Independent component analysis at the neural cocktail party

Glen D. Brown, Satoshi Yamada and Terrence J. Sejnowski

'Independent component analysis' is a technique of data transformation that finds independent sources of activity in recorded mixtures of sources. It can be used to recover fluctuations of membrane potential from individual neurons in multiple-detector optical recordings. There are some examples in which more than 100 neurons can be separated simultaneously. Independent component analysis automatically separates overlapping action potentials, recovers action potentials of different sizes from the same neuron, removes artifacts and finds the position of each neuron on the detector array. One limitation is that the number of sources – neurons and artifacts – must be equal to or less than the number of simultaneous recordings. Independent component analysis also has many other applications in neuroscience including, removal of artifacts from EEG data, identification of spatially independent brain regions in fMRI recordings and determination of population codes in multi-unit recordings.

A guest at a cocktail party must focus on one person's voice in a room filled with competing voices and other noises. This 'cocktail-party problem' is solved effortlessly by humans with binaural hearing[1]. Recently, independent component analysis (ICA) algorithms have provided an automated solution to the cocktail-party problem under certain idealized conditions[2–5]. The problem of isolating the electrical activity of single neurons in population recordings shares a number of similarities with the challenge of isolating voices at a cocktail party so we call it the 'neural cocktail-party problem'.

Wire-electrode and photodiode arrays record hundreds of individual neurons taking part in neural cocktail parties[6–10]. Action potentials from individual neurons can be related to behavior, either by recording from behaving animals implanted with electrode arrays or by optically recording fictive behavior patterns in reduced preparations. Unfortunately, extracting spike trains from these large data sets has remained labor intensive because automated methods for data analysis have not kept pace with advances in recording technology.

ICA (Fig. 1) promises to improve our ability to extract neural signals from recorded mixtures. We focus here on how ICA can be used to recover spike trains automatically from optical recordings made in the isolated brain of the seaslug, *Tritonia diomedea,* using a voltage-sensitive dye[11] (Fig. 2). Because the success or failure of separating spike trains is easy to judge, spike-train recovery provides a good real-world application for demonstrating the power and highlighting the limitations of ICA.

ICA finds spike trains by unmixing the optical signals into their most independent parts. Similar to its widely used predecessor, principal components analysis (PCA), ICA removes redundancy from multivariate data. PCA transforms data so that the co-variance between pairs of variables vanishes, whereas ICA attempts true redundancy reduction by minimizing mutual information (Box 1).

Because ICA traces its roots to blind source separation in signal processing[12], some signal-processing jargon must be introduced. The basic problem is to recover n sources given n different linear mixtures of sources (Fig. 1). The mixtures could be, for example, sound recordings from microphones at a cocktail party or in the case of optical recordings, output from photodiode detectors. The interesting sources are cocktail party guests or neurons, respectively, but noise sources must also be considered, as explained below. 'Sources' are also called 'latent variables,' 'hidden variables,' or 'causes' in other contexts. We use 'detectors' and 'mixtures' interchangeably, and these are also called 'measured variables' or 'manifest variables' in other contexts.

All sorting and filtering algorithms for recovering spike trains from multiple-detector recordings face similar challenges[11,13–16]. First, recording artifacts must be removed or ignored. Second, spikes corresponding to action potentials must be detected. Third, action potentials from neurons must be separated from each other in spite of similarities in their shape. Fourth, coincident action potentials that appear on the same detector(s) must be separated and assigned to appropriate neurons. Finally, action potentials from the same neuron should be put into a single group even when they change shape, for example during high-frequency bursts of spikes. The latter problem has defeated many conventional spike-sorting algorithms that assume a fixed spike shape[17,18]. Ideally, even subthreshold membrane potential fluctuations would be detected and assigned to the appropriate neuron.

Spike-sorting algorithms group spikes according to their shape because action potentials with similar shapes will often correspond to action potential trains from a single neuron[13–18]. However, there is also information about the identity of a spike from its spatial distribution on the detector array. If spikes are sorted on each detector separately, this potential source of information will be lost. Modern algorithms sort spikes based on events from multiple channels simultaneously, which makes some use of the spatial pattern of activity[14,19,20].

ICA uses only the spatial information[2–4]. Many ICA algorithms actually shuffle the time points before the

Glen D. Brown*
Terrence J. Sejnowski Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA. *e-mail: glen@salk.edu

Satoshi Yamada Advanced Technology R&D Center, Mitsubishi Electric Corporation, Amagasaki, Japan.

Terrence J. Sejnowski Dept of Biology, University of California San Diego, La Jolla, CA 92093, USA.
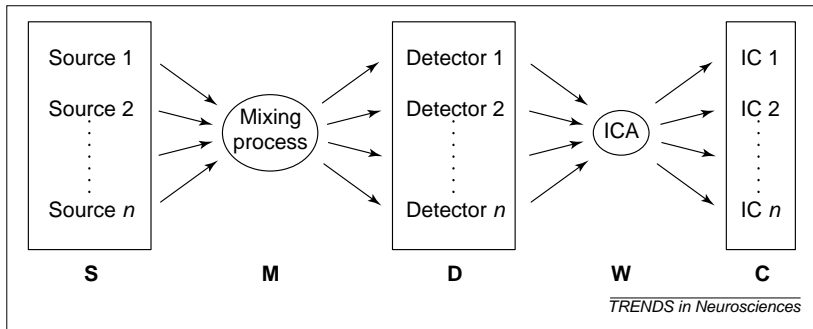
TRENDS in Neurosciences

Fig. 1. The independent component analysis model. *n* hypothetical source signals (S) are mixed linearly and instantaneously by an unknown mixing process (M). Mixed sources are recorded by a set of *n* detectors (D). Independent component analysis (ICA) (W) transforms the detected signals into *n* independent components (IC) (C). If the assumptions of the model are not violated, the independent components will be the original sources except that scale, sign, and order will not be preserved. The number of sources is assumed to be equal to or less than the number of detectors, which leads to the completeness problem described in the text.

analysis. The spatial pattern of activity on the detector array is assumed to be constant for each neuron. ICA recognizes this pattern of activity and collects redundant signals into a single channel. Each unmixed channel is an independent component (Fig. 1).

