



ELSEVIER

Clinical Neurophysiology 118 (2007) 2591–2604



www.elsevier.com/locate/clinph

Independent component analysis and clustering improve signal-to-noise ratio for statistical analysis of event-related potentials

Philip M. Zeman^{a,b,*}, Bernie C. Till^{a,b}, Nigel J. Livingston^a,
James W. Tanaka^c, Peter F. Driessen^b

^a *CanAssist, University of Victoria, BC, Canada*

^b *Department of Electrical and Computer Engineering, University of Victoria, BC, Canada*

^c *Department of Psychology, University of Victoria, BC, Canada*

Accepted 7 September 2007

Available online 29 October 2007

Abstract

Objective: To evaluate the effectiveness of a new method of using Independent Component Analysis (ICA) and k-means clustering to increase the signal-to-noise ratio of Event-Related Potential (ERP) measurements while permitting standard statistical comparisons to be made despite the inter-subject variations characteristic of ICA.

Methods: Per-subject ICA results were used to create a channel pool, with unequal weights, that could be applied consistently across subjects. Signals derived from this and other pooling schemes, and from unpooled electrodes, were subjected to identical statistical analysis of the N170 own-face effect in a Joe/No Joe face recognition paradigm wherein participants monitored for a target face (Joe) presented amongst other unfamiliar faces and their own face. Results between the Joe, unfamiliar face and own face conditions were compared using Cohen's *d* statistic (square root of signal-to-noise ratio) to measure effect size.

Results: When the own-face condition was compared to the Joe and unfamiliar-face conditions, the channel map method increased effect size by a factor ranging from 1.2 to 2.2. These results stand in contrast to previous findings, where conventional pooling schemes failed to reveal an N170 effect to the own-face stimulus (Tanaka JW, Curran T, Porterfield A, Collins D. The activation of pre-existing and acquired face representations: the N250 ERP as an index of face familiarity. *J Cogn Neurosci* 2006;18:1488–97). Consistent with conventional pooling schemes, the channel map approach showed no reliable differences between the Joe and Unfamiliar face conditions, yielding a decrease in effect size ranging from 0.13 to 0.75.

Conclusions: By increasing the signal-to-noise ratio in the measured waveforms, the channel pool method demonstrated an enhanced sensitivity to the neurophysiological response to own-face relative to other faces.

Significance: By overcoming the characteristic inter-subject variations of ICA, this work allows classic ERP analysis methods to exploit the improved signal-to-noise ratio obtainable with ICA.

© 2007 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.

Keywords: EEG; ICA; Signal-to-noise ratio; Effect size; Channel pooling; N170

1. Introduction

Responses to experimental manipulations, measured by Event-Related Potentials (ERPs), are considered distinct

when ERP peaks differ significantly in amplitude, latency, and/or scalp topography (Dien and Santuzzi, 2005). The effects of an experimental manipulation can be made more visible by employing any of a wide variety of signal processing techniques, such as spatial-, frequency- and/or time-domain filtering, to separate the signal of interest from background noise. A key issue in the design of such filters is consistency – the ERP data must be processed consistently across subjects and conditions to avoid

* Corresponding author. Address: Department of Electrical and Computer Engineering, University of Victoria, BC, Canada. Tel.: +1 250 589 4234; fax: +1 250 721 6611.

E-mail address: pzeman@ece.uvic.ca (P.M. Zeman).

introducing processing artefacts as a side effect of enhancing the measurability of the *bona fide* effects under study. Dimensionality-reduction techniques, such as Independent Component Analysis (ICA), can be very effective at increasing signal-to-noise ratio, but typically exhibit variations of scalp topography across subjects (Delorme and Makeig, 2004), which preclude the consistent use of standard statistical analysis techniques, such as ANOVA. We report a method of using ICA which restores consistency while retaining the increased signal-to-noise ratio.

Measurability of an effect is enhanced by increasing the signal-to-noise ratio, which amounts to increasing the effect size. This results either from an increased difference in the means of some statistic used to distinguish between experimental conditions, or from a reduced variance of that difference, or both. Since increasing the difference between the means is usually considered infeasible for ERP measurements, the reduction of variance is typically pursued and most often achieved by increasing the number of measurements taken (Tabachnick and Fidell, 1989).

Variance across trials can be reduced by increasing the number of trials. Indeed, trial-averaging of event-locked potentials is commonly used to suppress random background EEG while emphasising the EEG due to the cognitive processes under study. However, the number of useful trials is usually limited by habituation, fatigue and other extraneous effects that tend to increase with experiment duration. For example, in a virtual Morris Water Maze paradigm (Sorensen et al., 2006), the behavioural implications of repetition limit reliable data collection to only 30 trials or so. Similarly, variance across subjects can be reduced by increasing the number of subjects. However, it is often impractical even to double the number of subjects in many experiments due to scarcity of suitable subjects, limitations on total time available, and so on. In any case, doubling the number of subjects would only decrease the variance by a factor of 2. In contrast, this paper describes an example in which the proposed method increases the effect size by as much as 2.2 times. In terms of the constant-mean techniques discussed above, this is equivalent to reducing variance by a factor of 5 without requiring any additional data collection.

Another way of reducing variance involves channel pooling, or averaging of measurements taken from adjacent electrodes. Pooling is also often employed to simplify ERP analysis by reducing the number of degrees of freedom. In some cases, pooling may reduce problems associated with statistical testing of channels having correlated activity, on the assumption that correlation decreases with distance between electrodes. Generally, channels are selected for pooling based on commonly accepted practice, *ad hoc* observation, or channel \times condition interaction effects. Unfortunately, channel pooling often leads to blurring of amplitude effects, which attenuates statistical measures of difference between conditions (Dien and Santuzzi, 2005). Therefore, many ERP components are generally accepted as being best measured at particular

individual electrodes. The N170 ERP peak, which we use as our example, can be measured with or without pooling in face recognition paradigms. It has been measured from T5 and T6 (Rossion et al., 2000), or from P7 pooled with PO7 and P8 pooled with PO8 (Gauthier et al., 2003).

In this paper, we present a more general concept of pooling that avoids the shortcomings of conventional pooling while retaining the advantages, such as reduced variance and decreased sensitivity to outliers. In generally-accepted practice, pooling is applied to electrodes which are in adjacent positions on the scalp, or symmetrically located with respect to the midsagittal plane, and pooling is discrete; channels either do or do not belong to a given pool. In contrast, our approach permits partial membership of channels that are not necessarily adjacent and may even have opposite polarities. We assign a weight between -1 and $+1$ to each channel, which defines that channel's degree of membership in the channel pool. We refer to the set of all weights for a given pool as the channel map of that pool. In this paper, we demonstrate the construction and use of such channel maps to enhance the measured signal corresponding to a particular ERP. This signal is then used to evaluate the ERP amplitude peak uniformly across subjects in order to quantify differences among experimental conditions.

1.1. Related work

In some cases, channel selection reduces to a simple choice of reference electrodes (McFarland et al., 1997) to keep the reference far from the brain region of interest, determined *a priori*. For Brain-Computer Interface applications (Lal et al., 2004), channel selection is sometimes accomplished with the aid of clustering methods (Montgomery et al., 2005), but this differs from the present work in that the focus is usually on finding the minimal set of electrodes required to discriminate between a small number of conditions (typically two) whereas our aim is to make the most effective channel map possible, using all of the available electrodes.

A commonly-used method of separating an ERP into components is Principal Component Analysis (PCA), which decouples estimated sources based on their second-order statistics. The first component identified by PCA accounts for the largest fraction of the total energy of an ERP peak and is therefore usually physiologically quite plausible. Unfortunately, the physiological plausibility of the remaining PCA components becomes increasingly compromised by the restriction that each component be orthogonal to every previous component. Statistics beyond second-order are not used by PCA, and this leads to misallocation of variance and indeterminate factor interactions – in addition to which, PCA suffers from latency jitter (Achim and Marcantoni, 1997; Dien, 1998). Kayser and Tenke (2003) performed a systematic exploration of these issues and determined analysis methodologies which minimize the problems as much as possible, but found that the

problems can only be mitigated, not eliminated, because the limitations are inherent to the formal mathematical structure of PCA. In contrast, ICA implicitly takes into account statistics of all orders (Hyvärinen et al., 2001) to find components exhibiting maximal statistical independence, and thus the results of ICA reflect the structure of the sources themselves much more closely than do the results of PCA.

