International Journal of Electronics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tetn20

Audio analysis of statistically instantaneous signals with mixed Gaussian probability distributions

Ganesh R. Naik [a] & Wenwu Wang [b]

[a] RMIT University, Melbourne, Australia - 3001

[b] Centre for Vision, Speech and Signal Processing, University of Surrey, UK

Version of record first published: 12 Jul 2012.

International Journal of Electronics
Vol. 99, No. 10, October 2012, 1333–1350

Taylor & Francis
Taylor & Francis Group

Audio analysis of statistically instantaneous signals with mixed Gaussian probability distributions

Ganesh R. Naik[a]* and Wenwu Wang[b]

[a]*RMIT University, Melbourne, Australia – 3001;* [b]*Centre for Vision, Speech and Signal Processing, University of Surrey, UK*

(*Received 26 October 2009; final version received 21 March 2011*)

In this article, a novel method is proposed to measure the separation qualities of statistically instantaneous audio signals with mixed Gaussian probability distributions. This study evaluates the impact of the Probability Distribution Function (PDF) of the mixed signals on the outcomes of both sub- and super-Gaussian distributions. Different Gaussian measures are evaluated by using various spectral-distortion measures. It aims to compare the different audio mixtures from both super-Gaussian and sub-Gaussian perspectives. Extensive computer simulation confirms that the separated sources always have super-Gaussian characteristics irrespective of the PDF of the signals or mixtures. The result based on the objective measures demonstrates the effectiveness of source separation in improving the quality of the separated audio sources.

Keywords: blind source separation; probability distribution function; independent component analysis; kurtosis; signal to interference ratio; sub-Gaussian; super-Gaussian

1. Introduction

Audio analysis and separation is the problem of automated separation of audio sources present in a room, using a set of differently placed microphones, capturing the auditory scene (Foote 1999; Benaroya, Bimbot, and Gribonval 2006; Dubnov, Tabrikian, and Targan 2006; Wilson 2007). The whole problem resembles the task a human can solve in a cocktail party situation, where using two sensors (ears), the brain can focus on a specific source of interest, suppressing all other sources present (Hyvarinen, Karhunen, and Oja 2001; He, Clifford, and Tarassenko 2006; Morita and Nanri 2006). Recently, Blind Source Separation (BSS) using Independent Component Analysis (ICA) has received a great deal of attention for its potential in acoustics, telecommunication, medical and image signal processing (Cristescu, Ristaniemi, Joutsensalo, and Karhunen 2000; Stone 2002; He et al. 2006; De Martino et al. 2007). BSS is an emerging technique, which enables the extraction of target speech from observed mixed speeches without the need for source positioning, spectral construction or a mixing system. To achieve this, attention is focused on a method based on ICA. ICA is a method for finding underlying components that are statistically independent from multivariate statistical data. ICA extracts independent components from mixtures (Bell and Sejnowski 1997; Hyvarinen et al. 2001).

*Corresponding author. Email: ganesh.naik@rmit.edu.au

ISSN 0020–7217 print/ISSN 1362–3060 online

http://dx.doi.org/10.1080/00207217.2011.582450
http://www.tandfonline.com

ICA is being used routinely to separate signals from different independent and nearly non-Gaussian sources. The applications of this include defence, surveillance, security, communication and entertainment. In the recent fast many researchers have studied the impact and quality of the sub- and super-Gaussian sources (Zarzoso and Nandi 2002; Blaschke and Wiskott 2003; Eriksson and Koivunen 2004; Zarzoso, Murillo-Fuentes, Boloix-Tortosa, and Nandi 2006). While the assumption of independence is important for the success of ICA, the impact of Probability Distribution Function (PDF) of the sources has not yet been considered in detail. The other issue is the impact of the mixing environment on the quality of separation of the sources. To make the source separation more effective for security and surveillance, there is a need to evaluate the reliability of the use of ICA for obtaining the separated signals.

