Deblending random seismic sources via independent component analysis

Pawan Bharadwaj*, Laurent Demanet, and Aimé Fournier, Massachusetts Institute of Technology

SUMMARY

We consider the question of deblending for seismic shot records generated from simultaneous random sources at different locations, i.e., how to decompose them into isolated records involving one source at a time. As an example, seismic-while-drilling experiments use active drill-string sources and receivers to look around and ahead of the borehole, but these receivers also record noise from the operation of the drill bit. A conventional method for deblending is independent component analysis (ICA), which assumes a "cocktail-party" mixing model where each receiver records a linear combination of source signals assumed to be statistically independent, and where only one source can have a Gaussian distribution. In this note, we extend the applicability of ICA to seismic shot records with markedly more complex mixing models with unknown wave kinematics, provided the following assumptions are met.

1. The active source is fully controllable, which means that it can be used to input a wide range of non-Gaussian random signals into the subsurface.

2. The waves are a linear function of the source, have a finite speed of propagation, and follow finite-length paths.

The last assumption implies a scale separation, in frequency, between the mixing matrix elements (Green's functions) and the random input signals. In this regime, we show that the key to the success of ICA is careful windowing to frequency bands over which the Green's functions are approximately constant.

INTRODUCTION

There are situations where seismic experiments are to be performed in noisy environments. For example, the records of a look-ahead seismic system in a borehole environment are contaminated due to the noise generated by the operating drill bit (Rector III and Marion, 1991; Joyce et al., 2001; Aminzadeh and Dasgupta, 2013). As a result, the receivers will be recording *blended* data from both the active source and the drill-bit source. The drill-bit operation could be paused, while performing the seismic experiment, but this will increase the costs associated with the drilling. The blended records could be used for imaging if the drill-bit signal were exactly known *a priori* and used to perform interferometry (by template matching or deconvolution), but that is unrealistic. Estimating the drill-bit signal directly from data has traditionally been considered to be difficult, for the following reasons: 1. The drill-bit signal is not impulsive, so it lacks the distinguishing features that would allow event picking; 2. The wave-propagation model may contain more than one arrival, so estimating delays is not sufficient information to be able to extract the pulse; 3. It is impossible to tell a source signature vs. a Green's function with only one receiver, so any method that hopes to lift the ambiguity necessarily requires two or more receivers – and the number of receivers on a drill string cannot be large.

Conventionally, independent component analysis (ICA) is used for blind source separation (BSS) and blind deconvolution in audio signal processing (Hyvärinen and Oja, 2000). In this note, we apply and extend frequency-domain ICA for BSS (Makino et al., 2005) to deal with the seismic deblending problem.

The model that ICA can handle is the linear mixture

$$\begin{pmatrix} D_1(\omega) \\ \vdots \\ D_{n_r}(\omega) \end{pmatrix} = \mathbf{H} \begin{pmatrix} B(\omega) \\ S(\omega) \end{pmatrix}, \tag{1}$$

where $B(\omega)$ and $S(\omega)$ are the frequency-dependent source signals assumed to arise from statistically independent random processes;[1] $D_1(\omega)$ through $D_{n_r}(\omega)$ are the blended signals; and $\mathbf{H}$ is a $n_r \times 2$ matrix of numbers. In its simplest incarnation, ICA finds $\mathbf{H}$ as the linear transformation that makes $B(\omega)$ and $S(\omega)$ as close to statistically independent as possible (Hyvärinen and Oja, 2000; Bell and Sejnowski, 1995). At most one process may be Gaussian for ICA to work.