The approach which bears, perhaps, the strongest resemblance (in some of its more recent refinements) to that adopted in this paper is the Topographic Components Model (TCM), originally proposed by Möcks (1998a,b). This is a trilinear model, wherein the multi-channel EEG trace for each subject is modeled as a sum of terms, each of which is a product of three factors. The three factors composing each term are a scalp topography, a time course and a so-called “loading factor”. The loading factors indicate the degree to which the individual topographic ERP components, each represented by the product of a scalp topography and a time course, are present in the data. Thus the model represents a multi-channel EEG data set as a weighted sum of topographic ERP components.

Application of TCM to discrimination between normal and abnormal visual evoked potentials (Field and Graupe, 1990, 1991a) and classification of subjects based on the auditory P300 (Field and Graupe, 1991b) showed that TCM can give a better fit to real EEG data, using fewer parameters, than can be obtained with PCA. However, latency jitter proved to be no less a problem for TCM than for PCA, and the applications of TCM to real neurophysiological data are subject to a number of criticisms, not least among which is the fact that additional orthogonality constraints are required to obtain a solution. By way of remedy, Achim and Bouchard (1997) extended the model to permit variations in latency and time-scale across conditions. Their approach begins with Singular Value Decomposition (SVD) of concatenated data sets to produce topographies and initial estimates of waveforms for the components, followed by an iterative adjustment of amplitude, onset and duration parameters extracted from the waveforms by spline interpolation. The results reported are obtained only on simulated data sets and have not been generalized to more than two experimental conditions or more than one subject.

A relatively recent extension to TCM (Wang et al., 2000) relaxes the strict one-to-one correspondence between topographies and waveforms, which is characteristic of earlier versions of TCM, and permits the number of topographies to differ from the number of waveforms. In this approach, SVD is applied to a concatenation of EEG traces, each of which is obtained by averaging all trials for a given subject and condition. Topographies are determined by spatial SVD, and waveforms are determined by temporal SVD. Thus, though the use of SVD again leads to orthogonal waveforms, the net activation of a topography is expressed as a weighted sum of waveforms using a matrix of loading factors that need not be diagonal, and

therefore the activations of any pair of topographies will not generally be orthogonal. In order to obtain a unique solution, however, the components must be rotated such that averaging the loading matrices of all subjects results in a diagonal matrix. Once again, a constraint is imposed to make the model mathematically tractable at the expense of neurophysiological plausibility.

One characteristic which is common to PCA, SVD and ICA is that they estimate time courses and spatial profiles separately for each subject to minimize departures from spatial stationarity, and consequently fail to provide a framework for making the kind of between-subjects comparisons facilitated by TCM. One of the key objectives of the present work, therefore, is to combine the advantages of TCM and ICA. This is made possible by the fact that many of the topographies corresponding to the independent components bear a marked similarity to one another between subjects, which suggests clustering and averaging within clusters after ICA as a method of creating components which capture what is common to all subjects. Contreras-Vidal and Kerick (2004) clustered components using both k-means clustering and hierarchical clustering to make cross-subject comparisons. Using scalp topographies and spectral characteristics of ICA-derived sources as clustering criteria, they demonstrated that sources common across subjects will be clustered together while sources unique to specific subjects are highly unlikely to be placed in clusters.

A principle benefit of ICA is that it does not limit analysis to well-known ERP peaks, but permits mining of data for statistically independent sources. ICA not only provides for data-driven component selection but further enables analysis of events that are not prominent in averaged data. The scalp topographies provided by ICA are independent of time, and analysis is therefore relatively unhindered by latency jitter between trials. ICA has been demonstrated (Makeig et al., 2000) to be effective for recovering the waveforms and scalp topographies of statistically independent sources explicitly constructed to simulate the characteristics of the EEG signals typically encountered in ERP analysis. The statistical independence of the sources found by ICA dramatically reduces the need for repeated-measures corrections to mitigate the increased probability of type-I errors which would otherwise result from the application of multiple statistical tests to adjacent electrodes with partially correlated activity. In numerous studies using real data (e.g. Jung et al., 2001; Makeig et al., 2004; Onton et al., 2005), ICA has been used successfully to separate EEG into components. These are the main reasons why our work utilizes ICA.

We use ICA, clustering and averaging to construct channel maps which are, in effect, spatial filters derived from the data. We use the channel map corresponding to a given component to emphasise specific features in ERP data representing multiple conditions, collected from multiple subjects. Unlike methods based on statistical measures of ICA source waveforms (Delorme and Makeig, 2004), we use the

components found by ICA purely for constructing channel maps, with the objective of increasing effect size consistently across subjects in support of standard statistical analysis methods.

1.2. ICA in theory

The basic concept of ICA is perhaps best understood by considering the so-called cocktail party problem. There are n people in a room, all speaking simultaneously, and each voice is a source signal, $s_i(t)$. There are also m microphones situated about the room, and each one produces an observed signal, $x_i(t)$. This is a latent-variables model; the source signals are not observable in isolation because each microphone responds to some unknown mixture of the sources. The task, then, is to recover the $s_i(t)$ knowing only the $x_i(t)$.

In using ICA to solve this problem, it is assumed that the observed signals can be expressed as instantaneous, linear, time-invariant mixtures of statistically independent sources which are either non-Gaussian or non-white or non-stationary. Specifically, one introduces the model

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \quad (1)$$

where $\mathbf{x}(t)=[x_1(t), x_2(t), \dots, x_m(t)]^T$, $\mathbf{s}(t)=[s_1(t), s_2(t), \dots, s_n(t)]^T$, and the unknown $m \times n$ matrix, \mathbf{A} , is called the mixing matrix. Each row of \mathbf{A} determines how the sources combine to form a particular observed signal, and each column of \mathbf{A} determines how a particular source is distributed among the observed signals. This situation is depicted schematically in the online-only Figure S1.

Some of the assumptions leading to this example are somewhat unrealistic in the context of the cocktail party because, for instance, people will tend to move about the room and the room acoustics will exhibit significant delays and reverberations, which necessitate the treatment of the a_{ij} as time-dependent impulse response functions. On the other hand, the assumptions allowing the a_{ij} to be treated as simple, real constants are rather well satisfied by EEG signals. In this case, the observed signals are just the voltages on the electrodes, and the underlying sources are the potentials generated by the neural activity of anatomical structures in the brain which, for each subject, reside at fixed, though possibly unknown, locations within the skull. At the time and distance scales of interest in electroencephalography, electromagnetic propagation delays are entirely negligible and biological tissues behave essentially as linear isotropic media (Makeig et al., 1996). In addition, the statistical properties of EEG signals satisfy all of the desiderata: they are non-Gaussian and non-white and non-stationary, all at the same time (Parra and Sajda, 2003).

Eq. 1 is deceptively simple, because *both* of the factors on the right hand side are to be estimated. This results in two irreducible indeterminacies. First, any of the $s_i(t)$ may be multiplied by any nonzero constant, because the inverse of that constant can be absorbed into the corresponding column of \mathbf{A} without changing the left hand side

of the equation in any way. A commonly adopted convention is to scale the $s_i(t)$ to unit variance, but even so, a factor of ± 1 remains, which cannot be resolved. Second, the $s_i(t)$ may appear in any order, though this can be rendered irrelevant by adopting the convention that the $s_i(t)$ are ordered such that the magnitude of the columns of \mathbf{A} decreases with increasing column index.

Without loss of generality, the $x_i(t)$, and therefore the $s_i(t)$, may be assumed to have zero mean, but a number of additional assumptions are required in order to make this problem soluble. First, the columns of the mixing matrix must be linearly independent. Second, there must be at least as many signals as there are sources to be estimated: $m \geq n$. And finally, the sources are assumed to be statistically independent, which means that

$$p(s_1(t), s_2(t), \dots, s_n(t)) = p_1(s_1(t))p_2(s_2(t)) \dots p_n(s_n(t)), \quad (2)$$

where p is the joint distribution and the p_i are the marginal distributions of the $s_i(t)$. This should not be confused with the more commonly encountered condition of vanishing correlation,

$$\langle s_i(t)s_j(t) \rangle = \langle s_i(t) \rangle \langle s_j(t) \rangle, \quad i \neq j \quad (3)$$

which follows from it. In fact, Eq. 2 implies the much stronger condition

$$\langle f(s_i(t))g(s_j(t)) \rangle = \langle f(s_i(t)) \rangle \langle g(s_j(t)) \rangle, \quad i \neq j \quad (4)$$

for any measurable functions, f and g (Papoulis, 1991). It should be noted, however, that Eqs. 3 and 4 are completely equivalent for Gaussian sources, and therefore at most one of the p_i may be Gaussian.

