EEG/ERP literature (e.g., Luck, 2005; Luck’s ERP Workshop materials, http://www.erpinfo.org). Furthermore, it does not seem to be documented in any of the applied EEG methodological papers of Makeig and colleagues (e.g., Jung et al, 2000a, b).

**R3.7.** In the introduction, page 4, it is said that "fixed spatial or spectral filters or templates" cannot be applied to the correction of EMG. This statement is too general and unclear. Although I agree with the idea that fixed spatial only, or frequency only, etc, approaches may not do the job, there are creative ways (...) to use these features together for denoising. In fact, ICA-based EMG correction can be thought of as removing a set of fixed spatial patterns that are suspected to be non-neurogenic based on eye-balling the spectrum, time-course, and spatial topography itself. More quantitative approaches can be used with the same denoising goal and more automatically.

This concern again reflects an unfortunate word choice on our part. By ‘fixed’ we simply meant an unvarying canonical template used across all participants, conditions, etc.

Accordingly, we have revised the Introduction to read,
“Given marked individual differences in the spectral and anatomical profile of myogenic activity, spatial or spectral filters or templates that are fixed across subjects cannot be fruitfully applied to the correction of EMG artifact”

**R3.8.** On page 5, the description of source localization is not very clear. It seems to be mixing up the forward problem (modeling the predicted electric potentials that would be measured given an active current-density distribution, and volume conductor model) with the inverse problem (modeling and estimating the sources that generated the measurements).

We have revised the section to read,
“Source-estimation (“localization”) is a technique that estimates neurogenic signals from scalp EEG recordings. This is achieved by developing a forward-model that uses the biophysics of the EEG (e.g. the spatial filtering imposed on neurogenic signals by the cerebrospinal fluid, skull and scalp) to predict signals on the scalp given a particular neural generator. Source-estimation occurs when this model is inverted and used to estimate a probable neural generator given scalp-recorded signals (Pizzagalli, 2007). McMenamin et al (2009) speculated that the EMG-contaminated data cannot be properly localized because a solution space that only allows intra-cranial dipoles cannot account for a scalp-recording that contain both intra-cranial (neurogenic) and extra-cranial (myogenic) sources. The resulting attempt at localization will be corrupted and biased away from the true neurogenic solution. Removing signal due to extra-cranial sources from the data prior to source-estimation may circumvent this problem. Unfortunately, this is not possible using GLM-based techniques because they cannot reconstruct the artifact-free EEG time-series”

**R3.9.** On page 8, I think there's a typo on "Absent such evidence" (maybe just missing an "of").

We have revised the sentence to read,
“Without such evidence, poor validity might simply reflect inadequate training or an ambiguous classification protocol.”

**R3.10.** On the caption of figure 10, change "whereas a points" to "whereas points". Actually, this caption could be explained more clearly.

We strongly agree with this concern. Accordingly, we revised the caption to read,
“Figure 10. Source-space effects of EMG correction. Uncorrected (x-axis) compared to ICA-corrected t-values (y-axis) for voxels in the source-space, color-coded by ROI (Blue: Neurogenic ROI; Red: Myogenic ROI; Green: Both ROIs). Points close to the solid horizontal line are voxels where t approached zero after correction, indicating high sensitivity for EMG-contaminated contrasts and low specificity for the EMG-free neurogenic contrast. Conversely, points lying beyond the broken horizontal lines (p = .05) were significantly altered by correction. Points along the diagonal were unchanged by correction, indicating low sensitivity for EMG-contaminated contrasts and high specificity for the neurogenic contrast. Note that row B plots the uncorrected OR-CR contrast (x-axis) against the change in OR-CR after correction (i.e. correction induced error; y-axis).”
R3.11. It would have been nice to see more results at the source level, especially after each denoising strategy. Figure 10, is a bit too abstract to visually see the effects of denoising with the optimal methods (effects that were not shown visually on Fig 9).

We concur. Accordingly, several additional figures (Suppl. Figs. 16-17) have been added to the revised Supplemental Methods and Results to depict the myogenic (OR-OT) effect and error in the neurogenic effect (corrected OR-CR vs uncorrected OR-CR) for each of the two denoising protocols investigated in the intracerebral source-space (LORETA solutions). These contrasts are beneficial for highlighting the failures in sensitivity and specificity, respectively, for each of the two denoising methods.

R3.12. Last but not least, there are really two ways of denoising using ICA. One approach was used in this paper (i.e., subtract the projected times-series of each "noise" IC). The other approach is to use all normalized "noise" IC scalp projection vectors to form an orthogonal noise subspace, and then project the data out of that space (i.e., use Signal-Space projection (SSP)). This is a more aggressive denoising approach and may lead to cleaner data. This approach may also prove more robust at the source-space level, since to make both sides of the equation equal, both the data and the leadfield matrix are projected by the same SSP projector operator (something that is not done in the approach presented in this paper). This SSP of the leadfield matrix before computing the LORETA inverse operator could possibly improve the EMG-correction at the source level. LORETA assumes an uncorrelated noise model. Further improvements could be obtained by assuming a correlated noise model instead (e.g., using non-diagonal noise covariance and/or csd matrices).

We have revised the Discussion to include the following sentence, “It might also be fruitful to investigate the utility of signal-space projection methods (Nolte & Curio, 1999; Tesche, Uusitalo, Ilmoniemi, Huotilainen, Kajola & Salonen, 1995; Uusitalo & Ilmoniemi, 1997).”
Validation of ICA-Based Myogenic Artifact Correction for Scalp and Source-Localized EEG

Brenton W. McMenamin*
Department of Psychology
Center for Cognitive Science
University of Minnesota – Twin Cities

Alexander J. Shackman*
Laboratory for Affective Neuroscience
Waisman Laboratory for Brain Imaging and Behavior
University of Wisconsin – Madison

Jeffrey S. Maxwell
U.S. Army Research Laboratory, Aberdeen, MD
Laboratory for Affective Neuroscience
University of Wisconsin – Madison

David R. W. Bachhuber, Adam M. Koppenhaver, Lawrence L. Greischar, and Richard J. Davidson
Laboratory for Affective Neuroscience
Waisman Laboratory for Brain Imaging and Behavior
University of Wisconsin – Madison

Tables: 1
Figures: 10
Supplemental Files (Methods and Results): 1

*Brenton McMenamin and Alexander Shackman are co-first authors on this manuscript.

MS Correspondance:
Brenton W. McMenamin, Alexander J. Shackman or Richard J. Davidson
mcmen020@umn.edu, shackman@wisc.edu or rjdavids@wisc.edu
Laboratory for Affective Neuroscience
University of Wisconsin—Madison
1202 West Johnson Street
Madison, Wisconsin 53706
Abstract

Muscle electrical activity, or “electromyogenic” (EMG) artifact, poses a serious threat to the validity of electroencephalography (EEG) investigations in the frequency-domain. EMG is sensitive to a variety of psychological processes and can mask genuine effects or masquerade as legitimate neurogenic effects across the scalp in frequencies at least as low as the alpha band (8-13Hz). Although several techniques for correcting myogenic activity have been described, most are subjected to only limited validation attempts. Attempts to gauge the impact of EMG correction on intracerebral source models (source “localization” analyses) are rarer still. Accordingly, we assessed the sensitivity and specificity of one prominent correction tool, independent component analysis (ICA), on the scalp and in the source-space using high-density EEG. Data were collected from 17 participants while neurogenic and myogenic activity was independently varied. Several protocols for classifying and discarding components classified as myogenic and non-myogenic artifact (e.g., ocular) were systematically assessed, leading to the exclusion of one-third to as much as three-quarters of the variance in the EEG. Some, but not all, of these protocols showed adequate performance on the scalp. Indeed, performance was superior to previously validated regression-based techniques. Nevertheless, ICA-based EMG correction exhibited low validity in the intracerebral source-space, likely owing to incomplete separation of neurogenic from myogenic sources. Taken with prior work, this indicates that EMG artifact can substantially distort estimates of intracerebral spectral activity. Neither regression- nor ICA-based EMG correction techniques provide complete safeguards against such distortions. In light of these results, several practical suggestions and recommendations are made for intelligently using ICA to minimize EMG and other common artifacts.
Validation of ICA-Based Myogenic Artifact Correction for Scalp and Source-Localized EEG

Peri-cranial muscle or myogenic activity is distinguished by its relatively high amplitude, broad spectral and often anatomical distributions, and exquisite sensitivity to a variety of psychologically interesting processes. Consequently, it poses a serious inferential hazard for any electroencephalography (EEG) investigation in the frequency-domain. This artifact can compromise sensitivity by masking effects of interest or diminish specificity by masquerading as a neurogenic effect. Although several techniques have been developed to correct myogenic activity (Shackman, McMenamin, Slagter, Maxwell, Greischar & Davidson, 2009), many have been subjected to only limited attempts at validation, rendering their utility questionable. In particular, the sensitivity and specificity of one prominent electromyography (EMG) correction tool, independent component analysis (ICA), remains unclear.