Of course neurons are not independent of one another because they can be coupled by synapses or they can share inputs. Nevertheless, the individual spike trains might be the statistically independent components of population activity (Fig. 2). Note that dependence in this context refers only to instantaneous overlaps, not time-delayed dependencies. Sources that are not active simultaneously will not need to be separated anyway. Furthermore, ICA does not find truly independent components because independence does not exist in the real world, and the independent components are really only as independent as is
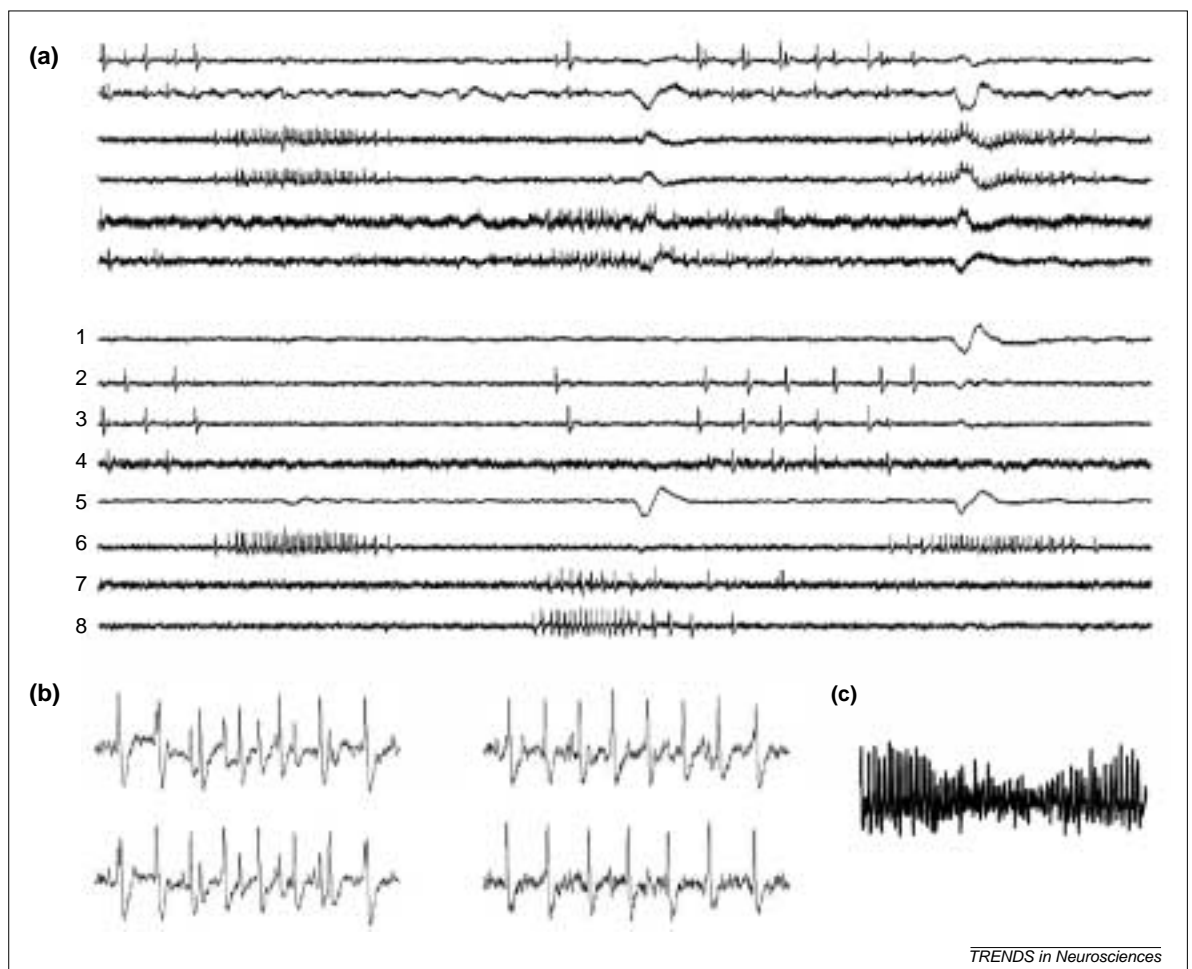

TRENDS in Neurosciences

Fig. 2. Independent components of optical recordings. (a) Upper traces show six channels of raw data selected from a group of 448 detectors recording the membrane potential of *Tritonia* neurons stained with a voltage-sensitive dye[6]. The *y* axis is proportional to transmitted light, which is, in turn, proportional to fluctuations in membrane potential[8]. Bursts of action potentials can be seen in the raw data, but each detector records from several different neurons, and artifacts are also present on many channels. Data were collected at 1 KHz. The independent component analysis (ICA) algorithm returned 448 independent components eight of which are shown in the lower panels (1–8). The *x* axis (time) is not changed by independent component analysis (ICA). Following ICA the *y* axis is arbitrary because scale is not preserved. Many independent components appeared to be spike trains from individual neurons (2,3,6–8). Other components were recording artifacts (1,5). (b) Coincident events were separated automatically. The left column shows raw data from two neighboring photodiode detectors (1 s total time). There appears to be action potential activity from two different neurons on both traces. The traces on the right show two independent components (1 s total time) that appear to be the unmixed action potential trains. The first two action potentials of each neuron in addition to the final one coincided. Coincident action potentials were common in these recordings because the neurons share presynaptic inputs. (c) ICA-separated nonstationary action potentials. Spikes changed shape during a high-frequency burst (shown after ICA, 1 s total time) and these types of changes actually aid the ICA separation (see text).

Box 1. Minimizing mutual information

Consider the following three sequences of ones and zeros, labelled A, B and C:

A   1110111010010000001011110000001011001...
B   0010011011000101101100100111111111010...
C   1100100001010101100111010111101000011...

Each sequence might be considered to be a measured variable or detector and each column a data point or time point. No two sequences co-vary. There is a probability of ~0.5 that B will be zero when A is zero, and that C will be one when A is zero, etc. However, zero co-variance does not imply independence. (C is the exclusive – or of A and B.) The columns contain only four of the possible eight combinations of ones and zeros. This is a third-order redundancy because it depends on consideration of all three variables.

Principal component analysis (PCA) could not reduce the dimensionality of this three-dimensional system because PCA only removes co-variance from

multivariate data. By contrast, ICA minimizes the mutual information among variables, which includes third-order redundancies of this type[a,b]. This system could be represented by two new sequences, for example X and Y:

X   0001000101101111101000011111010100110...
Y   0011011110101010011000101000010111100...

where (0,0) has replaced (1,0,1), etc.

The original sequences contained a third-order but no second-order relationship. In practice, mutual information normally includes co-variance in addition to third and higher order dependencies. ICA minimizes redundancies of all orders.

References
a   Comon, P. (1994) Independent component analysis, a new concept? *Sig. Proc.* 36, 287–314
b   Bell, A.J. and Sejnowski, T.J. (1995) An information-maximisation approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1004–1034

---

possible. If the spike trains are the main, or high-entropy, parts of the recorded signals, then they must be separated in order to make the resulting components as independent as possible.

In general, dependencies between sources will not be a major concern in deciding whether or not to use ICA on a particular data set. Interesting sources tend to be sufficiently independent. Instead, other assumptions of the ICA model – linear, instantaneous mixing of spatially stationary sources in which the number of sources is less than or equal to the number of detectors (Fig. 1 and discussed below) – must be considered.

Algorithms that solve the neural cocktail-party problem will be important for analyzing population recordings from closely spaced neurons, and these types of recordings will be of interest for several reasons. The distance between neurons connected by synapses has been minimized by natural selection to keep axons and dendrites as short as possible[21–23]. Neurons that send or receive the same information will tend to sit close to one another in the nervous system. If neurons share inputs, it is more probable that they will be neighbors in order to minimize axon length in presynaptic neurons. Similarly, axons carrying information with coincident action potentials might be adjacent in order to minimize dendrite length of shared postsynaptic targets.