Unfortunately, it is not generally possible to utilize Eq. 2 directly. In practice, therefore, the process reduces to the task of iteratively finding a linear transform, \mathbf{W} , such that

$$\mathbf{s}(t) = \mathbf{W}\mathbf{x}(t) \quad (5)$$

and the components of the left-hand side, the $s_i(t)$, are as independent as possible, in the sense that some measure of independence is maximized. The matrix, \mathbf{W} , is called the unmixing matrix. It can be shown (Comon, 1994) that such “maximally independent” components approach strict independence, and \mathbf{W} approaches \mathbf{A}^{-1} asymptotically, when $m = n$. Closely related to this fact is the empirical observation that the greater the number of sensors, the higher the statistical independence among the calculated sources. Essentially, this is because the ICA algorithm has a greater number of degrees of freedom to work with.

The various approaches to ICA differ primarily in which measure of independence is chosen to be maximized, and how the maximization is accomplished. One of the earliest approaches was based on maximum likelihood estimation (Pham et al., 1992). Another approach, called InfoMax (Bell and Sejnowski, 1995), maximizes the information flow through a neural network implementation of ICA. Alternatively, one may minimize the Gaussianity (Hyvärinen and Oja, 1997) of the estimated sources or minimize the mutual information between them (Cardoso, 1999). In fact, the

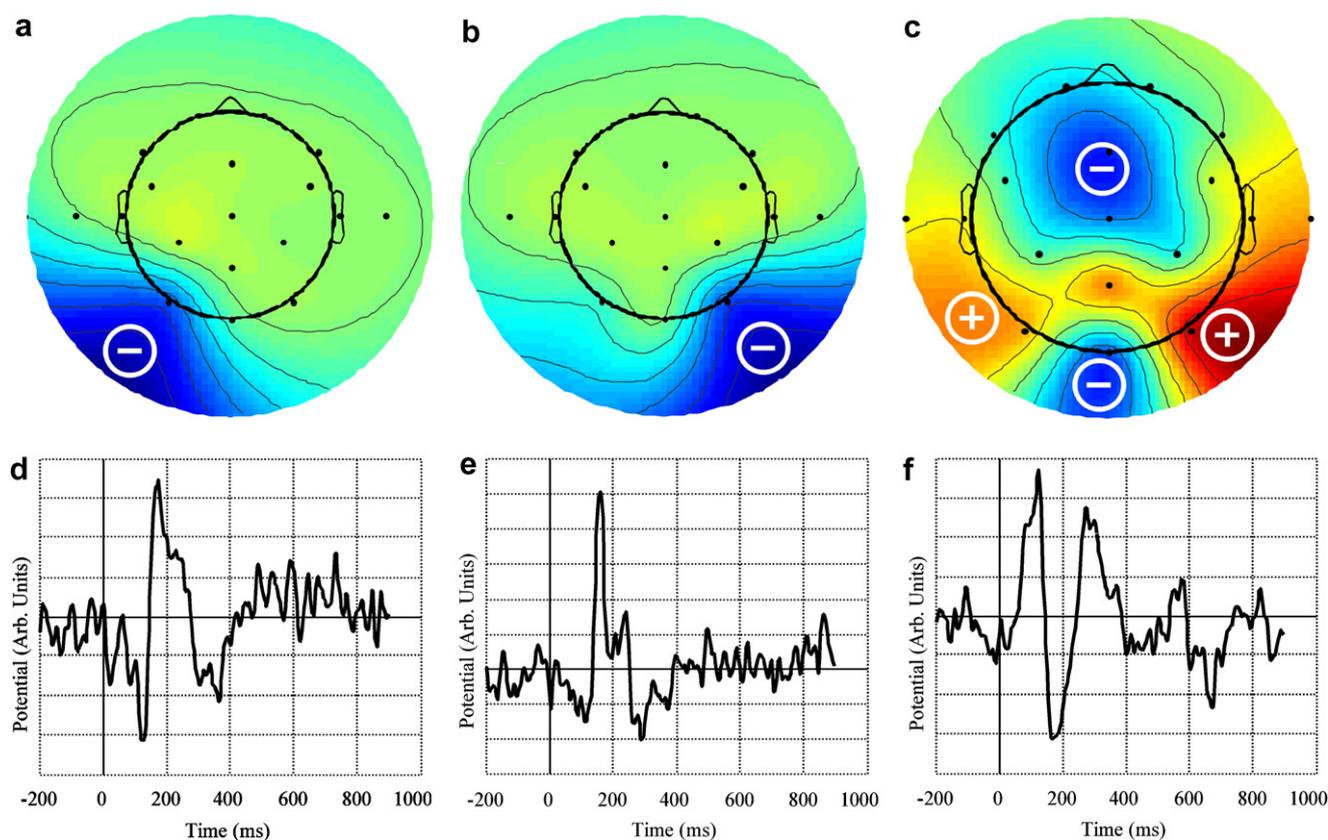


Fig. 1. The subtle relationships among components within and between subjects. (a) Channel map $\mathbf{a}_1^{(1)}$, (b) channel map $\mathbf{a}_2^{(1)}$, (c) channel map $\mathbf{a}_3^{(2)}$, (d) source $s_1^{(1)}$, (e) source $s_2^{(1)}$, (f) source $s_3^{(2)}$.

approach one chooses is largely a matter of taste, because it can be shown (Lee et al., 2000) that all these approaches are essentially equivalent.

Noise sources are not explicitly modeled in Eq. 1, but this is not a major issue in practice (Makeig et al., 1996) because any noise which is present should be extracted as one or more of the independent components. In consequence, particularly in EEG work, it is quite possible for the largest-magnitude component to be an artefact component. EEG data are generally not averaged prior to performing ICA, so that all available information can be treated in a uniform manner. Classification of components as *bona fide* sources or noise is done after the fact. This is a strength of the ICA methodology because it allows for rejection of artefacts that would not otherwise be separable from the data, and the extraction of sources that would otherwise be overwhelmed by strong artefacts. It does, however, place the onus on the experimenter to carefully scrutinize the components returned by ICA and decide which ones are, in fact, noise.

1.2.1. ICA in practice

Fig. 1 shows some of the results obtained via independent component analysis of data acquired in the experiment described in the next section. We present them here to illustrate some of the subtleties involved in interpreting

the results of ICA. Each vertical pair of panels in Fig. 1 (a/d, b/e, and c/f) is a geometric representation of the schematic depiction in the online-only Figure S1(b). The lower panel of this figure (d, e, or f) shows the time course, or waveform, of one of the sources, $s_i(t)$, extracted from EEG data by ICA; the upper panel (a, b, or c) shows how that source projects to the scalp, and is constructed as follows. Each element of the i th column of the mixing matrix, a_{ij} , is placed at the location of the electrode which measures signal, $x_j(t)$. This is a geometric layout of a channel map. Finally, the spaces between the electrode locations are interpolated to make a continuous scalp topography out of the discrete channel map. The largest positive values map to the deepest shades of red¹ (indicated by (+)), the most negative values map to the deepest shades of blue (indicated by (-)), and values near zero map to shades of green.

At each instant in time, the product of the source waveform and the associated channel map gives the contribution of that source to the potential at every point on the scalp. Taken over the entire time interval, this product is called a back-projected source in the ICA literature, and is known as a topographic component in the

¹ For interpretation of the references to color in this figure, the reader is referred to the web version of this paper.

TCM literature. The channels of a back-projected source will vary together with time and demonstrate perfect covariance, with each channel having the same waveform multiplied by a unique constant, either positive or negative.

Panels a/d and b/e show two different components from one subject, while panels c/f show a single component from another subject. Careful consideration of the relationships among them highlights some practical ramifications of independent component analysis, and demonstrates why some degree of visual inspection is completely unavoidable when dealing with components extracted by ICA.

Referring to Fig. 1, we see that, in all three waveforms, there is a peak at approximately 170 ms, flanked by two peaks of the opposite polarity. The time courses of sources $s_1^{(1)}$ and $s_2^{(1)}$ strongly resemble one another to within additive noise, and resemble the negative of $s_3^{(2)}$, also to within additive noise. We express this by writing

$$s_1^{(1)} \approx s_2^{(1)} \approx -s_3^{(2)} \quad (6)$$

where parenthesised superscripts identify the subject and subscripts identify the component.