In the geophysical context, $B(\omega)$ and $S(\omega)$ could respectively denote the drill-bit and active source signatures; the matrix $\mathbf{H}$ with columns $\mathbf{H}_1$ and $\mathbf{H}_2$ would encode wave propagation and scattering; $\mathbf{H}_1 B$ would be the drill-bit source contribution to the shot record; and $\mathbf{H}_2 S$ would be the active source contribution to the shot record. However, the simple instantaneous mixture model in Equation 1 is unrealistic in geophysics, because it assumes instantaneous propagation of the signals. Instead, we explain in the next section why the delays associated with multiple scattering contribute to a frequency-dependent mixing matrix $\mathbf{A}(\omega)$, so that Equation 1 would be generalized to a convolutive mixing model.

In the case of audio signals, blind source separation (BSS) for convolutive mixtures using ICA is performed in either the frequency or time domain (Pedersen et al., 2007). In frequency-domain BSS algorithms, the deblending problem is transformed into an instantaneous one in various narrow frequency bins, so that conventional ICA methods can be directly used (Makino et al., 2005).

If the goal is to go beyond deblending (i.e., recovery of $\mathbf{H}_1 B$ and $\mathbf{H}_2 S$) and perform deconvolution (i.e., recovery of B and S), the results from ICA from all such bands need to be combined together to get the final output. This piecing back or combination operation is not trivial as the outputs of the ICA algorithms in each frequency bin have unknown row order (permutation) and scaling, resulting in some fundamental problems in frequency domain BSS algorithms (Araki et al., 2003). This issue may jeopardize the applicability of ICA, but this note illustrates that these complications do not hinder the *deblending* goal of BSS.

---

[1]They are functions of frequency, but the mixture model can equivalently be written with functions of time.

Independent component analysis

We analyze the case where the active source signal is fully controllable, so that it can be programmed as a sample of a deliberately non-Gaussian random process, so that separation by ICA is possible.

MIXING MODEL

We consider a source at the location of the drill bit, $\mathbf{x}_b$, generating a band-limited random signal, $B(\omega)$. An active source is located at $\mathbf{x}_s$ and inputs a given random signal $S(\omega)$ into the subsurface. Here ω denotes the angular frequency that belongs to some interval Ω. The locations of the receivers are denoted by $\mathbf{x}_r$. In the Born approximation, the measured records at the receivers are given by

$$
\begin{aligned}
D_r(\omega) \quad = \quad & G(\mathbf{x}_b, \mathbf{x}_r; \omega)B(\omega) \\
& + \int_{\mathbf{x}} G(\mathbf{x}_b, \mathbf{x}; \omega)m(\mathbf{x})G(\mathbf{x}, \mathbf{x}_r; \omega)B(\omega)\mathrm{d}\mathbf{x} \\
& + G(\mathbf{x}_s, \mathbf{x}_r; \omega)S(\omega) \\
& + \int_{\mathbf{x}} G(\mathbf{x}_s, \mathbf{x}; \omega)m(\mathbf{x})G(\mathbf{x}, \mathbf{x}_r; \omega)S(\omega)\mathrm{d}\mathbf{x}, \quad (2)
\end{aligned}
$$

where m denotes the unknown subsurface reflectivity distribution and G denotes the subsurface Green's function in the frequency domain. The goal of seismic imaging is to estimate m, which is achievable when B and S are estimated by deconvolution. Otherwise, at least deblending has to be performed to decompose the recorded data, as if only one source were used at a time. Then the isolated active source records can be used for imaging after a cross-correlation with the known active source signals. When n_r receivers are considered, the data vector D can be written as the product of a source vector S and a mixing matrix $\mathbf{A}$ as:

$$
\underbrace{\begin{pmatrix} D_1(\omega) \\ \vdots \\ D_{n_r}(\omega) \end{pmatrix}}_{\mathsf{D}(\omega)} = \underbrace{\begin{pmatrix} A_{b,1}(\omega) & A_{s,1}(\omega) \\ \vdots & \vdots \\ A_{b,n_r}(\omega) & A_{s,n_r}(\omega) \end{pmatrix}}_{\mathbf{A}(\omega)} \underbrace{\begin{pmatrix} B(\omega) \\ S(\omega) \end{pmatrix}}_{\mathsf{S}(\omega)}. \quad (3)
$$