Several properties of cranial EMG are collectively responsible for its pernicious effects. First, EMG is sufficiently sizable to perturb all classic EEG bands. Goncharova, McFarland, Vaughan, and Wolpaw (2003) report myogenic artifact reliably as low as 2Hz, making even the widely used alpha band (8-13Hz) vulnerable to muscle artifacts (Lee & Buchsbaum, 1987; Willis, Nelson, Rice & Black, 1993; Van Boxtel, 2001). Second, EMG can often be detected across the entire scalp (Goncharova, McFarland, Vaughan & Wolpaw, 2003) due to volume conduction of myogenic activity independently generated by muscles across the head, face and neck. Anterior electrodes are sensitive to facial muscles, such as the corrugator supercilii and frontalis; lateral electrodes are sensitive to the muscles of mastication, masseter and temporalis; and posterior electrodes are sensitive to muscles at the intersection of the cranium, spine, and torso, such as occipitalis (Supplementary Figure 1). Third, EMG is temporally confounded with a variety of experimental manipulations. Facial EMG, in particular, is sensitive to numerous cognitive and affective processes, including cognitive load (Cohen, Davidson, Seulis, Saron & Weisman, 1992; Waterink & Van Boxtel, 1994), facial mimicry (Dimberg, Thunberg & Elmehed, 2000), and induced emotional states (Borden, Petersom & Jackson, 1991; Coan & Allen, 2003; Bradley, Codispoti, Cuthbert & Lang 2001).

EMG also exhibits less stereotypy than other biological artifacts. Ocular and cardiac artifacts, for example, arise from fixed sources and do not qualitatively differ across individuals. EMG, however, arises from the activity
of spatially distributed, functionally independent muscle groups, with distinct topographic and spectral signatures (Goncharova et al. 2003). For instance, *frontalis* activity peaks around 25Hz, whereas *temporalis* generates a low peak around 20 Hz and broad plateau centered around 40-80 Hz (Goncharova et al., 2003). The spectral composition of myogenic activity also varies as a function of contraction intensity (Goncharova et al., 2003) and fatigue (Chung, Kim & McCall, 2002). This is compounded by the fact that the relative contributions of each muscle group to the cranial EMG vary substantially across elicitors and individuals (Tassinary, Cacioppo & Vanman, 2007) and may differ somewhat between spontaneous and voluntary contractions (Davidson, Shackman & Maxwell, 2004; Morecraft & Tanji, 2009).

Given the inferential hazards posed by EMG, there is substantial interest in developing tools to remedy myogenic artifact. Generally, EEG artifacts can be addressed in one of two ways, rejecting contaminated epochs of data or filtering artifact from neurogenic activity. Rejection-based techniques are most appropriate for transient artifacts, such as blinks, that influence a small portion of the data record. The protracted time-course of EMG makes such a solution impractical—the high data rejection rate would markedly erode the signal to noise ratio (Jung, Makeig, Westerfield, Townsend, Courchesne & Sejnowski, 2000; Talsma, 2008). Moreover, because EMG covaries with cognitive and affective processes of interest, rejecting data laden with EMG artifact would likely entail discarding some of the most interesting, discriminative periods of neural activity (Davidson, Ekman, Saron, Senulis & Friesen, 1990). For these reasons, EMG mandates the use of filtering techniques capable of separating myogenic from neurogenic activity. Given marked individual differences in the spectral and anatomical profile of myogenic activity, spatial or spectral filters or templates that are fixed across subjects cannot be fruitfully applied to the correction of EMG artifact (cf. Frank & Frishkoff, 2006; Ille, Berg & Scherg, 2002; Koskinen & Vartiainen, 2009). Instead, a more flexible approach is required.

One class of techniques for correcting EMG artifact employs variants of the general linear model (GLM), such as multiple regression and ANCOVA, to identify and discard variance in a neurogenic band of interest (e.g., alpha) that is predicted by activity in an *a priori* EMG band (e.g., 70–80Hz). The advantage of this technique is that it does not require dedicated EMG channels or manual intervention, and, by performing separate corrections at each site, can accommodate individual differences in artifact topography. These GLM-based techniques have
proven quite popular (Allen, Coan & Nazarian, 2004; Davidson, Jackson & Larson, 2000), and McMenamin, Shackman, Maxwell, Greischar and Davidson (2009) have shown that at least one variant of this technique displays adequate sensitivity and specificity on the scalp.

Despite these strengths, GLM-based EMG correction techniques suffer from two key limitations. First, they do not permit reconstruction of the EEG time-series. Thus, while useful for investigations of tonic (“resting”) and induced changes in the EEG spectra (e.g., Lutz, Greischar, Rawlings, Ricard & Davidson 2004), GLM-based tools cannot be applied to studies relying on event-related spectral perturbation (ERSP) measures (Onton, Westerfield, Townsend & Makeig, 2006). Second, McMenamin et al (2009) reported that applying GLM-based techniques to source-estimated EEG in a voxelwise manner is not appropriate if the data has been corrupted by EMG prior to localization. Source-estimation (“localization”) is a technique that estimates neurogenic signals from scalp EEG recordings. This is achieved by developing a forward-model that uses the biophysics of the EEG (e.g. the spatial filtering imposed on neurogenic signals by the cerebrospinal fluid, skull and scalp) to predict signals on the scalp given a particular neural generator. Source-estimation occurs when this model is inverted and used to estimate a probable neural generator given scalp-recorded signals (Pizzagalli, 2007). McMenamin et al (2009) speculated that the EMG-contaminated data cannot be properly localized because a solution space that only allows intra-cranial dipoles cannot account for a scalp-recording that contain both intra-cranial (neurogenic) and extra-cranial (myogenic) sources. The resulting attempt at localization will be corrupted and the true neurogenic solution rendered unrecoverable. Removing extra-cranial source activity from the data prior to source-estimation may circumvent this problem. Unfortunately, this is not possible using GLM-based techniques because they cannot reconstruct the artifact-free EEG time-series.¹

A second class of EMG correction methods employs ICA to decompose the EEG time-series into a set of temporally independent components (Delorme, Sejnowski & Makeig, 2007; Onton et al., 2006; Onton & Makeig, 2006; Makeig, Debener, Onton, & Delorme, 2004). Components are inspected visually for the presence of artifact and those classified as predominantly artifactual (e.g. EMG or blinks) are discarded. Like GLM-based correction

¹ Source modeling in the frequency domain requires phase information in the form of the cross-spectra. Extant GLM-correction techniques operate on estimates of spectral power (squared amplitude) and discard information about the phase of EEG oscillations required to compute the cross-spectra.
techniques, ICA does not require dedicated EMG channels and can accommodate variation across the scalp. More importantly, unlike GLM-based techniques, ICA allows reconstruction of the artifact-filtered time-series, which can then be used for analyses employing averaging, spectral decomposition, or source modeling.

Although ICA shows great promise as a tool for correcting EMG and other kinds of biological artifact (e.g., Jung, Makeig, Humphries, Lee, McKeown, Iragui & Sejnowski, 2000), attempts to assess its validity have been limited. Many validation studies have relied on small samples of ad hoc data (Jung et al., 2000; Wallstrom, Kass, Miller, Cohn & Fox et al., 2004; Flexer, Bauer, Pripfl, & Dorffner, 2005; Ting, Fung, Chang, & Chan, 2006; Frank & Frishkoff, 2006). While others have used simulations (e.g., Crespo-Garcia, Atienza & Cantero, 2008; De Clercq, Vergult, Vanrumste, Van Hees, Palmini, Van Paesschen & Van Huffel, 2005; Delorme et al., 2007; Fitzgibbon, Powers, Pope & Clark, 2007; Frank & Frishkoff, 2006; Romero, Mananas & Barbanoj, 2008). In simulations, real or artificial EMG activity is mathematically “injected” into otherwise artifact-free EEG. The potential problem with this strategy is that the assumptions underlying injection (e.g., the degree of temporal and spatial correlation with neurogenic signals) may not characterize real EMG contamination, potentially limiting external validity and biasing the results in favor of correction techniques founded on similar assumptions (Grouiller, Vercueil, Krainik, Segebarth, Kahane & David, 2007; Hoffmann & Falkenstein, 2009).

Accordingly, the major aim of the present study was to quantitatively assess the quality of EMG artifact correction afforded by ICA. Ideally, validation would quantitatively establish that a technique possesses a high degree of sensitivity (i.e., attenuates myogenic artifact) and specificity (i.e., preserves neurogenic signals) in a reasonably large and varied dataset. This requires data in which the presence or absence of EMG (“ground truth”) is definitive or can be reasonably assumed. To this end, the dataset previously employed by McMenamin et al. (2009) for testing the validity of GLM-based correction techniques was reanalyzed using ICA. This had the advantage of facilitating direct comparisons across correction techniques. In this dataset, 128-channel EEG was acquired while neurogenic and myogenic activity were independently varied. Alpha band neurogenic activity was selectively increased or decreased by instructing participants to close or open their eyes, a procedure termed the “Berger maneuver.” Myogenic activity was manipulated by instructing participants to alternately tense and relax their cranial muscles. The sensitivity and specificity of ICA-based EMG correction were then quantitatively
ICA-based EMG correction has been assessed in the alpha band using methods similar to those described in our prior report (McMenamin et al., 2009).