To summarize, analysis of recordings from closely spaced neurons will be necessary to analyze neurons connected by synapses, to measure redundancy and to understand population codes in the nervous system. For example, it is of interest to record from as many neurons as possible in a cortical column, such as an orientation column in the primary visual cortex (V1) during a visual task. ICA might be able to help sort out recordings of such a neural cocktail party.

ICA in action

Optical recordings were made with a photodiode-detector array containing 448 elements in the isolated brain of the nudibranch seaslug *Tritonia diomedea* after staining with a voltage-sensitive absorbance dye[6]. During each recording, a fictive swimming episode was activated by electrical stimulation of a nerve root. Fictive swimming lasted ~30 s and was characterized by bursts of action potentials in many of the neurons that form the swimming network of *Tritonia*[24,25]. The raw data contained mixtures of action potentials and recording artifacts. The activity of most neurons was recorded by more than one detector, and one detector often recorded action potentials from multiple neurons.

Recovering spike trains from these large data sets has previously required up to a week of operator time, mostly for correcting errors made by the spike-sorting algorithms[26]. By contrast, ICA processes data automatically and requires only a few hours of operator time to proceed from raw data to separated spike trains[11] (Box 2).

ICA returned a continuous estimate of membrane potential for each neuron, and artifacts were assigned to separate channels (Fig. 2a). Overlaps were automatically resolved (Fig. 2b). Action potentials of different shapes and sizes from the same neuron were put into a single, independent component (Fig. 2c). Although subthreshold changes, such as postsynaptic potentials or intrinsic membrane fluctuations, were detected rarely in these recordings, ICA can recover subthreshold potentials in addition to spikes. This complements advances in voltage-sensitive-dye technology that make the detection of subthreshold potentials possible[27].

Changes in spike shape actually help the ICA algorithm. To understand why, consider the

Box 2. Using independent component analysis

Matrix multiplication

Linear transformations, such as independent component analysis (ICA), are matrix multiplications. Two matrices can be multiplied if their inner dimensions agree, that is if the number of columns in the first equals the number of rows in the second. Suppose matrix W has dimensions m by n (m rows by n columns) and matrix D has dimensions n by p. The resulting matrix (C) has outer dimensions m by p:

$$\begin{matrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & & & \\ & W & & \\ w_{m1} & \cdots & \cdots & w_{mn} \end{matrix} \quad \begin{matrix} d_{11} & d_{12} & \cdots & d_{1p} \\ d_{21} & & & \\ & D & & \\ d_{n1} & \cdots & \cdots & d_{np} \end{matrix} \quad = \quad \begin{matrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & & & \\ & C & & \\ c_{m1} & \cdots & \cdots & c_{mp} \end{matrix}$$

The individual entries in C are calculated as follows:

$$C_{ij} = \Sigma_{k=1\ to\ n}\ (w_{ik})(d_{kj})$$

where i takes values from 1 to m and j takes values from 1 to p. That is, each entry in C is calculated by summing the individual products when the ith row of W is multiplied by the jth column of D.

A few other matrix algebra rules must also be remembered. The identity matrix (I) has ones on the diagonal from top left to bottom right and zeros everywhere else. Multiplying by I does not change a matrix, e.g., $ID = D$. The transpose of a matrix, signified with a superscript T exchanges the rows with the columns of a matrix. By definition, multiplication with an inverse reverses multiplication so if $WD = C$, then $W^{-1}C = D$. Methods for inverting matrices can be found in a linear algebra text and computers running mathematical software can invert even large matrices.

Finding independent components

Consider the case of multiple detectors sampling the environment over time. If the signals from each detector form the rows of the data (or detector) matrix D, then each column of D is a time point. Independent components analysis (ICA) finds the square matrix W ($n = m$ = the number of detectors) such that $WD = C$. The rows of C are called 'independent components' because they are forced to be as independent as possible[a]. The independent components are the same length as the data and there are the same number of independent components as there are of detectors. This can be represented schematically:

$$\boxed{W}\ \boxed{\qquad D \qquad}\ =\ \boxed{\qquad C \qquad}$$

The hope is that the independent components (rows of C) will be of interest. W is called the 'unmixing matrix' because it unmixes the detected signals in D, which are assumed to be mixtures of signals from different sources. Each row of W unmixes the detected signals (D) into one independent component (row of C) so rows of W are sometimes called unmixing functions or ICA filters.

If the assumptions of the ICA model are correct, the rows of C will be the original source signals. However, neither the sign nor the scale will be preserved and the independent components will be shuffled with respect to the original sources. The trick of course is finding (or estimating) W, and this is accomplished by the ICA algorithm (see main text)[b]. One ICA algorithm[b] starts with a W that eliminates correlations, similar to a principle component analysis solution (Box 1), and iteratively refines the entries in W until it converges on the independent components (Box 3).

Viewing the columns of W^{-1}

ICA also reveals where the source signals were located with respect to the detectors. Postulated sources form the rows of a matrix S. If these sources are the same as the independent components, that is $S = C$, then $W^{-1} = M$, where M is the matrix that originally mixed the signals and:

$$\boxed{M}\ \boxed{\qquad S \qquad}\ =\ \boxed{\qquad D \qquad}$$

It might even be helpful to remember the mixing matrix M as something like an upside down W, which is the unmixing matrix. Of course a matrix inverse is not a matrix turned over. W originally stands for 'weight matrix' because the entries in W weight how much each detector will contribute to each independent component. Also, the entries in W became weights in a neural-network-based ICA algorithm[b].

The entries in M are determined by physical properties of the recording set-up, such as the distance between sources and detectors. One row of M mixes the original source signals for one detector, in the same way that one row of W unmixes one independent component. Now the useful part: a column of M, or in practice a column of W^{-1}, reveals which detectors recorded a particular independent component. The columns of M are sometimes called 'basis functions'.

Although the detectors are stacked on top of each other in one dimension in D, they might also have a two or three-dimensional structure. A column from W^{-1} can be visualized with respect to the original arrangement of detectors to reveal the pattern of detectors that recorded the corresponding independent component; this is a 'place map' (see main text).

Reconstructing raw data without artifacts

W^{-1} can also be used to reconstruct raw data without artifactual sources of signal. Each column of W^{-1} corresponds to one independent component, some of which are the source signal of interest, but many of which can be recording artifacts. If the columns of W^{-1} corresponding to artifacts are set to zero (the zero vector) resulting in the matrix W^{-1*}, and remembering that $W^{-1}C = D$, W^{-1*} is multiplied by C to reconstruct D without artifacts (now D^*):

$$\boxed{W^{-1*}}\ \boxed{\qquad C \qquad}\ =\ \boxed{\qquad D^* \qquad}$$

Ordering independent components

No standardized ordering method exists in ICA. Independent components could be ranked, similar to principal components, by the amount of variance each explains in the original data. Entropy can be used instead of variance (G.D. Brown, unpublished). However, a mathematical ordering for independent components is not necessarily important in applications. Similar to principal components, interesting independent components would not necessarily follow a generalized ordering scheme anyway because interest is application-specific. For example, in fMRI recordings, the ICA components with the largest variance tend to be artifacts[c].