The channel maps display an even more intriguing structure. In this case, it is the sum of $\mathbf{a}_1^{(1)}$ and $\mathbf{a}_2^{(1)}$ which resembles the negative of $\mathbf{a}_3^{(2)}$, again to within additive noise. That is,

$$\mathbf{a}_1^{(1)} + \mathbf{a}_2^{(1)} \approx -\mathbf{a}_3^{(2)} \quad (7)$$

Taken together, these two equations allow us to write, at each instant of time,

$$s_1^{(1)} \mathbf{a}_1^{(1)} + s_2^{(1)} \mathbf{a}_2^{(1)} \approx s_3^{(2)} \mathbf{a}_3^{(2)} \quad (8)$$

Evidently, the ICA algorithm found more-or-less the same bilateral component in the two subjects, but split the component further into left- and right-hemisphere contributions in the first subject, while the irreducible sign ambiguity was resolved differently in the second subject than in the first. This illustrates the fact that the sign ambiguity is inconsequential as long as one takes the product of the waveform and the channel map, and makes all comparisons between subjects at the level of back-projected topographic components. The application of the investigator's judgement at this point is unavoidable, because there is no mathematical reason to prefer 1×1 over $(-1) \times (-1)$. Only neurophysiological insight can provide that preference.

In summary, care must be taken when comparing ICA results between subjects. Each subject's functional brain anatomy will be at least somewhat different (for example, how tightly left and right occipital areas are coupled for a particular behaviour), and noise sources will also vary appreciably between experimental set-ups for different subjects, as will the non-task-specific neural background activity of each subject. Because ICA is rather sensitive to such differences and assumes spatial stationarity, it must be

applied to each subject individually. This introduces the potential for resolving the sign ambiguity differently from one subject to the next, which creates the need for the type of visual inspection and analysis described above, but such analysis becomes rapidly intractable with increasing numbers of subjects and experimental conditions, requiring the use of software to automate as much as possible of the process.

As of version 5.0 beta, the EEGLab software package (Delorme and Makeig, 2004) includes clustering of components based on topographies, modelled dipoles, spectral features and time-domain characteristics of extracted sources. The underlying assumption is that sources with similar scalp topographies and waveform characteristics are due to the same neurophysiology across subjects, and that differences within any cluster are due to slight anatomical variations among subjects. The methods offered in EEGLab make comparisons of waveforms across subjects, calculated using a different spatial filter and sphering matrix for each subject, and no mechanism is provided to include subjects in the statistical analysis for which ICA failed to find a component that could reasonably be included in a cluster. Thus the EEGLab clustering documentation (Delorme et al., 2006) explicitly warns that, "claims to discovery of physiological facts from component clustering should be accompanied by thoughtful caveat and, preferably, by results of statistical testing against a stable null hypothesis". However, no generally available software package creates channel pools from clustered ICA components so that standard statistical methods can be used for ERP analysis.

Our method differs from those available in EEGLab in that we combine ICA results from individual subjects to form canonical channel maps applicable to the entire group, which satisfies the consistency requirement outlined in the introduction and allows statistical tests to be made among experimental conditions across multiple subjects. To demonstrate this method, we use the "Joe/No Joe" task to examine the N170 face recognition ERP component (Tanaka et al., 2006) in which participants monitor for a target face (i.e., "Joe") presented among other unfamiliar faces and the participant's own face.

2. Methods

2.1. Participants

Twelve graduate students from the University of Victoria, six males and six females, aged 23–55 (31 ± 11) years, participated in this study. Eleven were right handed and one was left handed. All had normal or corrected-to-normal visual acuity and normal colour vision and none had a history of neurological disorders. Permission for data collection was granted by the University of Victoria Ethics Committee and informed consent was given by each participant.

2.2. Apparatus and stimuli

Stimuli consisted of gray-scale images of human faces, digitized at a resolution of approximately 120×180 pixels, presented on a 17-in. 1024×768 CRT monitor using E-Prime experiment control software (Psychological Software Tools, Pittsburgh, PA). These stimuli were viewed from a distance of about 70 cm in a quiet, dimly lit room, and subtended viewing angles of about 3.3° horizontally and 5° vertically.

Each participant viewed 12 different faces. One of these was the participant's own face, and the remaining 11 were completely unfamiliar faces prior to the experiment. One of the unfamiliar faces, of the same gender as the participant, was designated as the target face and named "Joe" or "Jane." Target faces were counterbalanced so that each of the unfamiliar faces served as the target face for each participant.

2.3. Experimental procedures

Subjects were seated comfortably in a chair and introduced to their target faces. Each trial began with the display of a fixation mark in the form of a '+' sign, which remained at the centre of the screen for 500 ms. A face image was presented immediately afterwards, and remained visible for 500 ms. This stimulus was followed by a blank screen for another 500 ms. Following the post-stimulus delay, the participant was visually prompted for a response by the single-word question "Joe?" (or "Jane?" according to gender), which was displayed for one full second in order to give the participant ample time to respond. The required response was to press a button with the left or right index finger (randomized between participants) if the picture was Joe/Jane, and otherwise to press another button with the index finger of the opposite hand. Immediately following the response, or after the maximum time allowed for a response had elapsed, the screen was again blanked for an inter-trial delay of one second. Participants were given a self-timed rest between sets, and longer breaks between blocks, while electrode impedances were checked and corrected if necessary. Each face appeared 70 times for a total of 840 trials. The trials were divided into four blocks and the order of presentation was random within blocks, with the proviso that no image ever appeared twice in a row.

2.4. Electrophysiological recordings

Electroencephalograms were recorded using a montage of 21 electrodes, placed at locations selected from the extended international 10–20 system (Jasper, 1958) using electrode caps fitted for each participant. Signals were acquired using the QuickAmp ERP tool kit (Brain Products GmbH, Gilching, Germany) and digitized at a rate of 250 samples per second to 22-bit accuracy, giving a resolution of 71.5 nV/bit and a full-scale range of ± 150 mV.

Digitized signals were recorded to disk using Brain Vision Recorder software (Brain Products GmbH, Gilching, Germany). Electrode impedances were maintained below 5 K Ω . A linked-earlobe reference was used, and electrooculogram (EOG) recordings were taken for artefact removal. Horizontal EOG was recorded from the external canthi of both eyes, while vertical EOG was recorded from the sub-orbital area of the right eye and referenced to Fp2.

2.5. Data analysis

The three experimental conditions were designated Own, Joe and Other. Selection of the Joe face and one face to represent the Other condition was alternated across subjects. The data were examined for statistical evidence of distinctions between these conditions in the N170 ERP component, against the null hypothesis that no such distinctions existed. This analysis was repeated eight times, using waveforms generated from the data in different ways. First (case 1), using a channel map constructed as described in the next section, which exhibited a pronounced bilateral symmetry. Second (cases 2 and 3), this channel map was decomposed, as described in a subsequent section, into a pair of channel maps with mirror symmetry across the midsagittal plane. As bases for comparison, channels known to be relevant to the N170 were pooled in three different ways. First (case 4), PO7 and PO8 were pooled together to form a bilaterally symmetric pool of the sort one might use to suppress hemispheric differences so as to allow examination of ERP components common to both hemispheres. Second (cases 5 and 6), PO7 and PO8 were considered separately; that is, unpooled. Third (cases 7 and 8), PO7 and PO8 were pooled with their nearest neighbours (PO7 with CP3 and PO8 with CP4), forming a pair of pools with mirror symmetry across the midsagittal plane, such as one might use in an attempt to reject noise while retaining hemispheric differences. These eight cases were distinguished *only* by the pooling method used.

The data were pre-processed identically in all cases, using Brain Vision Analyzer (Brain Products GmbH, Gilching, Germany); the procedure was typical for ERP analysis of early face-recognition ERP components (Tanaka and Porterfield, 2001). Epochs of 1100 ms length, beginning 200 ms prior to stimulus onset and extending to 900 ms afterwards, were segmented off-line. The digitized signals were filtered using a 4th order digital Butterworth filter having a pass band of 0.1–40 Hz. Trials with eye-movement artefacts were smoothed using the algorithm described in Gratton et al. (1983); trials with amplitudes exceeding ± 100 μ V or amplitude steps exceeding ± 50 μ V were excluded. The data were then averaged across trials for each subject and condition to generate a standard ERP trace for each electrode.

At this stage, the pooled waveforms were calculated quite differently in each of the eight cases, as described above. After pooling, however, data analysis continued identically in all cases.

Each pooled waveform was baseline corrected by subtracting the mean value of the signal in the 200 ms interval immediately preceding stimulus onset. N170 amplitudes were extracted from the trial-averaged waveforms for each subject by finding the most negative peak within a window of ± 50 ms, centred at 170 ms post-stimulus. For all subjects, this peak was visibly distinguishable from other large-amplitude variations in the ERP waveform. Means and variances of these amplitude peaks were then calculated across the twelve subjects for each experimental condition.