The mixing matrix is of dimension $n_r \times 2$ with elements

$$
\begin{aligned}
A_{j,r}(\omega) \quad = \quad & G(\mathbf{x}_j, \mathbf{x}_r; \omega) \\
& + \int_{\mathbf{x}} G(\mathbf{x}_j, \mathbf{x}; \omega)m(\mathbf{x})G(\mathbf{x}, \mathbf{x}_r; \omega)\mathrm{d}\mathbf{x}, \quad (4)
\end{aligned}
$$

where j can either be b or s. The deblended signals at a receiver with index r are given by

$$
\mathsf{Q}(\omega) = \begin{pmatrix} Q_{b,r} \\ Q_{s,r} \end{pmatrix} = \begin{pmatrix} A_{b,r}B \\ A_{s,r}S \end{pmatrix}. \quad (5)
$$

Note that a windowed Fourier transform with length T is applied to the time-domain data that are recorded in the field in order to form the data vector D in the first place. Also note that a similar mixing equation, similar to Equation 3, can be written even if there is multiple scattering, albeit with more complicated matrix $\mathbf{A}$ elements.

We now aim to estimate the random source signals B and S, and the elements of the matrix $\mathbf{A}$ of Equation 3. This is the subject of frequency-domain blind source separation methods (Pedersen et al., 2007), where a frequency bin $\Omega_a \subset \Omega$ is considered, in which the variations of $\mathbf{A}$ can be ignored. Therefore, we have $\mathbf{A}(\omega) \approx \mathbf{H}$, a constant matrix, $\forall \omega \in \Omega_a$. The maximum width of this frequency bin Ω_a, denoted by $|\Omega_a|$, is given by $\frac{2\pi}{\tau}$, where τ is the maximum traveltime of the waves propagating from sources to the receivers. In every Ω_a, we can write an instantaneous mixing model using the frequency-independent matrix $\mathbf{H}$ as $\mathsf{D}_a = \mathbf{H}\mathsf{S}_a$, where the subscript a is used to denote the element-wise multiplication of the data and source vectors with a boxcar function of support Ω_a. This instantaneous mixing model can be easily solved using ICA to output a separation or unmixing matrix $\mathbf{W}$ and its corresponding estimated source signal vector $\widehat{\mathsf{S}}_a$ such that $\widehat{\mathsf{S}}_a = \mathbf{W}\mathsf{D}_a$, where $\mathsf{S}_a = \mathbf{L}\mathbf{P}\widehat{\mathsf{S}}_a$. Here $\mathbf{L}$ and $\mathbf{P}$ are a diagonal scaling and a permutation matrix that are necessary to match the estimated source signals $\widehat{\mathsf{S}}_a$ to the actual source signals S_a; however, as is well known, $\mathbf{L}$ and $\mathbf{P}$ are individually undetermined by ICA.

Furthermore, in order to perform ICA in Ω_a, we need to have as many random samples of S_a as possible (see numerical example in Figure 1a). In other words, the frequency resolution $\Delta\omega = \frac{2\pi}{T}$, scale at which D_a and S_a oscillate, has to be much smaller than $|\Omega_a|$. This is possible by appropriately choosing T for the windowed Fourier transform such that $T \gg \tau$. On the other hand, T is limited by the total recording time at the receivers. The seismic imaging system will be impractical when the recording time is too large compared to the wave-propagation time. Note that the propagation time τ increases, when there is multiple scattering.