Several considerations led us to focus on the alpha band. First, it is relatively easy to manipulate neurogenic activity in this frequency. To our knowledge, comparably robust manipulations do not exist for the other classical EEG bands. Second, alpha activity has been among the most widely used spectral indices of neural activity, from the earliest EEG research (Berger, 1929/1969), to contemporary investigations of memory (Freuenberger, Fellinger, Sauseng, Gruber & Klimesch, 2009; Gevins & Smith, 2000; Hamidi, Slagter, Tononi & Postle, 2009), perception and attention (Romei Brodbeck, Michel, Amedi, ascual-Leone & Thut, 2008; Thut & Miniussi, 2009), emotion (Coan & Allen, 2003; Davidson et al., 1990), temperament and individual differences (Carver & Harmon-Jones, 2009; Shackman, McMenamin Maxwell, Greischar & Davidson in press), and psychopathology (Thibodeau, Jorgensen & Kim, 2006; DeRubeis, Siegle & Hollon, 2008).

The other major aim of this study was to test whether ICA-based techniques constitute a valid EMG correction technique for distributed intracerebral source modeling. Source modeling is an increasingly popular technique for maximizing the anatomical information yielded by scalp-recorded EEG (Pizzagalli, 2007) and the dissemination of commercial and freely available software for performing distributed source localization, such as Cartool (http://brainmapping.unige.ch/Cartool.htm), EMSE (http://www.sourceSignal.com), LORETA-KEY (http://www.unizh.ch/keyinst/) and SPM5 (http://www.fil.ion.ucl.ac.uk/spm/), is likely to accelerate this trend. Furthermore, prior work indicates that EMG correction techniques deemed valid on the scalp do not necessarily confer validity in the intracerebral source-space (McMenamin et al., 2009). Accordingly, ICA-based procedures that proved valid on the scalp were also tested with source solutions estimated using the low-resolution electromagnetic tomography (LORETA) algorithm.

A minor aim of this study was to evaluate the degree to which variation in the protocol for filtering non-myogenic artifacts, such as eye movements, impacts the quality of EMG correction. To date, existing methodological and empirical reports employing ICA provide little guidance on the question of which components ought to be discarded (Shackman et al., 2009). Furthermore, despite the fact that ICA requires trained raters to inspect hundreds or even thousands of components for a single high-density EEG study (number of components \(\approx\) channels \(\times\) participants; see Supplementary Method), the reliability of component classification has only rarely
been reported (Viola, Thorne, Edmonds, Schneider, Eichele & Debener, 2009). Without such evidence, poor validity might simply reflect inadequate training or an ambiguous classification protocol. Accordingly, the inter-rater reliability was computed.

Method

Participants.

The dataset consisted of seventeen individuals recruited from the University of Wisconsin–Madison campus (16 female; $M = 24.1$ years, $SD = 7.1$) and described in an earlier report assessing the validity of GLM-based EMG correction techniques (McMenamin et al., 2009). Each received US$20 for their participation. Participants provided informed consent in accord with guidelines prescribed by the local Institutional Review Board.

Design.

In order to independently manipulate neurogenic and myogenic activity in the alpha band (8-13Hz), the experiment took the form of a 2 (Eyes Open/Closed) × 2 (Muscles Tense/Relaxed) repeated-measures design. We anticipated that participants would generate greater broad-spectrum power, including increases in alpha power, indicative of reduced neural activity (Allen et al., 2004; Oakes et al., 2004), during the eyes-closed condition. We further expected participants to generate greater alpha power, indicative of increased muscle activity, during the muscles-tense condition. Hereafter, these four conditions are referred to using the following acronyms: Open-Relaxed (OR), Open-Tense (OT), Closed-Relaxed (CR), and Closed-Tense (CT).

Procedure

Procedures were identical to those detailed by McMenamin et al (2009). In brief, participants were instructed how to properly tense facial muscles at the outset of the session. Frontalis and corrugator muscles were contracted by lifting and squeezing the eyebrows together; masseter and temporalis were contracted by lightly
clenching the jaw. EEG was acquired during sixteen 32-second blocks (order counterbalanced; 4 blocks/condition).

Participants were continuously monitored via a closed-circuit audio-video circuit and real-time EEG.

**EEG Acquisition and Preliminary Reduction**

EEG were collected using a 128-channel Geodesic Sensor Net (GSN128; Electrical Geodesics Inc., Eugene, OR) referenced to vertex (Cz) and sampled at 500-Hz (analog anti-aliasing: 0.1 - 250 Hz). Data reduction used a combination of EEGLAB (Delorme & Makeig, 2004; http://www.sccn.ucsd.edu/eeglab) and in-house code written for MATLAB (http://www.mathworks.com). A zero-phase 60-Hz notch filter removed line noise from calibrated (µV) data, and bad channels (±100µV for >20s) or gross artifacts (±100µV for >4 channels) were manually identified and rejected. Such artifacts were rejected to better approximate the subtle contamination of signal that can occur when EMG covaries with an experimental treatment. Removal of non-stereotyped artifact also maximizes the quality of the ICA.

**ICA**

*Overview.* Prior to ICA, spatial Principal Components Analysis (PCA) was used to reduce the dimensionality of the EEG from 128 channels to 64 principal components (PCs). This was performed in a single step as part of the ICA using the EEGLAB *runica* command, implementing the extended Infomax algorithm (Bell & Sejnowski, 1995; Lee, Girolami & Sejnowski, 1999).

---

2 In hindsight, the acquisition parameters were not optimal for measuring high-frequency EMG effects. We would recommend that future studies of myogenic artifact use a higher sampling rate (e.g., 1000Hz) and more conservative anti-aliasing filter (e.g., 250Hz) to compensate for non-zero filter roll-off.

3 There were two reasons for doing so, aside from computational and classification efficiency. First, preliminary inspection of the 128 components extracted from the native electrode array indicated over-fitting, evidenced by fragmentation of artifacts across components (Li, Adali & Calhoun, 2007; Lawrence & Hancock, 1999). By contrast, exploratory analyses (not reported) showed that reduction to 48 or fewer PCs prior to ICA led to under-fitting, evidenced by cross-contamination of EEG, physiological artifacts, and noise (Fava & Velicer, 1996). Second, it has been suggested (Onton et al., 2006; Romero, Mananas & Barbanoj, 2008) that Infomax ICA requires a minimum of 20 x c² samples, where c is the number of channels or, equivalently, PCs. For the native electrode array, this would require 20 x 128² = 327680 samples, whereas we had at most 16 blocks x 32-s x 500Hz = 256000 samples. Reducing the model order by half allowed us to satisfy this criterion (20 x 64² = 81920 samples). Quantitative estimates of “model order,” the number of components required to adequately but parsimoniously describe the data, suggested that this was sufficient (see Supplementary Method and Results).
The primary aim of this study was to assess the validity ICA for EMG artifact correction. Accordingly, three protocols for the correction of EMG artifact, described below, were investigated. A secondary aim of this study was to investigate the degree to which the quality of EMG correction was dependent on the protocol for removing non-myogenic sources of variance (e.g., ocular artifact, noise components). Consequently, three ICA-based protocols for the correction of non-neurogenic/non-myogenic (NNNM) components, described below, were also examined. The quality of EMG artifact correction was evaluated for all nine factorial combinations of the EMG and NNNM protocols. Following removal of the relevant components, the filtered 128-channel time-series were reconstructed. Subsequent analyses used only the 107 cephalic electrodes. Exploratory analyses (not reported) using the complete 128-channel array indicated worsened performance when the peri-cephalic electrodes on the face and along the posterior edge of the array were retained (Supplementary Figure 1).

**Component classification.** Using in-house code, the variance accounted for by each of the ICs was assessed. By default, components that individually accounted for <0.2% of the variance were categorized as *Low-Variance*. In cases where the determination was unambiguous, exceptions were made. As detailed in the Supplementary Method and Results (Supplementary Figures 2-13), the remaining components were classified as neurogenic (*Neuro*), myogenic (*Myo*), a combination of the two sources (*Neuro-Dominant* or *Myo-Dominant*), or artifact (residual *Gross* or *Ocular*). Components classified as Gross included reference, ground, electrocardiographic, and alternating current artifacts. Components that met the minimum variance criterion, but proved impossible to unambiguously categorize were classified as *Noise*. Classifications were made by two raters based on inspection of the component’s time-series, power spectrum, and topography. When disagreements occurred, final classification was by consensus. Inter-rater reliability, assessed prior to consensus using Krippendorff’s alpha (Hayes & Krippendorff, 2007), was excellent, $\alpha=.98$ (for details, see Supplementary Method).

---

4 Preliminary inspection of the ICA results indicated that such low-variance (<0.2%) components were dominated by noise, making them difficult to reliably classify and leading raters to devote an undue amount of time to their consideration. The threshold of 0.2% was chosen to minimize the cumulative amount of variance that was automatically classified as “noise” (i.e., remained unclassified).
Correction of EMG artifact. Three different ICA-based protocols for removing myogenic artifact were assessed. The **Minimal-EMG** protocol discarded only those components that contained clear EMG activity in the absence of any identifiable neurogenic activity (i.e., rejected Myo components). The **Intermediate-EMG** protocol expanded this definition to include mixed components in which myogenic activity was *more* prominent than neurogenic activity (i.e. rejected both Myo and Myo-Dominant components). The **Maximal-EMG** protocol rejected any component containing myogenic signal, even if myogenic activity was *less* prominent than neurogenic activity (i.e., rejected Myo, Myo-Dominant, and Neuro-Dominant components). Thus, the Maximal-EMG protocol performs the strictest filtering of the data, at the potential expense of discarding neurogenic signals of interest.