Repeated-measures 1-way ANOVA was applied to the bilaterally symmetric cases (cases 1 and 4 above), to assess the significance of differences between conditions while factoring out variations between subjects. In the remaining cases, repeated-measures 2-way ANOVA was used to evaluate the effects of condition, hemisphere, and condition \times hemisphere interactions. As usual, a difference was taken to be significant if the probability was less than 0.05 that it was due to random effects. In addition, differences between experimental conditions were quantified using Cohen's d statistic (Cohen, 1992), which measures effect size by expressing the distance between the means of a pair of distributions in units determined by the standard deviation of that distance. Suppose we measure the value of some random variable under two different experimental conditions, and we obtain values of $\mu_1 \pm \sigma_1$ and $\mu_2 \pm \sigma_2$, where μ_1 and μ_2 are the means for the conditions and the corresponding standard deviations are σ_1 and σ_2 . Then the distance between the two means, Cohen's d , is defined as

$$d = \frac{\mu_1 - \mu_2}{\sigma}, \quad \text{where } \sigma = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}. \quad (9)$$

Evidently, Cohen's d statistic is just the square root of the signal-to-noise ratio. As such, it serves as an excellent basis for comparison between situations wherein the means and standard deviations are not directly comparable – such as between the proposed and conventional pooling methods. To provide a familiar basis of comparison for the effect sizes measured by Cohen's d , paired t -tests were also performed.

2.6. ICA-derived channel maps

The procedure for constructing a channel map responsive to the posterior N170 ERP component, and applicable across all subjects and conditions, is outlined in Fig. 2. The ICA implementation distributed with EEGLab v4.512 software (Delorme and Makeig, 2004) for MatLab was used to identify 18 maximally independent components for each of the 12 subjects. The only pre-processing used to prepare the data for ICA was 0.1–40 Hz band-pass filtering and average referencing. The ocular channels (LHEOG, RHEOG and VEOG) were excluded from the analysis, leaving 18 channels and thus limiting the number of sources to be

found by ICA to at most 18. In order to focus the analysis on activity occurring in the physiologically relevant time interval, the first and last 200 ms were trimmed from the dataset for each trial. For each of the 12 subjects, the shortened trials for all conditions (70 trials \times 3 conditions) were concatenated and subjected to ICA, so that the ICA algorithm could respond to the contrast between sources, both within and between experimental conditions (Makeig et al., 2000). This resulted in an 18×18 mixing matrix for each subject. In each matrix, each column represented the scalp topography of one source for one subject. A noteworthy aspect of this procedure is that the data were not explicitly segregated by condition. Because ICA finds statistically independent components, sources that were consistent between conditions were combined by the algorithm into a single component, while sources that were distinct from one condition to the next were assigned to separate components.

In order to resolve the sign ambiguity prior to clustering, the sources had to be back-projected to the scalp (recall Fig. 1 and Eq. 8). Thus, to calculate time-domain source waveforms representing individual source activities over the entire epoch of -200 to 900 ms, the trial-averaged data for each condition were multiplied by the inverse of the mixing matrix; this was done for each subject separately. The 18 resulting source waveforms for each of the 12 subjects were normalized to zero-mean and unit-variance. The final step in back-projecting the sources to the scalp was to multiply each source for each subject by its associated column of the mixing matrix for that subject.

The k-means clustering algorithm was then used to group components having similar spatio-temporal profiles. Given an integer, k , this algorithm groups the data into exactly k clusters, such that the sum of the intra-cluster variances is minimized. A description of the algorithm is provided in Karhunen et al. (2000) and references therein. Unfortunately, choosing the number of clusters can be quite complex; estimation of the optimal number of clusters is an area of research in itself (Burman, 1989; Xu, 1997). Currently, there is no generally applicable automatic procedure for selecting the optimal number of clusters. Therefore the selection is most effectively made by a human expert, relying on criteria that must necessarily remain somewhat arbitrary. For our purposes, the quality of each cluster was evaluated using the standard error of the cluster-average weight at each electrode, the covariance of the back-projected average sources, and the proportion of subjects involved. It was found empirically that the best balance between these figures of merit was obtained when the number of clusters was set at 10, which lead to clusters involving at least 9 of the 12 subjects.

Back-projected sources were clustered for each condition separately, according to their squared Euclidian distance. The pre-stimulus interval was included in the clustering procedure on the premise that the period of non-stimulus (average non-activity) is as important for grouping as was the post-stimulus period. For each cluster,

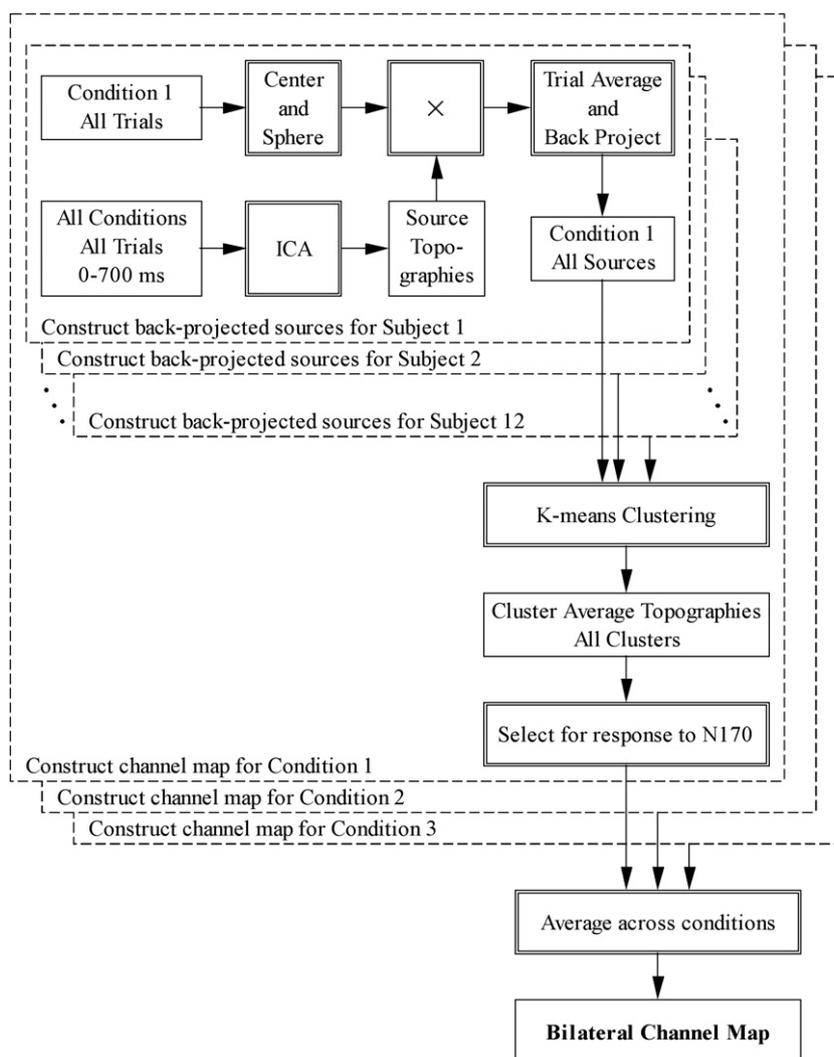


Fig. 2. Procedure for constructing the bilateral channel map.

the average of all the back projections in the cluster was calculated. It was found that, for each experimental condition, exactly one cluster exhibited a significant bilateral, occipito-parietal negative peak in the vicinity of 170 ms post-stimulus. The channel weights, shown in the online-only [Figure S2](#) with standard error indicated, were obtained by averaging within each of these three clusters, over a window of 16 ms length, centered on the peak. The scalp topographies shown in the figure were interpolated from the channel weights using spherical splines. The covariance of the waveforms resulting at each electrode when the back-projected sources are averaged together within each cluster is clearly evident in the figure.

As can be seen from the online-only [Figure S2](#), the clustering of components generated by ICA resulted in very similar relative channel weights for the three experimental conditions. At 170 ms, all three conditions exhibited a pronounced, bilaterally symmetric, occipito-parietal negativity and a corresponding positivity in the anterior central region of the scalp. This is reflected in the emphasis placed on the Fz, Cz, PO7, and PO8 electrodes by the channel weights.

The channel maps constructed for each condition were averaged to arrive at the canonical channel map for the N170, shown in [Fig. 3\(a\)](#). The pronounced bilateral symmetry of this channel map was not by design; it emerged as a result of clustering statistically independent topographic components. The cluster which responded most strongly at 170 ms also happened to be bilaterally symmetric.