Now, after estimating $\widehat{\mathsf{S}}_a$ in every Ω_a with an assumption that the elements of S_a are statistically independent, the next step is to combine the outputs together to form an estimated source signal vector over the entire interval Ω. We denote such a source signal vector by $\widehat{\mathsf{S}}$, it matches the actual source signal vector S up to a global permutation and scaling i.e., $\mathsf{S} = \mathbb{L}\mathbb{P}\widehat{\mathsf{S}}$, where $\mathbb{P}$ and $\mathbb{L}$ denote global permutation and scaling matrices. The combining operation is trivial only if the scaling and permutation matrices, $\mathbf{L}$ and $\mathbf{P}$, in every Ω_a are known. Otherwise, the reconstruction of $\widehat{\mathsf{S}}$ suffers from the fact that $\widehat{\mathsf{S}}_a$ has arbitrary scaling and row order depending on the choice of Ω_a (Araki et al., 2003).

Many authors propose using a priori information to estimate the permutation matrix $\mathbf{P}$ (Pedersen et al., 2007). For example, Soon et al. (1993) and Prasad et al. (2004) use the information about the direction of wave arrival at the receivers from each source to sort the elements of $\widehat{\mathsf{S}}_a$. Another common method to solve the permutation problem is to order the output source components based on their Gaussianity. Low et al. (2004) uses excess kurtosis as measure to differentiate a source signal of interest from an interference signal that is more Gaussian. In this note, we follow Low et al. (2004) and consider that the drill-bit source is closer to Gaussian than the active source.

For the choice of $\mathbf{L}$, we follow the simple prescription in Matsuoka (2002) and Makino et al. (2005), but note that the result of BSS (deblending only) does not depend on this choice. This approach estimates the vector Q of the deblended data at the receiver at $\mathbf{x}_r$, but not the source signatures. The isolated active source data in Q, which are $A_{s,r}S$, can be used for imaging after
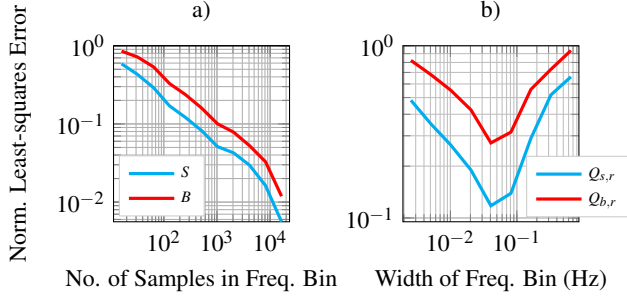
Figure 1: a). Instantaneous mixture example: relative least-squares error between the ICA estimated source signals and the actual signals is plotted as a function of the number of samples in each frequency bin. Red and cyan curves correspond to the drill bit and active sources, respectively. b). Same as a), but for the convolutive mixture in the Marmousi model. The error in the deblended records due to individual sources is plotted as a function of $|\Omega_a|$.

cross-correlation with known active source signals S.

CHOICE OF RANDOM SIGNALS

In this section, we discuss the random-signal models for B and S in every band Ω_a. We model the drill-bit signal as a Gaussian process using random i.i.d. variables X_i:

$$B(\omega) = \sum_i X_i \, \text{sinc}(T[\omega - \tfrac{2\pi}{T}i]), \qquad X_i \sim N(0, \sigma^2) \qquad (6)$$

Here, $N(0, \sigma^2)$ denotes Gaussian distribution with zero mean and standard deviation σ. The standard deviations can be different for each X_i in order to allow model colored drill-bit noise, and our method would apply to that case as well. We used a sinc function with width $\frac{2\pi}{T}$ in the above equation to limit the time-window length of windowed Fourier transform to T.

To model the active signal input in the subsurface, we use a similar equation as above with random variables Y_j (instead of X_i), obeying a *non-Gaussian* distribution. ICA requires the random signals S and B to be statistically independent i.e., their joint probability distribution function is given by the product of its marginals, $p(S, B) = p(S)p(B)$. Obviously, statistical independence X_i and Y_j implies independence of S and B too.
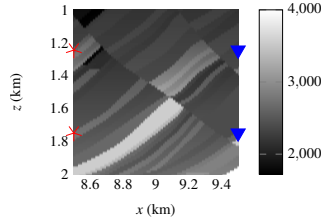