Filtering of non-neurogenic/non-myogenic (NNNM) signals. To provide a specific test of ICA’s utility for removing EMG artifact, it is necessary to first filter signals that are not clearly neurogenic or myogenic (cf. McMenamin et al., 2009). However, the choice of which components to remove is subjective and has a marked impact on the number of components and percentage of variance retained (see Results). Accordingly, three different ICA-based protocols for filtering non-neurogenic/non-myogenic signals were used. The **Minimal-NNNM** protocol made the fewest assumptions, filtering only those components that were explicitly classified as Gross or Ocular artifact, similar to the method used in McMenamin et al. (2009). The **Intermediate-NNNM** protocol made the additional assumption that components categorized as Noise do not contain meaningful neurogenic signal and filtered them as well. The **Maximal-NNNM** protocol further assumed that Low-Variance components do not contain significant neurogenic signal and filters them as well.

Scalp Spectral Power Density Estimation

Following reconstruction of the filtered time-series, epochs with residual artifact (i.e., deviations exceeding ±200 µV for more than half an epoch or variance exceeding 1000 µV²) or flat channels (epoch variance less than 0.25 µV²) were automatically rejected (Delorme et al., 2007). After residual artifact-rejection, the rejected channels were interpolated with a spherical spline when at least one neighboring electrode was usable (Greischar et al., 2004). Data were re-referenced to an average montage (Davidson et al., 2000; Dien, 1998) and spectral power
ICA-Based EMG Correction

Density (μV^2/Hz) estimated for the alpha (8-13Hz) band using Welch’s (1967) method on sliding Hanning-windowed epochs (50% overlap). Estimates were log_{10} transformed to normalize the distribution (Allen et al., 2004; Gasser, Bacher & Mocks, 1982).

LORETA Distributed Source Current Density Modeling

The modeling of distributed sources from scalp-recorded electrical activity was performed using previously published procedures (McMenamin et al., 2009; Shackman et al., in press) via in-house MATLAB code implementing the LORETA algorithm (Pascual-Marqui, Michel & Lehmann, 1994) to estimate intracerebral current density. LORETA has undergone extensive cross-modal validation (reviewed in Shackman et al., in press; Pizzagalli, 2007).

An inverse operator distributed with the LORETA-Key software suite (Pascual-Marqui, 1999; http://www.unizh.ch/keyinst/; \( \lambda=10^{-5} \)) was used to generate three-dimensional intracerebral current density estimates (A/m^2) from cross-spectra calculated using the artifact-free Hanning-windowed epochs from the scalp analyses. The forward-model is a 3-shell spherical head model using 107 cephalic EEG electrodes (Shackman et al., in press). The source-space is normalized to the Montreal Neurological Institute’s probabilistic MRI anatomical template (i.e., MNI305; Evans et al, 1993; Collins, Neelin, Peters & Evans, 1994), restricted to the cerebral gray matter, hippocampi, and amygdalae on a 7-mm^3 isotropic lattice. LORETA source-estimates were log_{10} -transformed prior to analysis (Thatcher, North & Biver, 2005). Results are displayed on the rendered canonical brain distributed with LORETA-Key.

Analytic Strategy

Overview. A valid correction technique should render EMG-contaminated data statistically equivalent to data collected under the same conditions in the absence of myogenic artifact (Frank & Frishkoff, 2006; Debener, Strobel, Sorger, Peters, Kranczioch, Engel & Goebel, 2007). Accordingly, each combination of the EMG correction and NNNM filtering protocols was evaluated in terms of its (i) sensitivity, the attenuation of myogenic
artifact (i.e., Tense vs. Relaxed) in the alpha band, (ii) specificity, the preservation of neurogenic effects (i.e., alpha-blocking: Eyes-Closed vs. Eyes-Open) in the alpha band, and (iii) the degree to which each protocol introduced correction artifacts, artificial effects generated by the correction. Sensitivity and specificity were assessed using regions of interest (ROIs) defined by the areas of peak myogenic and neurogenic activation, respectively. An ROI approach was used to constrain the number of comparisons in both scalp and LORETA source-space analyses. Only those filtering protocols that proved sufficiently valid on the scalp were assessed with LORETA. To permit a direct comparison of ICA- and GLM-based EMG correction techniques, key analyses reported in McMenamin et al. (2009) were recomputed using the identical validation techniques used here. These analyses are detailed in the Supplementary Method and Results.

*Sensitivity*. On the scalp, a myogenic ROI was created for each of the three NNNM filters using electrodes exhibiting a significant ($p < 0.05$) myogenic effect (OR-OT). Channels that were situated at the edge of the 107-channel electrode-array, were spatially discontiguous (i.e., lacked at least one nearest neighbor meeting the significance criterion), or also met the inclusion criteria for the neurogenic (i.e., specificity) ROI were excluded (range: 8-15 electrodes; many located at the posterior base of the array). The effect of the spatial contiguity criterion was minor, resulting in a single electrode being dropped. In the LORETA source-space, myogenic ROIs were created by identifying voxels in the OR-OT contrast with $p < .001$, using a cluster-extent threshold to correct for multiple comparisons (Nichols & Holmes, 2002; Shackman et al, in press).

Using the resulting ROIs, the degree to which each EMG correction protocol attenuated the myogenic contrast (i.e., EMG-corrected OR-OT vs. 0) was tested. Additional contrasts tested the degree to which each EMG-correction protocol removed myogenic effects using double differences that compared three EMG-corrected contrasts of interest and their uncorrected, artifact-free analogs. ICA’s ability to correct EMG artifact that negatively covaried with neurogenic signals was tested using the (EMG-corrected OT-CR) - (uncorrected OR-CR) contrast, and the ability to correct artifact that was positively covaried with neurogenic signals was tested using the (EMG-corrected OR-CT) - (uncorrected OR-CR) contrast. The amount of EMG artifact surviving each correction was indexed using median and peak ROI $t$-values, viewed as indices of typical and “worst-case” correction,
respectively. Significant $t$-tests for these contrasts indicate that the EMG-corrected EEG signals deviate from their artifact-free analogs, evidence of poor sensitivity.

Conversely, failure to reject the null hypothesis does not indicate the absence of residual myogenic activity. In order to rigorously test whether the EMG-corrected contrasts were *significantly* equivalent to artifact-free data, the Westlake-Schuirmann test (Seaman & Serlin, 1998) was employed as a follow-up test to non-significant contrasts. Sometimes termed the two one-sided tests (TOST) method, a number of fields (e.g., the US Food and Drug Agency; Department of Health and Human Services, 2001) consider TOST the gold standard for testing statistical equivalence. The null hypothesis for TOST is that the mean difference lies outside of the range $[-\varepsilon, \varepsilon]$, where $\varepsilon$ is an *a priori* error tolerance. To reject the null (i.e., demonstrate significant equivalence) for $\alpha = .05$, one must demonstrate that the $90\textsuperscript{th}$-percentile confidence interval of the mean difference between the artifact-free and EMG-corrected data lies completely within the interval $[-\varepsilon, \varepsilon]$. Following our prior report (McMenamin et al., 2009), $\varepsilon$ was set to 0.5 standard deviations of the artifact-free contrast (i.e., OR for the OR-OT contrast, OR-CR for positively/negatively covarying contrasts).

*Specificity.* Neurogenic ROIs were generated by thresholding the neurogenic contrast (OR-CR). Owing to the large size of this effect ($p < .001$ at all electrodes), it proved useful to threshold the contrast using a percentile approach. This had the advantage of creating ROIs that were similar in size to the myogenic ROIs used to interrogate sensitivity. On the scalp, this entailed selecting the upper tercile of channels. As before, channels that were situated on the edge of the array, were not spatially contiguous, or met the inclusion criteria for the myogenic (i.e., sensitivity) ROI were excluded. The latter two criteria led us to drop one electrode. In the source-space, voxels with absolute $t$-values in the upper tercile for the OR-CR contrast were selected.

As with the sensitivity analysis, $t$-tests and follow-up TOSTs were used to test whether neurogenic effects were preserved following EMG correction. First, to test the impact of correction *per se* on neurogenic activation, (EMG-corrected OR-CR) was compared against (uncorrected OR-CR). Second, to test correction’s impact on negatively covarying neurogenic and myogenic signals, (EMG-corrected OT-CR) was compared to (uncorrected OR-CR). Third, to test correction’s impact on positively covarying signals, (EMG-corrected OR-CT) was...
compared to (uncorrected OR-CR). The ε error tolerance for TOST follow-ups was defined using the uncorrected OR-CR contrast.