The precise evaluation of speech quality is a key research problem that has attracted attention in the field of speech communication for several years. The two major techniques utilised in the evaluation of speech quality are subjective and objective speech quality procedures. Subjective quality measures are more accurate and robust since they are given by professional personnel who have received special assessment training, but they are essentially time consuming and expensive. In contrast, objective quality measures, instigated by speech signal processing procedures, offer a proficient, economical alternative to subjective procedures. Even though it is not advocated to use objective quality measures to entirely restore subjective measures, objective quality measures do illustrate the strong aptitude to forecast subjective quality measures and the results do associate very well with those generated by subjective quality measures. In this article, a novel work on audio analysis of statistically instantaneous signals with mixed Gaussian probability distributions is reported. The Gaussian distortion introduced in the resulting speech is measured by approximating objective measures of perceptual speech quality such as log-likelihood ratio (LLR) measure, log-area-ratio (LAR) measure, Itakura–Saito distortion (IS) Measure and weighted spectral slope (WSS) measure (Hansen and Pellom, 1998). The discrepancy of these estimated objective measures of the spectral distortion is studied and analysed to see specific audio effects of the Gaussian PDF qualities.

2. Theory

2.1. *Non-Gaussian and Gaussian measures*

Non-Gaussianity is an important and essential principle in ICA estimation. To use non-Gaussianity in ICA estimation, there needs to be quantitative measure of non-Gaussianity of a signal. Before using any measures of non-Gaussianity, the signals should be normalised (Lee 1998; Cichocki and Amari 2002).

There are actually two types of non-Gaussian signals. The two non-Gaussian signals are known by various terms, such as super-Gaussian and sub-Gaussian or equivalently known as ‘platy kurtotic’ and ‘lepto kurtotic’ respectively.

- Super-Gaussian sources

  A signal with super-Gaussian PDF has most of its values clustered around zero. A speech signal is a typical example for a super-Gaussian source. Figure 1(a) shows a typical super-Gaussian source (speech signals). From the figure it is also evident that the super-Gaussian signals have PDFs that have more peaks than those of Gaussian signals.

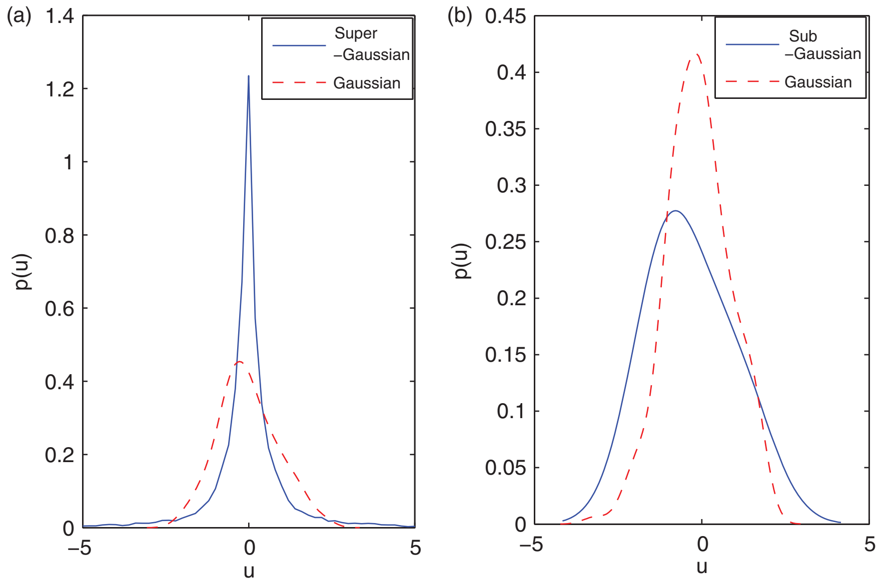

Figure 1. PDF of super- and sub-Gaussian sources. (a) super-Gaussian signal and (b) sub-Gaussian signal.

- Sub-Gaussian sources

  The signals with sub-Gaussian PDF have a wide distributed function, which is illustrated in Figure 1(b). A saw-tooth signal, polyphonic music signal and white noise signals are typical sub-Gaussian sources (Cristescu et al. 2000). The sub-Gaussian signals have PDFs that have less peaks than those of a Gaussian signals.

Some of the commonly used measures are kurtosis and entropy measures. Kurtosis is used as one of the measures in this article, which is explained next.