References

a Comon, P. (1994) Independent component analysis, a new concept? *Sig. Proc.* 36, 287–314
b Bell, A.J. and Sejnowski, T.J. (1995) An information-maximisation approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1004–1034
c McKeown, M.J. *et al.* (1998) Spatially independent activity patterns in functional magnetic resonance imaging data during the Stroop color-naming task. *Proc. Natl. Acad. Sci. U. S. A.* 95, 803–810
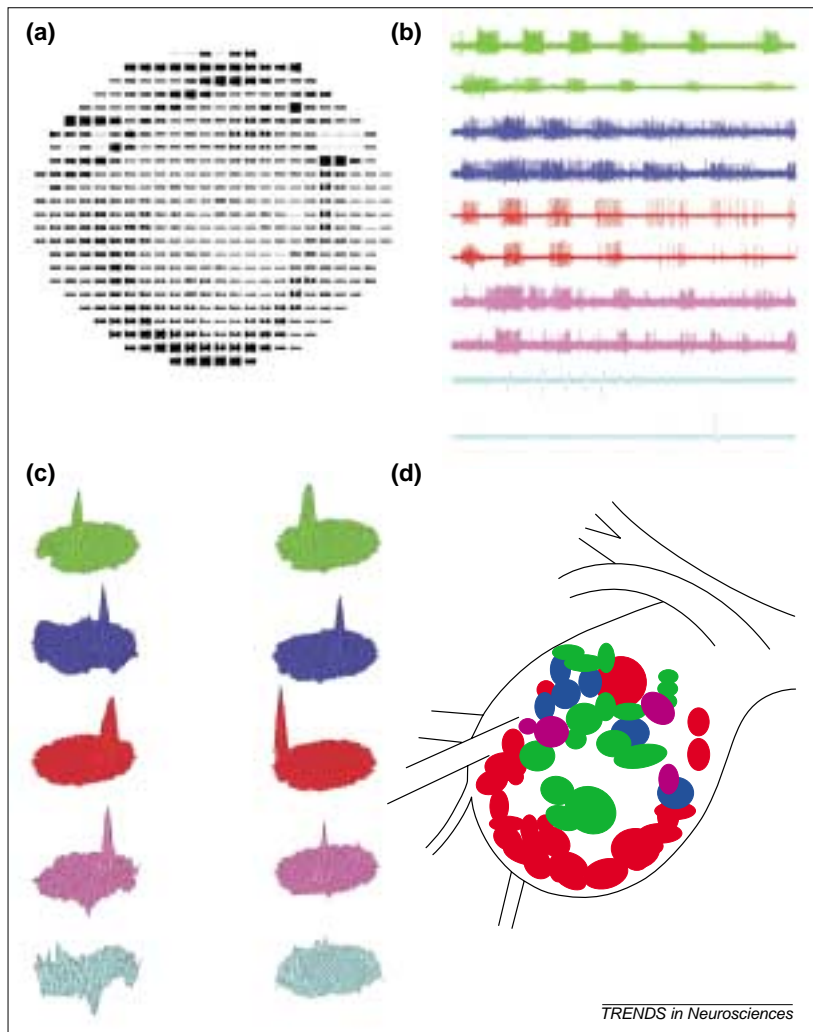
Fig. 3. Independent component analysis recovers the position of each neuron. (a) Eight seconds of raw optical recording data are shown at low resolution with the two-dimensional structure of the photodiode array preserved. Each pixel shows the activity recorded by one detector over 8 s. Bursting neurons can be seen. (b) Independent components were classified as one of four types of neurons[6,24,25] (top eight traces) or as artifacts (bottom two traces) and color coded (45 s total time). (c) For each of the four types of neurons and the artifacts represented in (b) the detector array is represented in the *x–y* plane of each place map. The *z* axis is the relative contribution of each detector to one independent component. As expected, independent components corresponding to single neurons formed tight clusters on the detector array. By contrast, recording artifacts (light blue) were distributed more widely. (d) The individual place maps for each type of neuron shown in (b) and (c) were thresholded and plotted in two dimensions as a colored spot. The recorded area of the *Tritonia* pedal ganglion shows the positions of the neurons in the brain. All data were collected at 1 KHz.

unrealistic example of two neurons that always fire synchronous action potentials and where the action potentials never change shape. ICA would consider these as one source, even if the spike shapes from the two neurons were different. However, assuming that the fluctuations themselves are independent, fluctuations in the shape of the spikes makes it possible to separate the two signals. Recalling the classic cocktail-party problem, it would be more difficult for a human to attend to an individual speaker if everyone were saying the same thing at exactly the same time. (Of course this is improbable, and computerized ICA algorithms could still separate such a chorus as the sound waves from the various speakers would arrive with different phases.)

ICA also recovered the spatial pattern of activity on the detector array for each neuron (Fig. 3, Box 2), which are often called 'place maps'. The brain region underneath the detector array was photographed so that the place maps revealed the location of each neuron. Place maps from neurons with similar spiking patterns can be combined to form topographic maps (Fig. 3).

The ICA algorithm itself puts no constraints on the shapes of the place maps[5]. However, we know *a priori* that detectors that record the activity of a single neuron should be close together. When the patterns associated with spike trains formed tight clusters of detectors on the array (Fig. 3c and 4b), this provided corroborative evidence that we had indeed recovered spike trains from individual neurons. Thus, knowledge of what the place maps should look like can help validate the ICA separation. In practice, the individual spike trains themselves are easily distinguished from artifacts or mixtures of spike trains (Figs. 2 and 3). Similarly, it is easy for a listener to distinguish a recording of one person talking from a recording of fragments of speech taken from several conversations.

Artifacts also have distinctive patterns of activity on the detector array (Fig. 3c), and ICA allows raw data to be reconstructed without artifacts (Fig. 5). This type of artifact rejection (Box 2) could be important, even when ICA is not needed to separate the spike trains from one another. For example, neurons might already be isolated on a single channel when using arrays of fine-wire electrodes[8,9], but artifacts might still appear on multiple channels. If artifacts can be isolated by ICA, then they can be eliminated from the data.

Choosing an ICA algorithm
The basic ICA step is a linear change of variables (Box 2). The difficult part is done by the ICA algorithm, which must find an unmixing matrix to separate the independent components (Box 3). Although ICA can be used effectively without a full understanding of the underlying algorithm, some knowledge of possible pitfalls of ICA is essential and will be discussed further in the next two sections.

Several ICA algorithms have appeared recently, including artificial neural networks[4,5] (Box 3), projection pursuit methods[2], maximum-likelihood approximations[28,29], fixed-point algorithms[30] and wavelet-based methods[31]. Computer code (often in MATLAB), which in most cases can be downloaded from the internet, is available from the creators of the algorithms. Existing algorithms are being improved, and newer, more-powerful algorithms will probably appear[32–34]. At the moment, it is unclear if there is a single best method or even what the application-dependent variables might be, so it is probably a good idea to compare several algorithms on a new data set.