Correction artifacts. To investigate the degree to which protocols generated artificial results, two kinds of tests were conducted. The first test determined whether correction of the EMG-contaminated myogenic contrast produced artificial effects in the neurogenic ROI (i.e., EMG-corrected OR-OT). The second test assessed whether correction of the EMG-free neurogenic contrast yielded artificial effects in the myogenic ROI (i.e., [EMG-corrected OR-CR] – [uncorrected OR-CR]). The presence of artifactual effects was assessed using the same logic as the sensitivity and specificity tests.

Performance ratings. For each contrast and correction protocol, sensitivity was rated as: poor ($p_{Median\,t-test} < .05$ or median $p_{TOST} > .05$), questionable ($.10 > p_{Median\,t-test} > .05$ or median $p_{TOST} < .05$), adequate ($p_{Median\,t-test} > .10$ and median $p_{TOST} < .05$) or excellent ($p_{Peak\,t-test} > .05$ and all $p_{TOST} < .05$). Thus, the significance of the peak $t$-test was only considered in cases where a particular EMG-correction protocol showed evidence of adequate or excellent sensitivity using the median-based tests.

Results

Scalp

Effects Prior to EMG Correction

Visual inspection indicated that the topography of the four alpha-band contrasts was similar across the three NNNM protocols (Figure 1).

Myogenic ROI. Consistent with expectation, scripted muscle tensing increased alpha power near facial muscles at midline-, left- and right-frontal electrodes. The myogenic contrast (OR-OT) was significant at 34-39 anterior electrodes (Figure 2) with qualitatively similar peak locations across the three NNNM filters. The median $t$-scores ($t_s = -2.30$ to -2.44, $ps < .04$; $\eta_p^2$ s = 0.24 to 0.27) and extreme $t$-scores ($t_s = -2.82$ to -3.21, $ps < .01$; $\eta_p^2$ s =
ICA-Based EMG Correction

0.33 to 0.39) were similar across the three protocols, indicating that the degree of EMG contamination was also similar. This contrast was used to form the Myogenic ROIs (Figure 3), resulting in clusters of 23-25 contiguous anterior electrodes, extending to mid-frontal and fronto-central leads (e.g., AF7, F2, F3, F5, F7, FC2, FC3, FT7, T7).

Neurogenic ROI. Consistent with expectation, the Berger maneuver (OR-CR) altered power at all electrodes ($ts < -4.37, ps < 0.001$). The peak difference occurred at midline parietal sites (Figure 1), and defined the neurogenic ROIs (Figure 3). The three neurogenic ROIs contained 21-26 contiguous electrodes (e.g., Pz, P1, P2, P3, P5, P7, P9, POz, PO3, PO4, PO8, Oz, CPz) with comparable median ($ts = -7.35$ to $-7.49, p < .001; \eta_p^2 = 0.77$ to 0.78) and extreme $t$-scores ($ts = -8.85$ to $-9.12, p < .001; \eta_p^2 = 0.83$ to 0.84), indicating that the strength of neurogenic effect was minimally affected by the choice of NNNM protocol.

Covarying effects. In the absence of EMG correction, myogenic activity distorted the magnitude of neurogenic effects in the alpha band. For instance, when changes in EMG and EEG negatively covaried (OT-CR), the effect remained significant at all electrodes; but, the magnitude of the neurogenic effect was significantly attenuated at 34-39 electrodes relative to the uncontaminated effect (OR-CR). This resulted in a slightly shifted topography that deemphasizes activation at anterior sites. Notably, significant attenuation was present at posterior electrodes far removed from the area of peak myogenic artifact (Figure 4). A parallel, albeit non-significant, pattern of amplification occurred at anterior sites (not shown) when changes in EMG and EEG positively covaried (OR-CT). In this case, significant attenuation was only observed at a small number of parietal sites (Figure 5).

Descriptive Statistics for ICA

Classification. Figure 6 depicts the relative frequency and percentage of scalp variance predicted by each class of components. A similar pattern was found using means. Visual inspection indicates that the most frequent classification, comprising about one-fifth of the total, was Noise (Figure 6a). The frequency of the other classifications was somewhat smaller, but similar to one another (11-15%). Neuro-Dominant and Gross Artifact
components were infrequent (2-3%). The frequency of the Myogenic and Myogenic-Dominant components showed marked variability and was positively skewed. That is, a few individuals displayed many more of these two EMG-related components than the remainder of the group.

Collectively, the Neurogenic (25%) and Ocular components (21%) accounted for nearly half of the variance in scalp electrical activity (Figure 6b). Myogenic, Myogenic-Dominant, and Noise components each accounted for another 8-10%, while the Neurogenic-Dominant, Low Variance, and Gross Artifact components each accounted for less than 2% of the variance. There was substantial variability and positive skew in the percentage of variance accounted for by several artifact components, particularly Ocular and Myogenic-Dominant. It is worth emphasizing that a sizable proportion of the variance—equal to that accounted for by the purely Myogenic component—was predicted by the more heterogeneous Myogenic-Dominant component.

Figure 7 depicts the median percentage of variance that was discarded or retained following each combination of NNNM filter and EMG correction protocol. The choice of NNNM filter and EMG correction protocol markedly affected the percentage of variance discarded. Application of the NNNM filters removed from one-quarter to slightly more than one-third of the variance in scalp electrical activity (Figure 7). There was little difference between Intermediate- and Maximal-NNNM filters (center and right columns), presumably owing to the small net contribution of the Low Variance components. Application of the Minimal-EMG protocol removed another 10% of the variance (second row). When paired with the different NNNM filters, application of either the Intermediate- or Maximal-EMG protocols removed two-thirds to three-quarters of the variance (bottom two rows). Put another way, artifact accounted for about twice as much variance as neurogenic activity in this sample. The small difference between the Intermediate- and Maximal-EMG protocols (bottom two rows), presumably stems from the infrequency of Neuro-Dominant components.

Validity of ICA-Based EMG Correction

As summarized in Table 1, only four protocols showed questionable or better performance across all tests of sensitivity (Figures 2, 4, and 5), specificity (Figures 4, 5, and 8), and correction-induced artifact (Figures 2 and
ICA-Based EMG Correction

18: the Minimal-EMG protocol paired with Minimal- or Intermediate-NNNM filtering and the Maximal-EMG protocol paired with Minimal- or Maximal-NNNM filtering. For detailed results, see Supplementary Tables 1-3. Moreover, inspection of Table 1 indicates that among these four, the Minimal-EMG/Intermediate-NNNM protocol invariably equaled or exceeded the performance of the Minimal-EMG/Minimal-NNNM protocol; likewise, the Maximal-EMG/Maximal–NNNM protocol always outperformed the Maximal-EMG/Minimal-NNNM protocol. Accordingly, these two combinations of EMG correction and NNNM filtering were subjected to additional testing in the intracerebral source-space using LORETA.

Intracerebral Source Modeling

Effects Prior to EMG Correction

Myogenic ROI. Muscle tensing increased current density in a large number of frontopolar and ventral prefrontal voxels near the facial muscles (Figure 9). Across correction protocols, the myogenic contrast (OR-OT) was significant at 717-789 voxels (~30% of the source-space). Myogenic ROIs (715-749 voxels) were created from this contrast after applying a cluster-extent threshold (Supplementary Figure 14). The median t-scores ($t_s = -3.75$ to $-3.83$, $p < .01$, $\eta_p^2 = 0.47$ to 0.48) and extreme t-scores ($t_s = -6.19$ to $-6.33$, $p < .001$, $\eta_p^2 = 0.71$) were similar across the two filters.

Neurogenic ROI. The Berger maneuver was associated with widespread attenuation of current density across the posterior cortex (Figure 9). Using the neurogenic (OR-CR) contrast, ROIs were generated for the Intermediate- and Maximal-NNNM filters (765-766 voxels) across posterior voxels (Supplementary Figure 15). Median ($t_s = -3.92$ to $-3.96$, $p < .01$, $\eta_p^2 = .49$) and extreme t-scores ($t_s = -7.26$ to $-7.68$, $p < .001$, $\eta_p^2 = .77$ to .79) were similar across the two filters.

5 Exploratory analyses indicated dramatically worse performance in the “EMG” band (70-80Hz). Prior to EMG correction, the myogenic contrast (OR-OT) in this frequency range was significant at 89-90 electrodes across NNNM filters ($p < .01$ threshold). Following correction, it remained significant at 68-86 electrodes.
Covarying effects. The presence of uncorrected EMG artifact altered the magnitude of the neurogenic effects produced by the Berger maneuver (Figure 9). In particular, when changes in neurogenic and myogenic activity negatively covaried (OT-CR) there was a substantial attenuation of effects in the myogenic ROI for both median ($t_s > 3.75, ps < .01, \eta_p^2 s > 0.47$) and extreme $t$-scores ($t_s > 6.19, ps < .001, \eta_p^2 s > 0.71$). The attenuation was reduced, but still observable, in the posterior neurogenic ROI (median $ps < 0.08, \eta_p^2 s > 0.18$; extreme $t_s > 4.53, ps < 0.01, \eta_p^2 s > 0.56$). Conversely, when changes in neurogenic and myogenic activity positively covaried (OR-CT), effects in the myogenic ROI were amplified as indexed by both the median ($t_s < -2.30, ps < .04, \eta_p^2 s > 0.25$) and extreme $t$-scores ($t_s < -4.32, ps < .001, \eta_p^2 s > 0.54$). These deleterious effects were weaker in the neurogenic ROI, reaching significance for the extreme ($t_s < -3.81, ps < .01, \eta_p^2 s > 0.48$) but not the median $t$-scores ($ps > .72, \eta_p^2 s < 0.01$).