In order to decompose the bilateral channel map into left and right hemispheric components, which could be compared directly to the signals obtained from PO7 and PO8, the channel map was modelled as the sum of a pair of dipoles having mirror symmetry across the mid-sagittal plane. DIPFIT, supplied with EEGLab, was used for this purpose. Starting from a random initial placement of the dipole pair, the algorithm optimized the fit by iteratively adjusting the positions and orientations of the dipoles, subject to the symmetry constraint. The fit turned out to be excellent, with a residual variance of only 1%. Nonetheless, it should be emphasized that this was not done for purposes of source localization. Indeed,

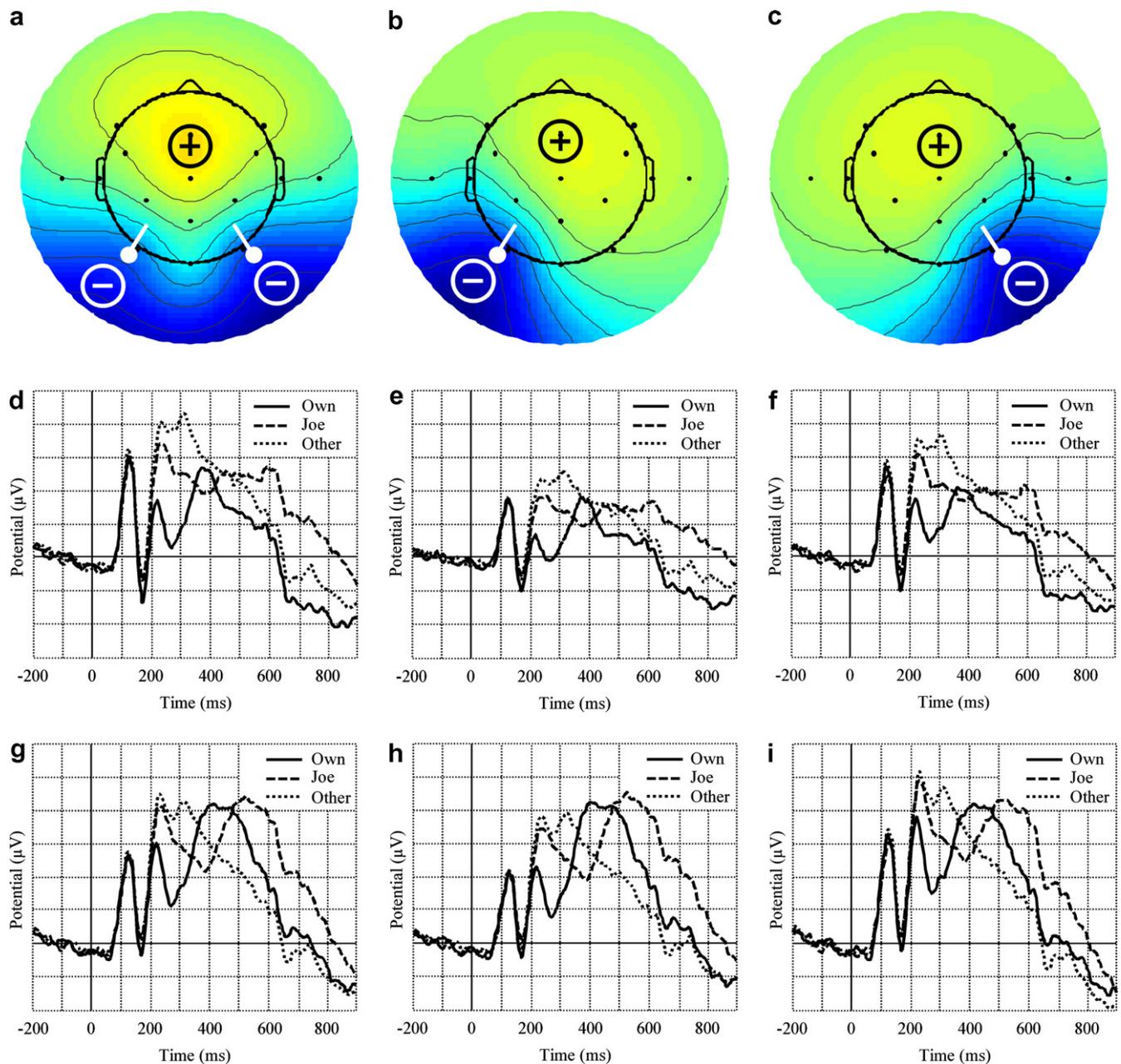


Fig. 3. Decomposition of bilateral channel map into left and right hemispheric maps. (a) Bilateral channel map, (b) left hemisphere channel map, (c) right hemisphere channel map, (d) waveform – channel map, (e) waveform – channel map, (f) waveform – channel map, (g) waveform – PO7 and PO8, (h) waveform – PO7, (i) waveform – PO8.

18 electrodes do not provide sufficient spatial resolution to warrant detailed conclusions regarding the location of sources. However, the resulting left and right hemispheric channel maps, shown in Fig. 3(b) and (c), did provide a suitable basis for comparison with PO7 and PO8 individually, while the bilateral channel map of Fig. 3(a) was comparable to the pool composed of PO7 and PO8 together. The waveforms shown in the second and third rows of Fig. 3 are averages across all subjects of the waveforms obtained by applying the indicated pools to the preprocessed EEG data.

3. Results

The mean values of the peaks at 170 ms are shown in Fig. 4, together with their standard errors. A striking aspect of this figure is that the means are greater than the width of the error bars only for the three channel maps; all other measurements are completely contained within their error bars except for PO7 in the Own and Other conditions. The improvement in signal-to-noise ratio obtained by using the channel maps is thus visually obvious. Additionally, it can be seen that the inclusion of CP3 and CP4

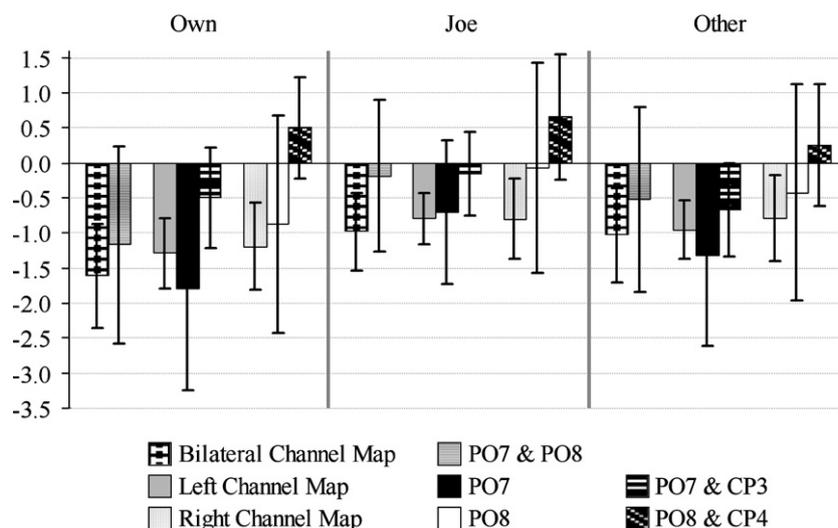


Fig. 4. Mean for each condition, with standard error, using each pooling method.

in pools with PO7 and PO8 not only moves the means in a positive direction, but does so to different extents in the left and right hemispheres, thereby appearing to introduce an effect of hemisphere, independent of condition. These qualitative observations will be born out quantitatively by further examination of the results.

The mean differences between conditions are shown in the online-only Figure S3. It is immediately apparent that all of them, including those obtained using the channel maps, are smaller than their respective error bars. Thus, the effect sizes are seen to be moderate at best, even when using the channel maps. However, the error bars on the measurements obtained with the channel maps are consistently smaller than those obtained using any of the other pooling methods, showing that the channel maps are indeed more effective at reducing variance than the other pooling methods.

That this variance reduction in fact corresponds to an increased effect size is evident from Fig. 5, which shows clearly that the channel maps have sharpened the contrast between Own and Joe, and even more so between Own and Other, but enhanced the similarity between Joe and Other. For Own versus either of the other two conditions, the left and right channel maps increased effect size as compared to the unpooled electrodes PO7 and PO8, whereas pooling CP3 with PO7 and CP4 with PO8 reduced the effect size. In regard to the difference between Joe and Other, the inverse is the case. The same observations apply to the bilateral channel map as compared to the pool of PO7 with PO8.