We compared ICA algorithms on optical recording data and also compared ICA to PCA (Fig. 4). PCA did not separate spike trains adequately, indicating that
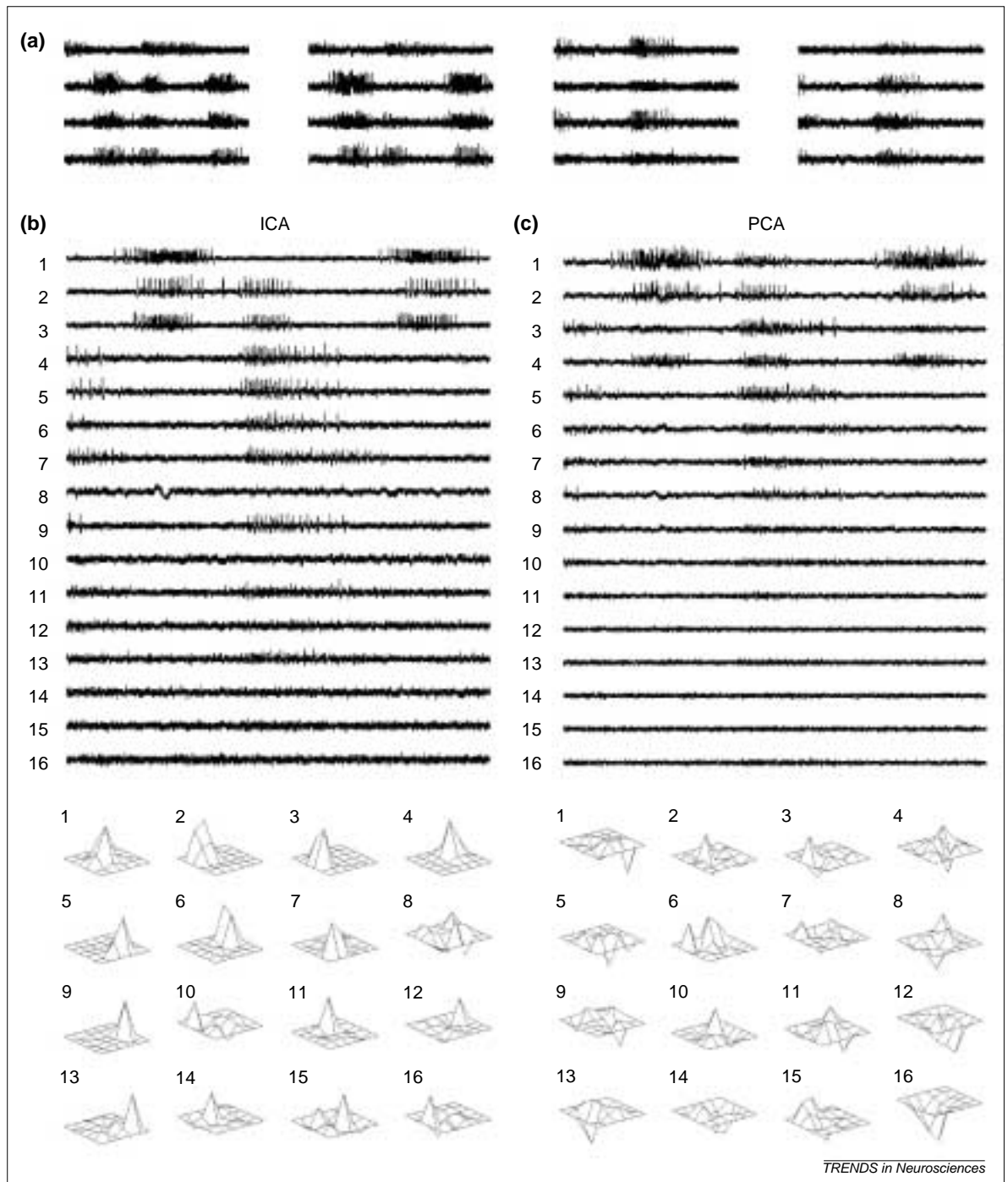
Fig. 4. Principal component analysis does not solve the neural cocktail-party problem. (a) Raw optical recording signals taken from a four by four group of photodiode detectors (8 s total). The arrangement of detectors has been preserved. (b) After independent component analysis (ICA), eight action potential trains from single neurons (traces 1–7 and 9) were apparent (8 s total, *y* axis arbitrary). Components 11 and 13 also showed action potential activity. Another component (8) appeared to be a recording artifact. In actual experiments[6], separation of spike trains was even higher quality because more detectors were used (see Figs 2 and 3). (c) Principal component analysis (PCA) was not able to isolate multiple spike trains from individual neurons. PCA also failed to separate recording artifacts from spike trains. In general, the PCA solution was unacceptable except in the most trivial cases.

statistical relationships between detectors beyond pairwise interactions are important for the separation (Box 1). Consistent with other studies, most ICA algorithms produced similar results[11,34,35]. There was some trade-off between computer time and the quality of the separation with the different algorithms, especially as the number of detectors increased, but more striking was the overall similarity of the results. Using the infomax algorithm (Box 3), the number of operations required to do ICA grows as a fourth power of the number of detectors and linearly with the number of data points. In practice, >100 channels can be routinely separated with a desktop PC.

The completeness problem
The ICA model assumes that there are the same number of detectors (e.g. photodiodes) as there are

Fig. 5. Data reconstructed without artifacts. Raw data can be reconstructed without artifacts (Box 2). The top trace is raw data with high and low frequency artifacts (2 s total; *y* axis proportional to transmitted light). The lower trace is the same signal after reconstruction but without the independent components associated with the artifacts. All data were collected at 1KHz.

TRENDS in Neurosciences
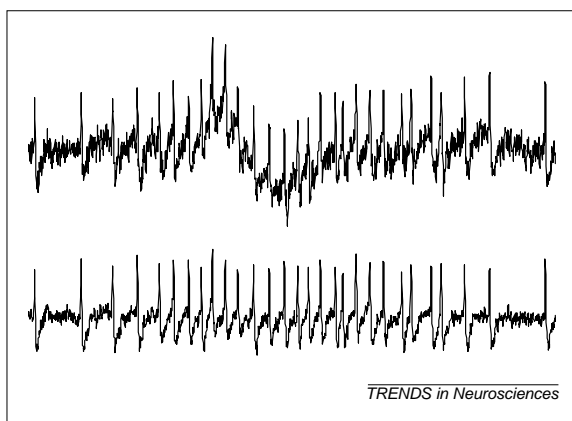
source signals[3] (Fig. 1). Thus, ICA finds a complete representation of the data. Standard ICA algorithms can recover fewer sources than detectors (the undercomplete case, but fail when there are more sources than detectors (the overcomplete case). In addition to the sources of interest, ICA considers unwanted nongaussian signal sources, such as movement artifacts or line noise, as independent components (Figs 2 and 3).

The completeness problem leads to a counter-intuitive result: more data are not always better (Fig. 6). When sources only appear sporadically during a recording, more data can mean more sources and a higher probability that data will be overcomplete. Conversely, ICA can be used for overcomplete data if the sources are not active all the time (Fig. 6b), that is, if subsets of the data are either complete or undercomplete. For example, artifacts can be removed from overcomplete optical recording data (Fig. 5) if not all neurons are active when the artifact appears.