Sensitivity

Myogenic contrast (OR-OT). Although both protocols quantitatively reduced EMG contamination in the myogenic ROI (Figure 10a, red points), their sensitivity was poor (Supplementary Table 4, Supplementary Figures 16-17).

Negatively covarying contrast (OT-CR). The Maximal-EMG/Maximal-NNNM pairing showed adequate sensitivity (Supplementary Table 4), whereas the Minimal-EMG/Intermediate-NNNM pairing evidenced poor sensitivity. Across protocols, voxels in the myogenic ROI showing weaker EMG contamination were more resistant to correction (Figure 10c).

Positively covarying contrast (OR-CT). Again, the Maximal-EMG/Maximal-NNNM pairing demonstrated adequate sensitivity, whereas the Minimal-EMG/Intermediate-NNNM pairing showed poor sensitivity (Supplementary Table 4). Inspection of the myogenic ROI voxels that yielded peak errors indicated that the Minimal-EMG/Intermediate-NNNM pairing tended to undercorrect the data (negative $t$-scores), whereas the Maximal-EMG/Maximal-NNNM pairing tended to overcorrect the data (positive $t$-scores), albeit to a lesser degree. This is depicted in Figures 8c and 10d.
**Specificity**

*Neurogenic contrast (OR-CR contrast).* The Maximal-EMG/Maximal-NNNM pairing showed questionable specificity, whereas the Minimal-EMG/Minimal-NNNM pairing showed adequate specificity (Figure 10b, blue points, Supplementary Table 5, Supplementary Figures 18-19). Inspection of the neurogenic ROI voxels that yielded peak errors indicated that in the absence of EMG artifact both protocols attenuated changes in activation associated with the Berger maneuver. This was particularly evident for the Maximum-EMG protocol paired with Minimal-NNNM filtering.

*Negatively covarying contrast (OT-CR).* Both protocols tended to attenuate neurogenic activity, yielding questionable or worse specificity (Supplementary Table 5 and Figure 10c).

*Positively covarying contrast (OR-CT).* Both pairings exhibited acceptable or better specificity (Supplementary Table 5). Inspection of the neurogenic ROI voxels that yielded peak errors indicated that the Minimal-EMG/Intermediate-NNNM pairing led to undercorrection (negative t-scores), whereas the Maximal-EMG/Maximal-NNNM pairing led to overcorrection (positive t-scores). This is depicted in Figure 10d.

**Correction Artifact**

*Myogenic contrast (OR-OT) in the neurogenic ROI.* The Maximal-EMG/Maximal-NNNM pairing yielded an acceptable amount of correction-induced artifact when EMG was present (Supplementary Table 6 and Figure 10a, blue points), whereas the Minimal-EMG/Intermediate-NNNM pairing showed poor performance.

*Neurogenic contrast (OR-CR) in the myogenic ROI.* Both pairings showed an acceptable level of correction artifact when EMG was absent (Supplementary Table 6 and Figure 10b, red points).

---

6 We interpreted differences in the size of the neurogenic contrast as evidence of non-specificity. Nonetheless, such differences might instead reflect amplification of the neurogenic statistical effect (“signal-to-noise”) following EMG correction. This would occur if EMG correction led to reductions in error (“noise”) that were substantially larger than reductions in the mean difference across conditions (“signal”). If this were so, the magnitude of the t-test following EMG correction should increase. Contrary to this alternative account, the t-test for the neurogenic contrast was decreased following Maximal-EMG correction (Maximal-NNNM: \( t=-3.92 \); Maximal-EMG/Maximal-NNNM: \( t=-3.70 \)).
Given the substantial inferential threat posed by EMG contamination, there is a pressing need for valid correction tools. In recent years, ICA has rapidly become a popular tool for correcting EMG artifact, despite limited work assessing its validity for this purpose. Accordingly, we quantitatively tested the sensitivity and specificity of ICA-based EMG correction using a naturalistic dataset in which neurogenic (opening and closing the eyes, i.e. the “Berger maneuver”) and myogenic activity (scripted muscle tensing and quiescence) were independently manipulated. This design allowed investigation of ICA-based EMG correction under conditions in which changes in neurogenic and myogenic activity covaried, as they do in many experimental settings.

Low-intensity clenching of the face increased EEG spectral power across much of the scalp, extending as far as the fronto-central electrodes. Smaller pockets of myogenic activity were also present at the posterior edge of the electrode array (Figure 1). In contrast, the Berger maneuver altered power across the entire array, peaking at midline parietal sites. And while the effect-size of the anterior myogenic effect was only one-third that associated with the posterior neurogenic effect, it was sufficient to alter the magnitude of neurogenic effects. For instance, when changes in EMG and EEG negatively covaried, as they would be expected to do in many experiments, neurogenic effects associated with the Berger maneuver were attenuated at electrodes across the array, including posterior sites far removed from the area of peak myogenic activity (Figures 1 and 4).

Consistent with these findings, uncorrected EMG artifact increased alpha-band current density across large regions (~30%) of the intracerebral source-space located near the facial muscles (i.e., frontopolar and ventral prefrontal cortex, insula, temporal poles; Figure 9). This effect reflects the fact that LORETA-Key and many other distributed modeling packages restrict the source space to the cerebral cortex. Consequently, activity generated in the cranial muscles tends to be explained by dipoles fitted to the proximal region of cortex. The Berger maneuver was associated with attenuated current density across the posterior cortex (Figure 9). Furthermore, covariation in neurogenic and myogenic activity significantly altered the magnitude of neurogenic effects in anterior regions of the brain (Figure 9). Similar, albeit less dramatic, distortions were found in posterior regions, particularly for negatively covarying activity.

From these data, independent components were extracted and classified for each participant. Neurogenic components typically accounted for the most variance in scalp electrical activity (median: 25%), whereas those
classified as entirely or predominantly myogenic each accounted for another 8-10%. Participants showed
pronounced variability in the amount of variance accounted for by components exhibiting characteristics of both
myogenic and neurogenic activity (interquartile range for “myogenic-dominant” components: 3-33%; see Figure
6). Descriptively, the choice of which components to exclude had a marked impact on the amount of variance
retained for analysis (Figure 7). Nine different protocols for determining which components to discard were
examined, reflecting the factorial pairing of three for removing non-neurogenic/non-myogenic (NNNM)
components with three for removing EMG. NNNM filtering removed between one-quarter and one-third of the
total variance and EMG correction removed between a tenth and one-third of the total. Together, these two filters
led to the exclusion of as little as one-third to as much as three-quarters of the variance in scalp electrical activity.

Not surprisingly, the validity of these protocols was also quite variable. In the present study, sensitivity
(i.e., attenuation of myogenic effects), specificity (i.e., preservation of neurogenic effects), and correction artifacts
(i.e., generation of effects in the absence of artifact) were each quantitatively assessed using ROIs corresponding to
areas of peak myogenic and neurogenic activation. Results indicated that most of the nine protocols did a
reasonable job removing EMG artifact, evidenced by adequate or excellent sensitivity (Table 1). Unfortunately,
many of these pairings also altered neurogenic activity, evidenced by inadequate specificity or excessive
correction-induced artifact (Table 1). In fact, only two pairings—Maximal-EMG correction combined with
Minimal- or Maximal-NNNM filtering—showed adequate or excellent performance across all three measures.
Furthermore, the Maximal-EMG/Maximal-NNNM pairing tended to outperform the Maximal-EMG/Minimal-
NNNM pairing. None of the nine procedures consistently displayed excellent performance.

On the scalp, the most sensitive and specific procedure for removing EMG artifact from the alpha band was
among the strictest, indexed by the amount of variance discarded (median: 71%). That is, the Maximal-
EMG/Maximal-NNNM pairing entailed rejecting any component containing myogenic signal—including those
where myogenic activity was less prominent than neurogenic activity—in addition to those indexing gross or
ocular artifacts, noise, or unclassifiable low-variance signals. The only other procedure that consistently showed
adequate performance—the Maximal-EMG/Minimal-NNNM pairing—was similarly strict (median variance
discarded: 63%), differing only in the retention of noise and unclassifiable components.
In contrast to the scalp analyses, Maximal-EMG correction paired with Maximal-NNNM filtering was associated with inadequate performance in the intracerebral source-space (Figures 10 and Supplementary Tables 4-6). In particular, it failed to adequately remove EMG when neural activity was fixed and overcorrected neurogenic activity when EMG was absent. Poor sensitivity was also evident for the other procedures examined in the source-space (i.e., Minimal-EMG/Intermediate-NNNM).