The online-only Figure S4 shows the results of paired *t*-tests applied to the effects of Fig. 5. As expected, the *p*-value is small wherever the effect size is large, and vice versa. Once again, the use of the channel maps provides stronger evidence supporting the same conclusions that can be deduced from the signals on PO7 and PO8, whether individually or pooled. Namely, that Joe and Other do not

differ significantly, while Own is distinctly different from the other two conditions, and that this pattern is more pronounced in the left hemisphere than the right.

All relevant descriptive statistics, including effect sizes and results of paired *t*-tests, are tabulated in Table 1. Probabilities shown in boldface type are those for which comparisons between conditions yielded statistically significant ($p \leq .05$) differences. In keeping with the interpretations of the preceding figures, the pool of PO7 with PO8 showed a significant own-face effect ($p = .05$), which was even more apparent using the bilateral channel map ($p = .04$). The data for PO7 and PO8 individually were too noisy to permit the effect to be localized to a hemisphere, though the effect was larger in the left hemisphere. In contrast, the increased effect-size resulting from the use of the left and right channel maps allowed the effect to reach statistical significance in the left hemisphere ($p = .05$), but not in the right hemisphere ($p = .10$). The inclusion of CP3 and CP4 in pools with PO7 and PO8 reduces the effect size for Own as compared to either of the other two conditions, and increases the effect size for Joe versus Other, with the result that the effect of condition is completely obscured by this pooling method.

For the bilaterally symmetric pools, a one-way repeated measures ANOVA was performed, and the results appear in Supplemental Table T1. The results bear out the comments regarding Table 1. The effect of condition is found to be significant ($p = .04$) when using the bilateral channel map, but not when using the pool of PO7 and PO8.

For the pools having midsagittal mirror symmetry, a two-way repeated measures ANOVA was performed, and the results appear in Supplemental Table T2. Once again, a significant effect of condition is apparent using PO7 and PO8 unpooled ($p = .04$), and the same effect is more pronounced using the left and right channel maps ($p = .01$). An effect of hemisphere, which appears significant using PO7 and PO8 ($p = .01$), is completely

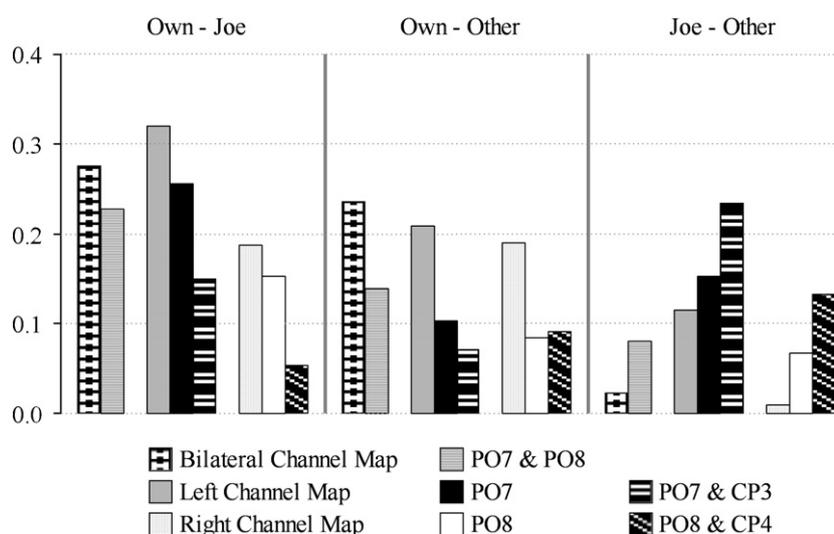


Fig. 5. Effect size (Cohen's *d*) of the differences measured by each pooling method.

invisible when using the left and right channel maps ($p = .44$) but is greatly emphasised ($p = .001$) when pooling PO7 with CP3 and PO8 with CP4. Referring back to Fig. 4, it can be seen that CP3 and CP4 introduce a very large hemispheric difference, which is essentially independent of condition. This difference reflects hemispheric specialization effects which are interesting in their own right, but which have nothing to do with distinguishing between conditions in regard to processing that is occurring occipito-parietally at 170 ms. This difference is so large that it is tempting to conclude that it gives rise, through volume conduction, to the hemispheric difference seen on PO7 and PO8. The complete disappearance of this effect when using the channel maps tends to support this conclusion.

Finally, the Supplemental Table T3 compares the various pooling methods by listing numerical ratios of effect sizes obtained using them. Compared to the pool of PO7

and PO8, the bilateral channel map increased effect size by factors of 1.2 and 1.7 for Own versus Joe and Other, and decreased effect size by a factor of .29 for Joe versus Other. Compared to PO7, the left channel map increased effect size by factors of 1.3 and 2.0 for Own versus Joe and Other, and decreased effect size by a factor of .75 for Joe versus Other. Compared to PO8, the right channel map increased effect size by factors of 1.2 and 2.2 for Own versus Joe and Other, and decreased effect size by a factor of .13 for Joe versus Other. In each case, use of channel maps moved the effect size in the direction of reduced uncertainty. Where conventional pooling schemes showed a difference, the channel maps showed a greater difference; where conventional pooling schemes showed a similarity, the channel maps showed a greater similarity. By contrast, the pools involving CP3 and CP4 always moved the effect size in the direction of greater uncertainty.

Table 1
Descriptive statistics and *t*-test results for all pooling methods

Pooling method	<i>p</i> -value (<i>t</i> -test)			Cohen's <i>d</i>			Mean, variance		
	Own vs. Joe	Own vs. other	Joe vs. other	Own vs. Joe	Own vs. other	Joe vs. other	Own	Joe	Other
Bilateral channel map	0.04	0.07	0.78	0.28	0.24	0.02	-1.6 6.6	-1.0 3.7	-1.0 5.4
PO7 + PO8	0.05	0.17	0.47	0.23	0.14	0.08	-1.2 24	-0.2 14	-0.5 21
Left channel map	0.05	0.10	0.41	0.32	0.21	0.11	-1.3 3.1	-0.8 1.6	-1.0 2.1
PO7	0.06	0.25	0.28	0.26	0.10	0.15	-1.8 25	-0.7 13	-1.3 20
PO7 + CP3	0.34	0.63	0.25	0.15	0.07	0.23	-0.5 6	-0.2 4	-0.7 5
Right channel map	0.10	0.14	0.88	0.19	0.19	0.01	-1.2 4.6	-0.8 3.9	-0.8 4.5
PO8	0.10	0.38	0.44	0.15	0.08	0.07	-0.9 29	-0.1 27	-0.4 29
PO8 + CP4	0.70	0.60	0.20	0.05	0.09	0.13	0.5 6.3	0.7 9.7	0.3 9.0

Probabilities in boldface are statistically significant ($p \leq 0.5$); not corrected for multiple tests.

4. Discussion

In this paper we have shown how to use ICA, clustering and averaging to construct channel maps which can be applied consistently across subjects and which enhance the effects of experimental condition in specific brain regions while filtering out other effects. In contrast to conventional pooling schemes, these channel maps increase effect size without introducing extraneous effects, and thereby enhance the discriminatory power of standard statistical analysis techniques.

We demonstrated the use of such channel maps by using them to investigate the effects of experimental condition on the posterior N170 ERP component. When subjected to the same statistical analysis, the data obtained using the channel maps enhanced both the contrasts and the similarities between experimental conditions that were exhibited by the noisier data obtained using the best conventional pooling schemes, including unpooled electrodes where appropriate. This led to reduced uncertainty in asserting the presence of differences, where they existed, as well as in asserting their absence, where they did not. In contrast to a previous study employing conventional pooling procedures (Tanaka et al., 2006), the current channel map approach revealed differences in the ERP response to the Own condition in comparison to the Joe and Other conditions.

Given the montage used in this experiment, the only possible conventional pooling choice for studying bilaterally symmetric effects is to pool PO7 with PO8. However, use of this pool inevitably results in an effect size which is intermediate between the sizes of the effects observed on the unpooled electrodes, and therefore results in a loss of information. On the other hand, the bilateral channel map yields a net information gain in that it sharpens the contrast between Own vs. Joe and Other, and enhances the similarity between Joe and Other.

Again, given the montage, CP3 and CP4 are the best choices for constructing conventional pools with PO7 and PO8 that have midsagittal mirror symmetry. These pools turn out to be ill-advised; not only do they reduce the signal-to-noise ratio, but – worse – they introduce a strong hemispheric effect originating outside the brain regions of interest. There is thus no way, using conventional pooling, to improve upon the unpooled signals from PO7 and PO8. The left and right channel maps, however, being derived from the bilateral channel map by dipole decomposition, are highly selective for signals with occipito-parietal origins and yield a net information gain over the unpooled electrodes.