Figure 2: A $1\,\text{km} \times 1\,\text{km}$ section of the Marmousi II P-wave velocity model used for numerical experiments involving convolutive mixtures. The positions of sources and receivers are indicated by red stars and blue triangles, respectively.
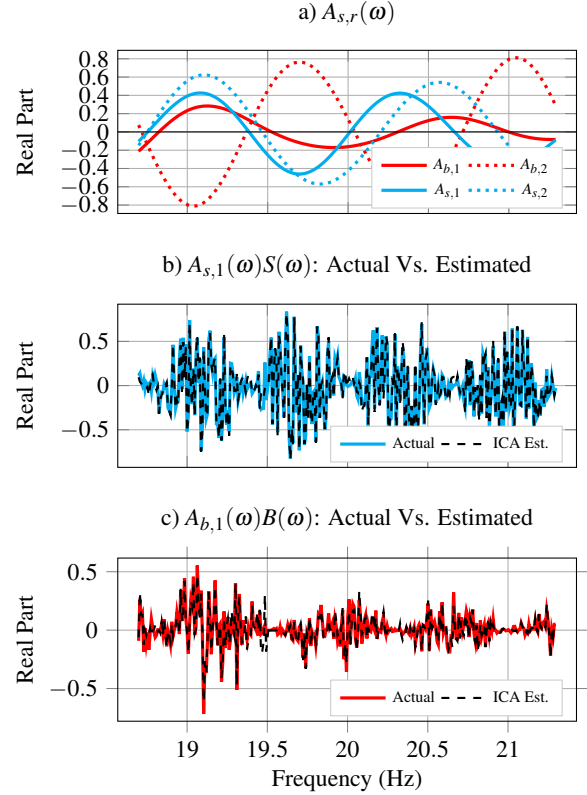

Figure 3: For the convolutive mixture in the Marmousi model: a) The elements of the mixing matrix are plotted as a function of frequency. Solid lines correspond to the receiver with index $r = 1$. b). Deblended records after scaling with $\mathbb{L}$ are compared to the active source (cyan) records. Only 1 in every 100 samples is considered for plotting. c). Same as b), except for the drill bit source (red).

NUMERICAL EXAMPLES

To estimate the separation matrix in a given Ω_a, in our examples, we used the FastICA algorithm (Hyvärinen, 1999) from the multivariate statistical analysis package in Julia (Bezanson et al., 2014). FastICA seeks an orthogonal rotation of pre-whitened data by minimizing negentropy of the rotated components. It uses the fact that a Gaussian random variable has minimum negentropy among all distributions with fixed first and second moments. We limited the number of FastICA iterations to 200.

Instantaneous Mixture

We consider a simple numerical example, where we choose the frequency band Ω of both the active and drill-bit sources to have 16384 samples. The goal of this example is to determine the minimum number of random samples that are necessary to estimate the statistics accurately using ICA. We generated the random signals B and S of the source vector S by picking samples from Gaussian and uniform distributions, respectively. Then, synthetic data at two receivers are modeled by assuming an instantaneous mixing. We divided Ω into various bins of equal sizes and perform ICA in each bin individually. Finally, the es-
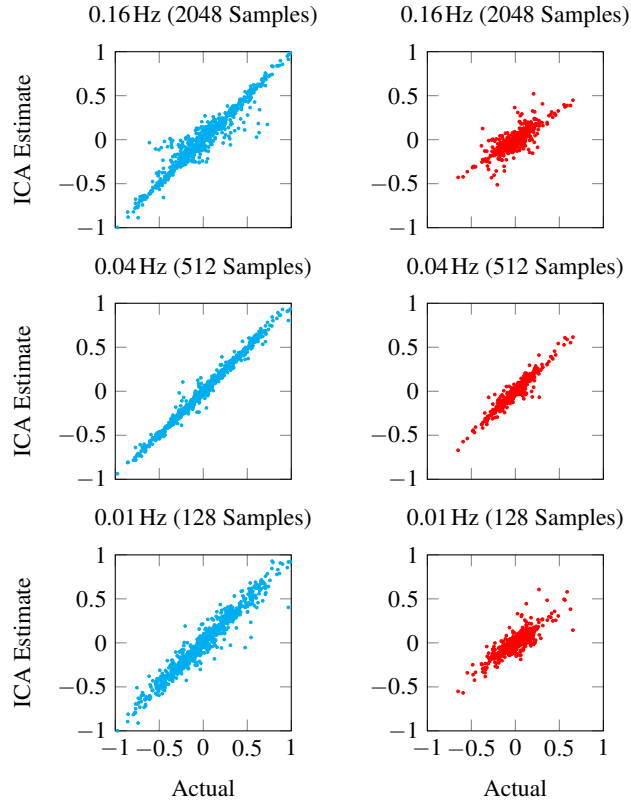
Figure 4: For the convolutive mixture in the Marmousi model: scatter plots between the deblended records Q and the actual drill bit (red) and active source (cyan) records for different frequency bin widths. Only 1 in every 40 samples is considered for plotting.