Prospects for ICA-Based EMG Correction

As noted in the Introduction, many studies have documented, with varying degrees of rigor, the utility of infomax ICA for attenuating various physical and biological artifacts. Nevertheless, several studies have found ICA to exhibit worse performance than alternative source separation algorithms for ocular (Wallstrom et al., 2004; Romero et al., 2008) and EMG artifacts (Crespo-Garcia et al., 2008; Fitzgibbon et al., 2007). More recently, Debener et al. (2007) showed that ICA displays low specificity for ballistocardiogram artifacts, evidenced by attenuation of event-related neurogenic activity, under some circumstances (Debener, Mullinger, Niazy & Bowtell, 2008). Paralleling Debener and colleagues’ research, the present findings suggest that ICA is a valid means of correcting EMG in some, but not all, cases. On the scalp, some of the ICA-based protocols we tested displayed adequate sensitivity and specificity, whereas other did not. Furthermore, in the intracerebral source-space, even those protocols that showed the most promising performance on the scalp failed.

The inadequate performance of ICA in the source-space likely reflects two factors. First, source modeling, implemented here using the LORETA algorithm and a three-shell head model, makes use of data from all electrodes in the array, not just those in the scalp ROIs. Less-than-perfect EMG correction at even a modest number of electrodes in- or outside of the scalp ROIs could, therefore, exert a substantial influence on the EEG source model. As such, source modeling can be viewed as providing a global check on the quality of EMG correction performed on the scalp, complimenting the more local ROI analyses.

Second, several lines of evidence suggest that ICA failed to adequately separate myogenic from neurogenic sources on the scalp, causing under-correction (low sensitivity) and over-correction (low specificity) in the source-space. In particular,
1. The amount of variance accounted for by “mixed” components, those displaying characteristics of both neurogenic and myogenic activity, was equivalent (e.g., myogenic-dominant vs. pure myogenic components, Figure 6). If ICA had cleanly separated the two sources, one would instead expect the mixed components to be infrequent and to account for little variance.

2. Quantitatively, none of the nine protocols consistently showed excellent performance on the scalp. That is, a reasonably small number of “worst-case” electrodes located inside of the scalp ROIs (Figure 3) evinced under- or overcorrection (Supplementary Tables 1-3).

3. Qualitatively, visual inspection of the topographic maps created following EMG correction indicated significant distortions outside of the scalp ROIs. For instance, significant residual artifact was present at the edge of the array for the corrected myogenic contrast depicted in Figure 2. Likewise, evidence of overcorrection (blue regions) and, to a lesser degree, undercorrection (red regions) was present outside of the defined ROIs when neurogenic and myogenic activity covaried (Figures 4-5).

It is worth emphasizing that these observations cannot be attributed to limitations in the procedure for classifying or rejecting components. The manual classification protocol was detailed (see Supplementary Method), and raters were extensively trained and highly reliable in its application ($\alpha=.98$). In contrast to prior studies, the impact of systematically varying criteria for rejecting both non-myogenic and myogenic components was examined. Of the nine protocols examined, only four consistently showed questionable or better performance on the scalp, and none did so in the source-space. Systematic biases in the classification or rejection of components cannot explain the combination of low sensitivity and specificity in the source-space, whereas a failure to cleanly separate myogenic and neurogenic sources does.

Two factors could plausibly account for the inability of infomax ICA to fully separate myogenic from neurogenic sources. Systematically testing these hypotheses represents a key challenge for future research. First,
inadequate separation might reflect non-optimal specification of the number of components to extract ("model order"). Like most other high-density studies, a combination of PCA and ICA was used to extract fewer components, 64, than the limit imposed by the number of electrodes in the array, 128. As detailed in Footnote 3, the number of components was subjectively optimized using a trial-and-error approach. Model order was also constrained to be identical across participants. Notably, specifying too many components or too few (i.e., over- or under-fitting) might explain poor source separation (Naeem, Brunner & Pfurtscheller, 2009). This possibility could be tested using an information theoretic approach to objectively identify the optimal model order for each participant, as is typical in the functional magnetic resonance imaging literature (Beckmann & Smith, 2004; Calhoun, Adali, Pearlson, & Pekar, 2001; Li, Adali & Calhoun, 2007) and has occasionally been done in the EEG literature (Mouraux & Iannetti, 2009; see also Supplementary Method and Results). Alternatively, the deflationary approach implemented in the fastica package (http://www.cis.hut.fi/projects/ica/fastica) could be used (Mantini, Franciotti, Romani, & Pizzella, 2008). Procedures for optimizing model order on an individual basis would also facilitate the development of criteria for discarding problematic participants (e.g., those with an unusually large number of components). And, from a more practical perspective, application of such procedures would reduce the need to classify large numbers of closely related or uninformative components.

Second, inadequate separation could reflect EMG violating key assumptions of the infomax ICA algorithm (Bell & Sejnowski, 1995). For instance, infomax assumes that the activation of each component is sparse and non-Gaussian. The lengthy blocks of neurogenic and myogenic activation in the present study—and in many studies of emotion (e.g., Coan & Allen, 2003; Davidson et al., 1990; Shackman, Sarinopoulos, Maxwell, Pizzagalli, Lavric & Davidson, 2006)—may not adequately satisfy this assumption. Infomax also assumes that sources are mutually temporally independent. To the degree that neurogenic and myogenic activity are too closely coupled in the time domain they would violate this assumption (but cf. Ohla, Hudry & le Coutre, 2009). Future studies could test the degree to which second-order blind source separation algorithms, that do not require strong assumptions, such as Second Order Blind Identification (SOBI) or Algorithm for Multiple Unknown Signals Extraction (AMUSE), produce better separation (Joyce, Gorodnitsky & Kutas, 2004; Romero et al., 2008; Tang, Liu & Sutherland, 2005). It might also be fruitful to investigate the utility of signal-space projection methods (Nolte & Curio, 1999; Tesche,
important to examine the degree to which these conclusions generalize to paradigms characterized by punctuate bursts of time-locked activation, as is typical of ERSP studies.

**GLM-Based EMG Correction**

The present results would seem to partially contradict prior reports (McMenamin et al., 2009; Shackman et al., 2009) in which the use of one form of GLM-based EMG correction, “epoch-wise regression,” was recommended for scalp analyses. Epoch-wise regression removes epoch-to-epoch variance in alpha-band activity predicted by contemporaneous EMG-band activity (e.g., 70-80Hz) separately for each electrode and participant. To facilitate a more direct comparison of GLM- and ICA-based correction techniques, the validity of epoch-wise regression was re-examined using the identical methods used for testing ICA (see Supplementary Method and Results). Consistent with McMenamin et al (2009), epoch-wise regression exhibited adequate performance across nearly all validation tests on the scalp (Supplementary Tables 7-9), with the lone exception of poor specificity in the presence of positively covarying neurogenic activity. The performance discrepancy between these reports is likely to be a consequence of using much larger neurogenic ROIs for the present analyses (i.e., 7 vs. 23 electrodes). Larger ROIs were employed with the aim of indexing the impact of EMG artifact and EMG correction in regions characterized by less extreme signals (myogenic and neurogenic) with the idea that such signals would be more representative of real-world changes in spectral activity. By contrast, the small, highly focused neurogenic ROI used by McMenamin et al (2009) was blind to distortions outside of areas of peak neurogenic activity.

**Future Challenges**

Three limitations of the present study represent additional avenues for future research. First, the impact of EMG artifact and EMG correction on individual differences in state or trait brain electrical activity remains unknown. In particular, the degree to which either ICA- or GLM-based correction techniques preserve hemispheric asymmetries in tonic (“resting”) frontal activity, a neural marker of individual differences in emotional reactivity
ICA-Based EMG Correction

(Shackman et al., in press; Coan, Allen & McKnight, 2006) and affective disorders (Thibodeau et al., 2006) remains untested.

Second, our conclusions derive from analyses of the alpha band (8-13Hz) during extended blocks of myogenic activity. While this represents a reasonable analog to studies using blocked manipulations of emotion (e.g., threat of shock, emotional films), the degree to which these conclusions generalize to event-related designs or other frequency bands is unclear. Still, it seems likely that the quality of performance for the neighboring theta band (4-8Hz) would be similar to that observed for alpha. In contrast, we anticipate that the quality of correction would be lower for bands, such as gamma (>30Hz), that lie closer to peak EMG activity. Indeed, exploratory analyses indicated poor performance for ICA in the 70-80Hz range (Footnote 5). Finally, the degree to which the present conclusions generalize to stronger or weaker EMG contamination is unknown (Fitzgibbon et al., 2007).

Third, like nearly all prior validation studies, the tests of specificity are founded on the assumption that myogenic activity was absent when participants were instructed to relax (Shackman et al., 2009). If, in fact, modest amounts of EMG were present during this condition, estimates of specificity would be artifactually reduced. We consider this a minor concern; the location of the neurogenic ROI (~Pz) for testing specificity on the scalp (Figure 3) should minimize contributions from the anterior, lateral, and posterior muscle groups (Supplementary Figure 1) when myogenic activity is weak. Nevertheless, it would be informative to validate ICA and other EMG correction techniques using data obtained during neuromuscular blockade (Whitham et al. 2007, 2008).