The consistent applicability across subjects of our channel maps stands in sharp contrast to other approaches involving ICA and clustering, such as those provided with EEGLab, which result in subject-specific scalp topographies. In cases where compelling evidence exists for significant variations in functional anatomy from one subject to the next, such subject-specific approaches may be war-

ranted, but then the applicability of standard statistical analysis is called into question; the meaning and validity of statistical comparisons between data that are not pre-processed in a uniform manner must be carefully established on a case-by-case basis. Use of channel maps constructed as described in this paper, however, is completely consistent across subjects and therefore such concerns do not arise.

The method described in this paper thus establishes a link between novel ICA-based methods and standard statistical techniques which are well established in the field of ERP research. We have shown that ICA-derived channel maps increase the power of statistical analysis by increasing signal-to-noise ratio as measured by Cohen's *d* statistic. Of course, we have only applied these channel maps in a single paradigm as a demonstration of the method, though there is nothing in the method which is specific to the paradigm. Nonetheless, it remains for future research to apply the method in a wide variety of paradigms to verify its general applicability, and to determine the degree to which further increases in signal-to-noise ratio can be obtained by increasing the density of electrodes on the scalp.

Acknowledgements

We thank the PSYC 576C ERP class at the University of Victoria for making their data available. We also thank Sharon Livingstone, who inspired the idea of reducing variance to improve contrast between the experimental conditions in this paradigm. Funding for this research was provided by the Natural Sciences and Engineering Research Council of Canada.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.clinph.2007.09.001](https://doi.org/10.1016/j.clinph.2007.09.001).

References

- Achim A, Bouchard S. Toward a dynamic topographic components model. *Electroencephalogr Clin Neurophysiol* 1997;103(3):381–5.
- Achim A, Marcantoni W. Principal component analysis of event-related potentials: misallocation of variance revisited. *Psychophysiology* 1997;34:597–606.
- Bell A, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. *Neural Comp* 1995;7:1129–59.
- Burman P. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika* 1989;78(3):503–14.
- Cardoso JF. Entropic contrasts for source separation. In: Haykin S, editor. *Adaptive unsupervised learning*. New York: Wiley; 1999.
- Cohen J. A power primer. *Psychol Bull* 1992;112:155–9.
- Comon P. Independent component analysis – a new concept?. *Signal Process* 1994;36:287–314.
- Contreras-Vidal JL, Kerick SE. Independent component analysis of dynamic brain responses during visuomotor adaptation. *NeuroImage* 2004;21(3):936–45.

- Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 2004;134:9–21.
- Delorme A, Serby H, Makeig S. EEGLab tutorial: VI. Independent component clustering. Retrieved October 11, 2006 from <http://www.sccn.ucsd.edu/eeglab/clusttut/clustertut.html>.
- Dien J. Addressing misallocation of variance in principal components analysis of event-related potentials. *Brain Topogr* 1998;11(1):43–55.
- Dien J, Santuzzi AM. Application of repeated measures ANOVA to high-density ERP datasets: a review and tutorial. In: Handy TC, editor. *Event-related potentials: a methods handbook*. Cambridge: MIT Press; 2005. p. 57–82.
- Field AS, Graupe D. Topographic components analysis of evoked potentials: parameter estimation and some preliminary results. *IEEE Int Symp Circuits Syst* 1990;3:2049–52.
- Field AS, Graupe D. Topographic component (parallel factor) analysis of multichannel evoked potentials: practical issues in trilinear spatiotemporal decomposition. *Brain Topogr* 1991a;3(4):407–23.
- Field AS, Graupe D. Trilinear analysis applied to multichannel evoked potentials. *IEEE Int Symp Circuits Syst* 1991b;1:726–9.
- Gratton G, Coles MGH, Donchin E. A new method for off-line removal of ocular artifacts. *Electroencephalogr Clin Neurophysiol* 1983;55:468–84.
- Gauthier I, Curran T, Curby KM, Collins D. Perceptual interference supports a non-modular account of face processing. *Nat Neurosci* 2003;6:428–32.
- Hyvärinen A, Karhunen J, Oja E. *Independent component analysis*. New York: John Wiley and Sons; 2001.
- Hyvärinen A, Oja E. A fast fixed-point algorithm for independent component analysis. *Neural Comput* 1997;9(7):1483–92.
- Jasper H. The ten–twenty system of the international federation. *Electroencephalogr Clin Neurophysiol* 1958;10:271–5.
- Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. Analysis and visualization of single-trial event-related potentials. *Human Brain Map* 2001;14(3):166–85.
- Karhunen J, Mäläröiu S, Ilmoniemi M. Local linear independent component analysis based on clustering. *Int J Neural Syst* 2000;10(6):439–50.
- Kayser J, Tenke CE. Optimizing PCA methodology for ERP component identification and measurement: theoretical rationale and empirical evaluation. *Clin Neurophysiol* 2003;114:2307–25.
- Lal TN, Schröder M, Hinterberger T, Weston J, Bogdan M, Birbaumer N, et al. Support vector channel selection in BCI. *IEEE Trans Biomed Eng* 2004;51(6):1003–10.
- Lee TW, Girolami M, Bell AJ, Sejnowski TJ. A unifying information-theoretic framework for independent component analysis. *Comput Math Appl* 2000;39:1–21.
- Makeig S, Bell A, Jung T, Sejnowski T. Independent component analysis of electroencephalographic data. In: Touretzky D, Mozer M, Hasselmo M, editors. *Advances in neural information processing systems* 1996;vol. 8. Cambridge, MA: MIT Press; 1996. p. 145–51.
- Makeig S, Delorme A, Westerfield M, Jung TP, Townsend J, Courchesne E, et al. Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biol* 2004;2(6):0747–62.
- Makeig S, Jung T-P, Ghahremani D, Sejnowski TJ. Independent component analysis of simulated ERP data. In: Nakada T, editor. *Integrated human brain science: theory, method, applications (music)*. Amsterdam: Elsevier; 2000.
- McFarland DJ, McCane LM, David SV, Wolpaw JR. Spatial filter selection for EEG-based communication. *Electroencephalogr Clin Neurophysiol* 1997;103(3):386–94.
- Möcks J. Decomposing event-related potentials: a new topographic components model. *Biol Psychol* 1998a;26:199–215.
- Möcks J. Topographic components model for event-related potentials and some biophysical considerations. *IEEE Trans Biomed Eng* 1998b;35(6):482–4.
- Montgomery Jr EB, Huang H, Assadic A. Unsupervised clustering algorithm for N-dimensional data. *J Neurosci Methods* 2005;144(1):19–24.
- Onton J, Delorme A, Makeig S. Frontal midline EEG dynamics during working memory. *NeuroImage* 2005;17(2):341–56.
- Papoulis A. *Probability, random variables, and stochastic processes*. New York: McGraw-Hill; 1991.
- Parra L, Sajda P. Converging evidence of linear independent components in eeg. *Proc 1st Int IEEE EMBS Conf Neural Eng* 2003:525–8.
- Pham DT, Garrat P, Jutten C. Separation of a mixture of independent sources through a maximum-likelihood approach. *Proc European Sig Process Conf* 1992:771–4.
- Rossion B, Gauthier I, Tarr MJ, Despland P, Bruyer R, Linotte S, et al. The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Cogn Neurosci* 2000;11:69–74.
- Sorensen PL, Zeman PM, Sutherland RJ. Differing patterns of synchronous cortical activity during a virtual spatial navigation task. *Canadian Society for Brain, Behaviour and Cognitive Science, 16th Annual Meeting, Paper presentation*; 2006.
- Tabachnick BG, Fidell LS. *Using multivariate statistics*. 2nd ed. New York: HarperCollins Publishers; 1989.
- Tanaka J, Porterfield A. The search for self-identity: The own-face effect (Abstract). *J Vis* 2001;1(3):334a.
- Tanaka JW, Curran T, Porterfield A, Collins D. The activation of pre-existing and acquired face representations: The N250 ERP as an index of face familiarity. *J Cogn Neurosci* 2006;18:1488–97.
- Wang K, Begleiter H, Porjesz B. Trilinear modeling of event-related potentials. *Brain Topogr* 2000;12(4):263–71.
- Xu L. Bayesian Ying-Yang machine, clustering and number of clusters. *Pattern Recogn Lett* 1997;18:1167–78.