The completeness problem applies locally in addition to applying globally: the number of sources recorded by any subset of detectors, where these sources are not recorded by any other detectors, must be less than or equal to the number of detectors in that subset. ICA will fail on any part of the data that is overcomplete. For example, ICA cannot separate two neurons that are recorded by one detector if their activity is not recorded by other detectors. Thus, all independent components must be examined after unmixing to determine if the analysis has been successful.

Encouragingly, algorithms that find overcomplete representations have been introduced[29,32,34]. In the future, it will be possible to separate more spike trains than there are detectors. Of course no matter how much algorithms improve, there will be some limit on the number of spike trains that can be accurately separated when recording with a limited number of detectors.

Other constraints
Although ICA could be applied to any type of multivariate data, some algorithms require many

data points to do the analysis. In signal processing, researchers often sample many times per second so this might not be a problem. However, an ecologist might sample an ecosystem once per year. Nevertheless, there is reason to be optimistic about using ICA on categorical and other types of data[35–37]. The ability of an ICA algorithm to transform data in an interesting way depends on the number of measured variables, the number of latent variables and the amount of data collected.

The automated cocktail-party problem, that is, using a machine to isolate multiple voices recorded by multiple microphones, can be used to illustrate other assumptions of the ICA model. First, sources must be stationary. Guests cannot move relative to the microphones during recording. Second, there can be no delays, introduced, for example, because a speaker is closer to some microphones than others. Similarly, the method can be defeated by echoes. These assumptions might be reasonable for the neural cocktail party (Figs 2–6) where the recording configuration is stationary and mixing occurs at the speed of light, but not for sound recordings at a real cocktail party. The ICA algorithm can be generalized to data with time delays, but only at the expense of introducing many more parameters that must be estimated[38,39]. The ICA model also assumes that the source of the signals are nongaussian, which should be true for most interesting source signals. Finally, ICA assumes that sources are mixed linearly, which might not be true generally, but is a good first approximation in many applications.

Who is the loudest person at the cocktail party? Another limitation is a failure of the algorithm to preserve sign or scale (Box 2). When recovering trains of action potentials from optical recordings (Figs 2–5), the size and spatial extent of the signals depend on factors unrelated to membrane voltage (e.g., the size of the neurons), therefore ordering of the independent components is not critical (Box 2).

Other successes
The problem of dimensionality is fundamental to any reductionistic science because interesting problems almost always require the analysis of multiple variables. If too few variables are studied, the system cannot be adequately characterized. If too many variables are considered, analysis and interpretation of data become unwieldy. ICA addresses the latter problem by eliminating redundancy such that the independent components often occupy a lower dimensional space than the original data. Even when dimensionality is not reduced, that is when all of the independent components are retained, ICA helps to insure that the same data is not processed multiple times by, for example, detecting and counting action potentials from the same neuron on multiple detectors.

We have had success applying ICA to recordings [provided by E.J. Chichlinsky (Salk Institute for Biological Studies, CA, USA)] from planar electrode
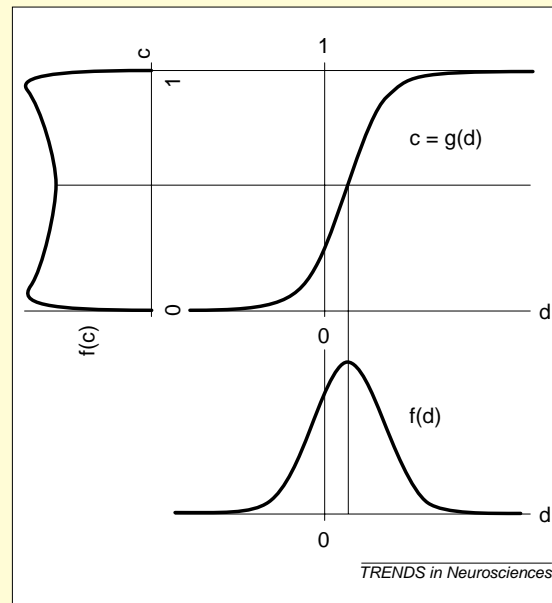

Box 3. Infomax

Fig. I. Maximizing entropy in a sigmoidal neuron. The input, d, with the density function f(d) is passed through a nonlinear function, c=g(d). The maximum amount of information the output can provide about the input is realized when the entropy of the output distribution f(c) is maximized, i.e., when the output distribution is flat.

The basic step in independent component analysis (ICA) is a linear transformation of data, D, into independent components, C, by a weight matrix W (see Box 2):

$$WD = C$$

The goal of any ICA algorithm is to find the weight matrix W.

Infomax falls into the large category of neural-network algorithms that minimize or maximize some cost function[a]. Input neurons represent the data, output neurons are the independent components and the synaptic weights between inputs and outputs form W. The true cost function for ICA is the independence of the outputs, but this is difficult to measure directly. Infomax maximizes the mutual information between the inputs and outputs to approach independence for the outputs. [Remember that mutual information is just a higher-order generalization of correlation (see Box 1) and that entropy is something like variance.]

The principle behind infomax is illustrated for one input (d) and one output (c) in Fig. I. Output neurons have a sigmoidal activation function. The steepest part of the sigmoid is aligned to the mode of the input distribution and the slope is scaled to match the variance. This flattens the output distribution, maximizing its entropy. Because the output is driven only by the input, the maximum entropy state of the output also corresponds to maximum mutual information between the input and the output.

For multiple inputs and outputs, the outputs are spread in a hypervolume instead of the single output dimension shown in Fig. I. Output entropy is still maximized, but there is a term in the cost function that prevents two outputs from reaching the same high-entropy state. Thus, redundancy between outputs is also minimized, which is the goal of ICA.

Before using infomax, weights in W are initialized to the principal components solution, in which outputs are uncorrelated (Box 1). Then, parameters are adjusted or 'learned' in a stepwise manner following a gradient (the derivative of the cost function) leading to the point where the mutual information between inputs and outputs is maximized.

We actually use the 'natural-gradient' version of infomax because it simplifies the learning rules used to adjust the weights:

$$C = WD$$
$$y = 1/(1+e^{-c})$$
$$\Delta W = \varepsilon[I+(1-2y)\,C^{T}]W$$

where ε is a learning rate parameter, and I is the identity matrix[b]. C is denoted in italics here because the rows of C do not become the independent components until after learning is complete. On each step, the learning-rate parameter is reduced and the changes in the weights (entries in W) become smaller. Typically, the algorithm converges after 50–200 steps on the optical recording data analyzed in the main text.

References

a Bell, A.J. and Sejnowski, T.J. (1995) An information-maximisation approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1004–1034
b Amari, S. *et al.* (1996) A new learning algorithm for blind signal separation. *Adv. Neural Inf. Proc. Syst.* 8, 757–763

---

arrays[7]. However, in several other recordings from wire-arrays, including tetrodes and octrodes, individual wires often record from different parts of the same neuron. This adds time delays and presents serious challenges to ICA algorithms. As mentioned above, new algorithms are being developed to address the delay problem[38,39].

Because ICA algorithms use the pattern of activity in space to compute independent components, it would be useful to position detectors to collect information in all three spatial dimensions. For recordings done in an optical plane, a two-dimensional array might be sufficient. However, the spread of electrical signals in three dimensions could be advantageous in processing data from wire-electrodes[40].