Recommendations and Conclusions

Consideration of these observations and the extant literature yields four recommendations. First, if widespread EMG artifacts are suspected, Maximal-EMG correction protocol with Maximal-NNNM filtering should be employed. This method exhibited the best combination of sensitivity and specificity across all tests (Table 1). It also outperformed GLM-based EMG correction, which never demonstrated excellent performance in any of our validation tests (Supplementary Tables 7-9). Second, given its merely adequate sensitivity and inconsistent specificity, we no longer recommend the use of GLM-based EMG correction techniques (cf. McMenamin et al., 2009) for studies characterized by widespread EMG artifact. Third, for investigations where
specificity is a smaller concern than sensitivity, any of the Intermediate- or Maximal-EMG correction ICA-based protocols represent reasonable choices. Fourth, the use of distributed modeling techniques, such as LORETA, to estimate the intracerebral sources of spectral EEG in studies with prominent myogenic activity is not recommended. It remains to be seen whether it is reasonable to do so using other approaches, such as dipoles or beamformers (Michel, Murray, Lantz, Gonzalez, Spinelli & Grave de Peralta, 2004; Nazarpour, Wongsawat, Sanei, Chambers & Orantara, 2008).

Recent years have witnessed a resurgence of interest in using scalp-recorded and source-modeled EEG to answer fundamental questions on the neural implementation of cognitive processes (Makeig et al. 2004; Pizzagalli 2007). The continued development and careful validation of techniques for separating myogenic from neurogenic signals will have substantial benefits this endeavor.
ICA-Based EMG Correction

References


ICA-Based EMG Correction


http://psyphz.psych.wisc.edu/~shackman/ajshackman_bis_AcceptedPsychSci2009a.pdf


Clinical EEG, 24(3), 123-126.
Acknowledgements

The first two authors (BWM and AJS) contributed equally to this research. We thank Alanna Clare, Donna Cole, Isa Dolski, and Aaron Teche for assistance and three anonymous reviewers for helpful comments. This work was supported by the NIMH (P50-MH069315 and R37/R01-MH43454 to RJD; BWM was supported by T32-HD007151).
Figure Captions

Figure 1. *Alpha-band contrasts prior to correction.* Topographic maps depict spline-interpolated t-maps for each condition-contrast (columns) and non-neurogenic/non-myogenic (NNNM) artifact filter (rows). There were four conditions, reflecting the factorial manipulation of myogenic (Muscles: Relaxed, Tensed) and neurogenic activity (Eyes: Open, Closed). Contrasts were computed to isolate myogenic (OR - OT), neurogenic (OR - CR), positively-covarying (OT - CR), and negatively-covarying (OR - CT) activity. Note the less extreme values for the myogenic contrast (OR – OT; first column).

Figure 2. *Myogenic contrast (OR-OT) after EMG correction.* Topographic maps depict thresholded p-values at each electrode after applying each method of non-neurogenic/non-myogenic (NNNM) artifact filtering and ICA-based EMG correction. Negative values are depicted in blue (dark-blue: $p < .05$; light-blue: $p < .10$; green: $p > .10$). Note that the row labeled “None” depicts the thresholded OR-OT contrast from Figure 1.

Figure 3. *Scalp regions of interest (ROIs).* Gray circles depict electrodes included in the ROIs.

Figure 4. *Negatively-covarying contrast: Post-correction error.* Topographic maps depict p-values corresponding to the corrected OT-CR minus uncorrected OR-CR contrast. Positive values, indicating an increase in magnitude for the contrast, are shown in red (dark-red: $p < .05$; light-red: $p < .10$; yellow: $p > .10$).

Figure 5. *Positively covarying contrast: Post-correction error.* Topographic maps depict p-values corresponding to the corrected OR-CT minus uncorrected OR-CR contrast. Positive values, indicating an increase in magnitude for the contrast, are shown in red (dark-red: $p < .05$; light-red: $p < .10$; yellow: $p > .10$).

Figure 6. A) *Median frequency and B) variance (in percent) accounted for by each class of components.*

Frequencies and percentages were computed within participants. Medians were then computed across participants. Error bars indicate the 25th and 75th percentiles.
Figure 7. Median variance retained and discarded for each combination of NNNM filter and EMG correction protocol.

Figure 8. Neurogenic contrast: Post-correction error. Topographic maps depict thresholded $p$-values corresponding to the corrected OR-CR minus uncorrected OR-CR contrast. Negative values, indicating an increase in magnitude for the neurogenic contrast, are shown in blue (dark-blue: $p < .05$; light-blue: $p < .10$; green: $p > .10$). Positive values, indicating correction-induced magnitude reduction of the neurogenic contrast, are shown in red (dark-red: $p < .05$; light-red: $p < .10$; yellow: $p > .10$).

Figure 9. Contrasts of interest in the source-space after applying the Intermediate-NNNM filter. A) Myogenic contrast (OR-OT), B) Neurogenic contrast (OR-CR), C) Error induced by negatively-covarying artifact prior to EMG correction ([OT-CR] minus [OR-CR]), and D) Error induced by positively-covarying artifact ([OR-CT] minus [OR-CR]) prior to EMG correction.

Figure 10. Source-space effects of EMG correction. Uncorrected (x-axis) compared to ICA-corrected $t$-values (y-axis) for voxels in the source-space, color-coded by ROI (Blue: Neurogenic ROI; Red: Myogenic ROI; Green: Both ROIs). Points close to the solid horizontal line are voxels where $t$ approached zero after correction, indicating high sensitivity for EMG-contaminated contrasts and low specificity for the EMG-free neurogenic contrast. Conversely, points lying beyond the broken horizontal lines ($p = .05$) were significantly altered by correction. Points along the diagonal were unchanged by correction, indicating low sensitivity for EMG-contaminated contrasts and high specificity for the neurogenic contrast. Note that row B plots the uncorrected OR-CR contrast (x-axis) against the change in OR-CR after correction (i.e. correction induced error; y-axis).
Table 1. Validity Ratings on the Scalp

<table>
<thead>
<tr>
<th>EMG Correction</th>
<th>Myogenic ROI</th>
<th>Neurogenic ROI</th>
<th>Negativity Covarying</th>
<th>Positivity Covarying</th>
<th>Lowest Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal</td>
<td>/ / +</td>
<td>+/ +</td>
<td>/ / +</td>
<td>++ / +</td>
<td>/</td>
</tr>
<tr>
<td>Intermediate</td>
<td>/ / +</td>
<td>+/ **</td>
<td>/ / +</td>
<td>++ / +</td>
<td>?</td>
</tr>
<tr>
<td>Maximal</td>
<td>/ / +</td>
<td>+/ +</td>
<td>/ / +</td>
<td>++ / +</td>
<td>-</td>
</tr>
<tr>
<td>Intermediate</td>
<td>/ / +</td>
<td>/ / +</td>
<td>++ / +</td>
<td>++ / +</td>
<td>-</td>
</tr>
<tr>
<td>Maximal</td>
<td>/ / +</td>
<td>/ / +</td>
<td>++ / +</td>
<td>++ / +</td>
<td>-</td>
</tr>
<tr>
<td>Maximal</td>
<td>/ / +</td>
<td>/ / +</td>
<td>++ / +</td>
<td>++ / +</td>
<td>-</td>
</tr>
<tr>
<td>Maximal</td>
<td>/ / +</td>
<td>/ / +</td>
<td>++ / +</td>
<td>++ / +</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: *Results from the myogenic contrast (corrected OR-OT) in the myogenic and neurogenic ROI. *The Corrected OR-CR vs. Uncorrected OR-CR contrast. *Corrected GT-CR vs. uncorrected OR-CR. *Corrected OR-CT vs. uncorrected OR-CT. *Ratings from each ROI: + poor, ? questionable, ++ adequate, ** excellent (see Method).
Figure 1. Alpha-band contrasts prior to correction.
Click here to download 5. Figure: baselines_alpha.eps

- Minimal NNNM
- Intermediate NNNM
- Maximal NNNM

OR-OT  OR-CR  OT-CR  OR-CT
Figure 2. Myogenic contrast (OR-OT) after EMG correction.

Click here to download 5. Figure: orotThresh_Alpha.eps
Figure 3. Scalp regions of interest (ROIs).
Click here to download Figure: EMG_alpha_ROIs.eps
Figure 4. Negatively-covarying contrast: Post-correction error.
Click here to download 5. Figure: otcrThresh_Alpha.eps
Figure 5. Positively covarying contrast: Post-correction error.
Click here to download 5. Figure: orctThresh_AAlpha.eps
Figure 6. A) Median frequency and B) variance (in percent) accou
Click here to download 5. Figure: fig6.eps
Figure 7. Median variance retained and discarded for each combination of NNNM Filter and EMG Protocol.
Figure 8. Neurogenic contrast: Post-correction error.
Click here to download 5. Figure: orcrThresh_Alpha.eps
A Uncorrected myogenic effect (OR-OT)

B Uncorrected neurogenic effect (OR-CR)

C Uncorrected error for negatively covarying EMG (OT-CR)-(OR-CR)

D Uncorrected error for positively covarying EMG (OR-CT)-(OR-CR)
Figure 10. Source-space effects of EMG correction.

Click here to download 5. Figure: voxScatterFig.eps
6. Supplementary Material
Click here to download 6. Supplementary Material: EMGFIX2_Rev1_supplement.doc