2.2. *Kurtosis*

Kurtosis is the classical method of measuring non-Gaussianity. When data is preprocessed to have unit variance, kurtosis is equal to the fourth moment of the data. The kurtosis of signal (s), denoted by $kurt(s)$, is defined by

$$kurt(s) = E(s^4) - 3(E\{s^2\})^2 \tag{1}$$

This is a basic definition of kurtosis using higher order (fourth order) cumulant, this simplification is based on the assumption that the signal has zero mean. To simplify things, we can further assume that (s) has been normalised, so that its variance is equal to one: $E\{s^2\} = 1$. Hence Equation (1) can be further simplified to

$$kurt(s) = E\{s^4\} - 3 \tag{2}$$

Equation (2) illustrates that kurtosis is a normalised form of the fourth moment $E\{s^4\} = 1$. For Gaussian signal, $E(s^4) = 3\{(E\{s^2\})^2\}$ and hence its kurtosis is zero. For most non-Gaussian signals, the kurtosis is nonzero. Kurtosis can be both positive or negative. Random variables that have positive kurtosis are called *super-Gaussian*, and those with negative kurtosis are called *sub-Gaussian*. Non-Gaussianity is measured using the absolute value of kurtosis or the square of kurtosis.

Kurtosis has been widely used as a measure of non-Gaussianity in ICA and related fields because of its computational simplicity. Theoretically, it has a linearity property such that

$$kurt(s_1 \pm s_2) = kurt(s_1) \pm kurt(s_2) \tag{3}$$

and

$$kurt(\alpha s_1) = \alpha^4\, kurt(s_1) \tag{4}$$

where α is a constant. Computationally, kurtosis can be calculated using the fourth moment of the sample data, by keeping the variance of the signal constant (Lee, Lewicki, and Sejnowski 1999). Kurtosis is extremely simple to calculate, however, it is very sensitive to outliers in the dataset (Pham, Garrat, and Jutten 1992).

2.3. *Independent component analysis*

ICA is a new statistical technique that aims at transforming an input vector into a signal space in which the signals are statistically independent (Lee, Girolami, Bell, and Sejnowski 2000; Hyvarinen et al. 2001; Jang and Lee 2003).

ICA assumes the mixing process as linear, so it can be expressed as:

$$x(t) = As(t) \tag{5}$$

where $x = [x_1(t), x_2(t), \ldots, x_n(t)]$ are the recordings, $s = [s_1(t), s_2(t), \ldots, s_n(t)]^T$ are the original signals and A is the $n \times n$ mixing matrix. This mixing matrix and each of the original signals are unknown. To separate the recordings to the original signals, an ICA algorithm performs a search of the un-mixing matrix W by which observations can be linearly translated to form independent output components so that:

$$s(t) = Wx(t) = WAs(t) \tag{6}$$

For this purpose, ICA relies strongly on the statistical independence of the sources, s. The block diagram approach of ICA for source separation is shown in Figure 2. The ICA technique iteratively estimates the un-mixing matrix using the maximisation of independence of the un-mixed signals as the cost function. Signals are statistically independent if the joint probability density of those components can be expressed as a product of their marginal probability density. It is important to observe the distinction between independence and uncorrelatedness, since decorrelation can always be performed by transforming the signals with a whitening matrix to get the identity covariance matrix. Independent signals are always uncorrelated, but uncorrelated, signals are not always independent. However, in the case of Gaussian signals, uncorrelatedness implies independence. Transformation of a Gaussian signal with any orthogonal un-mixing matrix or transform results in another Gaussian signal, and thus the original signals cannot be separated. Hence Gaussian signals are forbidden for ICA. Thus, the key to