Besides sorting out what each individual is saying at the neural cocktail party, ICA can also help determine
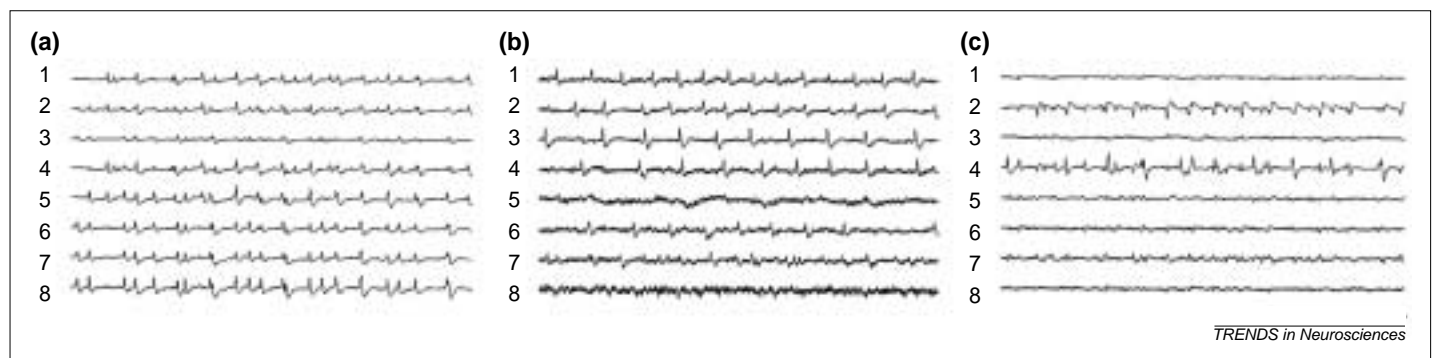
Fig. 6. The overcomplete problem. (a) Two seconds of raw data from eight (1–8) nearby channels were separated into spike trains by independent components analysis (ICA) shown in (b). The ICA showed that traces 1–4 and 6 were clearly spike trains and trace 7 appeared to contain a mixture of spikes from different neurons. (c) Spike trains were not separated when more time points, including the 2 s of data in (a), were used as input to the ICA algorithm. Although ICA was carried out on a much longer recording, only the results from the 2 s corresponding to (a) and (b) are shown. The failure in this case was as a result of the overcomplete problem (the number of neurons and artifacts exceeded the number of raw data channels when more time points were used for the analysis).

which neurons are speaking as a group[36]. Population codes (correlated firing of ensembles of neurons) selected with ICA can predict behavior, such as correct and incorrect trials in a motor-learning task[37].

One of the first applications of ICA to neuroscience was to the EEG, which is simply the neural cocktail-party problem at a different level of organization[41]. Although exactly what constitutes a source in EEG recordings remains unclear, the independent components of EEG might turn out to be clinically useful[42].

Another neuroscience application, ICA of functional magnetic resonance imaging (fMRI) data, has normally used the spatial as oppose to the time domain[35]. Imaging data are collected from many voxels (cubic regions of three-dimensional space) at multiple time points. Time points, rather than detectors, are used as the measured variables and ICA collects signals that are redundant in time into a single channel. The independent components of fMRI data are spatial maps forced to be as independent as possible that often correspond to known brain structures or to a particular type of artifact.

The time course of each independent brain region can be recovered in the same way that the spatial pattern of activity on the detector array was recovered for each neuron in optical recordings (Box 2; Fig. 3). Independent three-dimensional fMRI maps show that widespread brain areas can vary together, which contrasts with the localized pattern of activity of *Tritonia* neurons on the photodiode-detector array (Fig. 4).

Can nervous systems find independent components?
The infomax ICA algorithm (Box 3) reflects principles found to operate in the insect eye and incorporates biologically realistic principles of neurons[4,43–45]. However, the algorithm also depends on a global calculation that is not realistic biologically. Nevertheless, independent

components can be calculated in several ways and it seems natural to wonder if nervous systems ever do ICA.

The brain as a whole is able to solve the cocktail-party problem, especially using binaural input, but the neural mechanisms for doing so are unclear[1,46]. Interestingly, there are more neurons in the auditory cortex that monitor the location of sound sources than there are for spatial localization *per se*[46]. One possibility is that this information is used to separate sources of activity so that individuals can attend to one speaker or process one source of sound at a time.

Barlow has emphasized the importance of redundancy reduction in all sensory modalities[48]. Redundant signals carry common information, whereas sensory systems typically seek novel, high-information inputs. Receptor adaptation, habituation and predictive filtering are all mechanisms for ignoring low-information stimuli[49,50]. However, these do not necessarily result in an ICA-like representation of the incoming sensory information.

The retina removes some redundancy from visual inputs, perhaps to reduce the number of retinal ganglion cell axons required to convey visual information centrally[51,52]. Redundant information tends to be local in visual scenes, so that redundant information is often collected by neighboring photoreceptors. To reduce local redundancies, neurons inhibit their neighbors, which could be accomplished by lateral inhibition from local inhibitory interneurons.

The wire-length considerations discussed above indicate that local redundancies also exist in the CNS (Refs 21–23). It might sometimes be necessary to remove redundancy in order to reduce wiring volume, especially in long-distance projections from one population of neurons to another, and redundancy reduction might be one general function of local inhibitory interneurons.

Redundancy reduction in the visual pathway also allows the cortex to use an efficient representation of visual information[47]. Properties of the receptive fields of neurons in the primary visual cortex of mammals resemble the regularities found in natural scenes by ICA (Refs 53–55). According to this view, the

connections from the thalamus to the primary visual cortex and the neurons in the cortex itself (e.g., layer IV pyramidal cells) form something that resembles an ICA unmixing matrix (Box 2; Fig. 1). Cortical neurons represent patterns of inputs that are each approximately independent over the ensemble of natural images.

Conclusions

Although some classical algorithms for data analysis date back more than a century, ICA is only about a decade old. Until we know more about how each of the ICA algorithms performs on different data sets, we should be cautious when applying and interpreting results. In particular, assumptions, such as no time delays and as many detectors as sources, should be checked, and results from a range of data sizes should be compared. As with any other technique, ICA can be misused, but if it is used properly it can provide a powerful tool for eavesdropping on neural-cocktail parties.