ulated time of $\tau = 1.2\,\mathrm{s}$. We only consider an arbitrary frequency band $[18.7, 21.3]\,\mathrm{Hz}$, which includes the dominant frequency, for simplicity, but without loss of generality. The random drill-bit signal B is generated by assuming $\sigma = 1$ in Equation 6. In order to generate the random active signals S, we picked samples of Y_i from a uniform distribution with $Y_{\min} = -2$ and $Y_{\max} = 2$. The samples for both these signals are picked such that the maximum recording time as well as the length of the time window for the short-time Fourier transform is $T = 1.2 \times 10^4\,\mathrm{s}$. In order to generate the synthetic data, D, at the receivers, the band-limited Green's functions are then convolved with the source signals, according to Equation 2.

Given D, we now aim to perform deblending at a receiver with index $r = 1$, such that the output vector Q contains isolated records due to individual sources. Therefore, the model in the Equation 3 has to be solved by dividing the entire frequency band into bins Ω_a, estimating statistically independent components in each bin, followed by combining all the outputs. As shown in the previous example, we expect the error in the deblended records to decrease with an increase in the width of each frequency bin $|\Omega_a|$. However, the assumption that the mixing matrix is independent of frequency is violated when a large $|\Omega_a|$ is chosen. Therefore there is an optimal choice of the frequency-bin width to best perform deblending. We plotted the relative least-squares error between the estimated vs. actual deblended records as a function of $|\Omega_a|$ in Figure 1b. We see that $|\Omega_a| = 0.04\,\mathrm{Hz}$ with 512 samples performs optimal deblending in this case. The deblended records for the optimal choice of $|\Omega_a|$ are compared to the actual records in Figures 3b and 3c, where 1 in every 100 samples is plotted. The scatterplots, in Figure 4, compare the distance between the estimated and actual Q for values of $|\Omega_a|$ greater and lesser than the optimal choice. It can be observed that the distance between the signals is greater for a non-optimal choice.

CONCLUSIONS

With an assumption that the active source is fully controllable in a drilling environment i.e., it can input any given random signal into the subsurface, we propose a deblending method that uses independent component analysis (ICA) to separate the active and drill-bit source signals. While ICA is conventionally used to solve the instantaneous cocktail-party mixing problem, the physically accurate mixing model is a more complex convolution with the Green's functions. Fortunately, there is a scale separation between those Green's functions and the random sources themselves, which enables ICA after a proper division of the frequency axis into small bins. We show the potential of the proposed method using simple numerical examples.

ACKNOWLEDGEMENTS

timated components in each bin are combined together to output $\widehat{\mathsf{S}}$. The local scaling ambiguity for instantaneous mixing is resolved by normalizing each column of the ICA-estimated mixing matrix, and the local permutation ambiguity is resolved by ordering the source signals using kurtosis. Before comparing the outputs, $\widehat{\mathsf{S}}$, with original signals, S, the overall scaling matrix $\mathbb{L}$ and permutation matrix $\mathbb{P}$ are determined by minimizing the norm $\|\mathbb{L}\mathbb{P}\widehat{\mathsf{S}} - \mathsf{S}\|_2$. Figure 1a plots the relative least-squares error between $\mathbb{L}\mathbb{P}\widehat{\mathsf{S}}$ and S. We see that a minimum number of samples should be maintained in each frequency bin in order to estimate the statistics accurately. If we aim for a relative error in the reconstructed signals below 10^{-1}, there should be at least 10^3 samples in each bin.

Convolutive Mixture using the Marmousi Model

Now we consider a convolutive mixture with two sources and two receivers. The active and drill bit sources are located at depths $1.75\,\mathrm{km}$ and $1.245\,\mathrm{km}$, respectively. The receivers are also present at these depths, but in a well $1\,\mathrm{km}$ away from the source well. We used a time-domain staggered-grid finite-difference acoustic solver for wave-equation modeling. Here, the P-wave velocity model in Figure 2 is used to generate the elements of the frequency-dependent mixing matrix **A** in Equation 3. These elements, plotted in Figure 3a, are numerically modeled with a Ricker source wavelet (peak frequency of $20\,\mathrm{Hz}$) and a total sim-

Independent component analysis

REFERENCES

Aminzadeh, F., and S. N. Dasgupta, 2013, Geophysics for petroleum engineers: Newnes, **60**.

Araki, S., R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, 2003, The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech: IEEE Transactions on Speech and Audio Processing, **11**, 109–116.

Bell, A. J., and T. J. Sejnowski, 1995, An information-maximization approach to blind separation and blind deconvolution: Neural computation, **7**, 1129–1159.

Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah, 2014, Julia: A fresh approach to numerical computing: arXiv preprint arXiv:1411.1607.

Hyvärinen, A., 1999, Fast and robust fixed-point algorithms for independent component analysis: IEEE Transactions on Neural Networks, **10**, 626–634.

Hyvärinen, A., and E. Oja, 2000, Independent component analysis: algorithms and applications: Neural Networks, **13**, 411 – 430; http://dx.doi.org/http://dx.doi.org/10.1016/S0893-6080(00)00026-5.

Joyce, B., D. Patterson, J. Leggett, and V. Dubinsky, 2001, Introduction of a new omni-directional acoustic system for improved real-time LWD sonic logging-tool design and field test results: Presented at the SPWLA 42nd Annual Logging Symposium, Society of Petrophysicists and Well-Log Analysts.

Low, S. Y., S. Nordholm, and R. Togneri, 2004, Convolutive blind signal separation with post-processing: IEEE Transactions on Speech and Audio Processing, **12**, 539–548.

Makino, S., H. Sawada, R. Mukai, and S. Araki, 2005, Blind source separation of convolutive mixtures of speech in frequency domain: IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, **88**, 1640–1655.

Matsuoka, K., 2002, Minimal distortion principle for blind source separation: SICE 2002. Proceedings of the 41st SICE Annual Conference, IEEE, 2138–2143.

Pedersen, M. S., J. Larsen, U. Kjems, and L. C. Parra, 2007, A survey of convolutive blind source separation methods: Multichannel Speech Processing Handbook.

Prasad, R., H. Saruwatari, and K. Shikano, 2004, An ica algorithm for separation of convolutive mixture of speech signals: International Journal of Information Technology, **2**, 273–283.

Rector III, J., and B. P. Marion, 1991, The use of drill-bit energy as a downhole seismic source: Geophysics, **56**, 628–634.

Soon, V., L. Tong, Y. Huang, and R. Liu, 1993, A robust method for wideband signal separation: Circuits and Systems, 1993., ISCAS'93, 1993 IEEE International Symposium on, IEEE, 703–706.