References

1 Cherry, E.C. (1953) Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 24, 975–979

2 Cardoso, J-F. and Souloumiac, A. (1993) Blind beamforming for non-Gaussian signals. *IEE Proceedings-F* 140, 362–370

3 Comon, P. (1994) Independent component analysis, a new concept? *Sig. Proc.* 36, 287–314

4 Bell, A.J. and Sejnowski, T.J. (1995) An information-maximisation approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1004–1034

5 Amari, S. *et al.* (1996) A new learning algorithm for blind signal separation. *Adv. Neural Inf. Proc. Syst.* 8, 757–763

6 Nakashima, M. *et al.* (1992) 448-detector optical recording system: Development and application to *Aplysia* gill-withdrawal reflex. *IEEE Trans. Biomed. Eng.* 39, 26–36

7 Meister, M. *et al.* (1994) Multi-neuronal signals from the retina: Acquisition and analysis. *J. Neurosci. Methods* 51, 95–106

8 Nordhausen, C.T. *et al.* (1996) Single unit recording capabilities of a 100 microelectrode array. *Brain Res.* 726, 129–140

9 Nicolelis, M.A.L. *et al.* (1997) Reconstructing the engram: Simultaneous, multisite, many single neuron recordings. *Neuron* 18, 529–537

10 Wu, J.Y. *et al.* (1998) Voltage-sensitive dyes for monitoring multineuronal activity in the intact central nervous system. *Histochem. J.* 30, 169–187

11 Brown, G.D. *et al.* Motor pools in the *Tritonia* swimming neural network revealed by optical recording. *J. Neurophysiol.* (in press)

12 Jutten, C. and Herault, J. (1991) Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture. *Sig. Proc.* 24, 1–10

13 Abeles, M., and Goldstein, H. M. (1977) Multispike train analysis. *Proc. IEEE* 65, 762–772

14 Andreassen, S. *et al.* (1979) Application of optimal multichannel filtering to simulated nerve signals. *Biol. Cybern.* 32, 25–33

15 Jansen, R.F. and Ter Maat, A. (1992) Automatic wave form classification of extracellular multineuron recordings. *J. Neurosci. Methods* 42, 123–132

16 Lewicki, M.S. (1998) A review of methods for spike sorting: the detection and classification of neural action potentials. *Network* 9, 53–78

17 Fee, M.S. *et al.* (1996) Variability of extracellular spike waveforms of cortical neurons. *J. Neurophysiol.* 76, 3823–3833

18 Snider, R.K. and Bonds, A.B. (1998) Classification of non-stationary neural signals. *J. Neurosci.* 84, 155–66

19 McNaughton, B.L. *et al.* (1983) The stereotrode: A new technique for simultaneous isolation of several single units in the central nervous system from multiple unit records. *J. Neurosci. Methods* 8, 391–397

20 Rebrik, S.P. *et al.* (1997) Cross channel correlations in tetrode recordings: Implications for spike-sorting. *Neurocomputing* 26–27, 1033–1038

21 Mitchison, G. (1992) Axonal trees and cortical architecture. *Trends Neurosci.* 15, 122–126

22 Cherniak, C. (1995) Neural component placement. *Trends Neurosci.* 18, 522–527

23 Chklovskii, D. (2000) Optimal sizes of dendritic and axonal arbors in a topographic projection. *J. Neurophysiol.* 83, 2113–2119

24 Hume, R. I. *et al.* (1982) Motor organization of *Tritonia* swimming. I. Quantitative analysis of swim behavior and flexion neuron firing patterns. *J. Neurophysiol.* 47, 60–74

25 Brown, G.D. (1997) Isolated-brain parallels to simple types of learning and memory in *Tritonia*. *Physiol. Behav.* 62, 509–518

26 Yamada, S. *et al.* (1992) Data processing for multi-channel optical recording: Action potential detection by neural network. *J. Neurosci. Methods* 43, 23–33

27 Gonzalez J.E. and Tsien R.Y. (1997) Improved indicators of cell membrane potential that use fluorescence resonance energy transfer. *Chem. Biol.* 4, 269–77

28 Pearlmutter, B.A. and Parra, L.C. (1996) In *International Conference on Neural Information Processing (Hong Kong)* pp. 151–157, Springer–Verlag

29 Lewicki, M.S. and Sejnowski, T.J. (2000) Learning overcomplete representations. *Neural Comput.* 12, 337–365

30 Hyvarinen, A. and Oja, E. (1997) A fast fixed-point algorithm for independent component analysis. *Neural Computat.* 9, 1483–1492

31 Koehler, B.U., and Orglmeister, R. Independent Component Analysis Using Autoregressive Models. First International Workshop on Independent Components Analysis and Blind Signal Separation–ICA '99. pp. 359–363

32 Attias, H. (1998) Independent factor analysis. *Neural Computat.* 11, 803-851

33 Amari, S. (1999) Superefficiency in blind source separation, *IEEE Trans. Sig. Proc.* 47, 936–944

34 Zibulevsky, M. and Pearlmutter, B.A. (1999) *Blind Source Separation by Sparse Decomposition.* University of New Mexico Technical Report No. CS99–1, Albuquerque, NM, USA.

35 McKeown, M.J. *et al.* (1998) Spatially independent activity patterns in functional magnetic resonance imaging data during the Stroop color-naming task. *Proc. Natl. Acad. Sci. U. S. A.* 95, 803–810

36 Laubach, M. *et al.* 1999) Independent component analyses for quantifying neuronal ensemble interactions. *J. Neurosci. Methods* 94, 141–154

37 Laubach, M. *et al.* (2000) Cortical ensemble activity increasingly predicts behavior outcomes during learning of a motor task. *Nature* 405, 567–571

38 Torkkola, K. (1996) Blind separation of convolved sources based on information maximization. In *Neural Networks for Signal Processing VI, IEEE.* New York

39 Attias, H. and Schreiner, C.E. (1998) Blind source separation and deconvolution: The dynamic component analysis algorithm. *Neural Comput.* 10, 1373–1424

40 Recce, M.L. and O'Keefe, J. (1989) The tetrode: A new technique for multi-unit extracellular recording. *Soc. Neurosci. Abstr.* 15, 1250

41 Makeig, S. *et al.* (1997) Blind separation of auditory event-related brain responses into independent components. *Proc. Natl. Acad. Sci. U. S. A.* 94, 10979–10984

42 McKeown, M.J. *et al.* (1999) *Brain Topogr.* 12, 1–10

43 Laughlin, S. (1981) A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch., C* 36, 910–912

44 Linsker, R. (1992) Local synaptic learning rules suffice to maximize mutual information in a linear network. *Neural Comput.* 4, 691–702

45 Hebb, D.O. (1949) The organization of behavior, a neuropsychological theory. Wiley, New York

46 von der Malsburg, C. and Schneider, W. (1986) A neural cocktail-party processor. *Biol. Cybern.* 54, 29–40

47 Bushara, K.O. *et al.* (1999) Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nat. Neurosci.* 2, 759–766

48 Barlow, H.B. (1989) Unsupervised learning. *Neural Comput.* 1, 295–311

49 Srinivasan, M.V. *et al.* (1982) Predictive coding: A fresh view of inhibition in the retina. *Proc. R. Soc. London Ser. B* 216, 427–459

50 Rao, R.P. (1999) An optimal estimation approach to visual perception and learning. *Vis. Res.* 39, 1963–1989

51 Atick, J.J. and Redlich, A.N. (1990) Toward a theory of early visual processing. *Neural Comput.* 2, 308–320

52 Meister, M. (1996) Multineuronal codes in retinal signaling. *Proc. Natl. Acad. Sci. U. S. A.* 93, 609–614

53 Olshausen, B.A. and Field, D.J. (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 607–609

54 Bell, A.J. and Sejnowski, T.J. (1997) The 'independent components' of natural scenes are edge filters. *Vis. Res.* 37, 3327–3338

55 van Hateren, J.H. and Ruderman, D.L. (1998) Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proc. R. Soc. London Ser. B* 265, 2315–2320