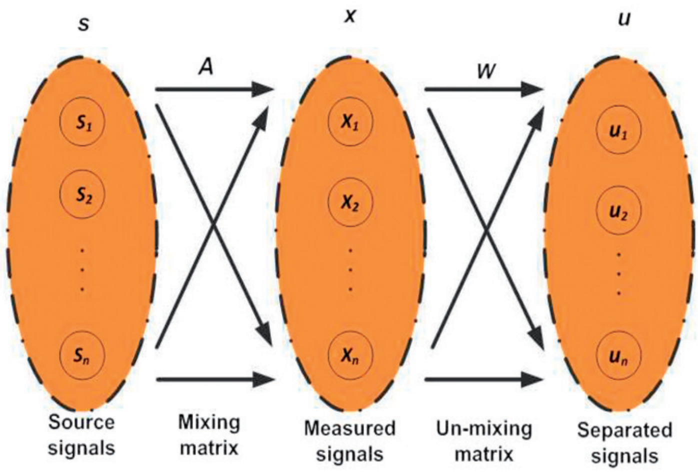

Figure 2. ICA block diagram. $s(t)$ are the sources. $x(t)$ are the recordings, $\hat{s}(t)$ are the estimated sources, A is mixing matrix and W is un-mixing matrix.

independent component estimation is measuring the non-Gaussianity of the signals (Cardoso 1998). In most applications such as the cocktail party problem, these are not serious problems. The supervisor is able to identify the different sources and determine the quality of the separation by listening to the sounds. To summarise from the above, the signals that can be separated need to be non-Gaussian and independent. For the purpose of applying ICA to audio recordings, there is a need to determine the conditions under which these signals can be considered as independent and non-Gaussian.

2.4. *Objective methods of speech quality measure*

Computational methods for objective evaluation of speech and audio quality are usually designed for measuring quality loss due to just a few specific types of signal degradation. Seeking an effective, yet general, objective method for speech and audio seems to be a pertinent and challenging R&D goal. How to measure the distortion between original source and the estimated one has no completely trivial solution. It depends on the mixing system and the separation process as well as the field of application. It is still hard to evaluate the separation algorithm because of the lack of appropriate performance measures even in the very simple case of linear instantaneous mixtures (Vincent, Gribonval, and Fevotte 2006; Campbell, Jones, and Glavin 2009).

In most cases, speech enhancement or noise reduction is measured in terms of improvement in signal-to-noise ratio (SNR), but in reality, this may not be the most suitable performance criteria for enhancement of perceptual speech quality. Humans do have a perceptive understanding of spoken language quality, nevertheless this may not be easy to quantify. From numerous studies, it has been shown that impact of noise on degradation of speech quality is non uniform. An objective speech quality gauge shows, the level of distortion for each frame, across time. Objective methods depend on mathematically based measure between reference signal and the signal under concern.

The objective measures are based on different parametric illustration of the speech, and be different due to inclusion or non-inclusion of several constraints and the different weightage given to them, in order to emulate auditory model and insight as closely as possible. The particulars of each one is specified below.

- SNR
- IS distortion measure,
- LAR measure
- LLR measure and
- WSS measure

In audio applications, SNR is a popular tool to measure the quality separation (Cichocki and Amari 2002). The SNR is explained next.

2.4.1. *Signal-to-noise ratio*

SNR is the ratio of the strength of the desired signal to the amplitude of noise signals at a given point in time, and can be measured in terms of the amplitude or power. It is an important measure to determine the quality of the signal. SNR is often expressed in decibels which is 20 times the logarithm of the amplitude ratio, or 10 times the logarithm of the power ratio. In this work, SNR ratio is computed as defined below:

$$\mathrm{SNR}_{dB} = 10\log_{10}\frac{P_s}{P_n} \tag{7}$$

2.4.2. *IS distortion measure*

If for an original clean frame of speech with linear prediction (LP) coefficient vector, $\vec{a}_\phi$, correlation matrix is R_ϕ. And for processed LP coefficient vector is $\vec{a}_d$, correlation matrix is R_d, then the IS distortion measure is given by

$$d_{\mathrm{IS}}(\vec{a}_d, \vec{a}_\phi) = \left[\frac{\sigma_\phi^2}{\sigma_d^2}\right]\left[\frac{\vec{a}_d R_\phi \vec{a}_d^T}{\vec{a}_\phi R_\phi \vec{a}_\phi^T}\right] + \log\left[\frac{\sigma_d^2}{\sigma_\phi^2}\right] - 1 \tag{8}$$

where σ_ϕ^2 and σ_d^2 represents all pole gain for the processed and clean speech, respectively.

2.4.3. *LAR measure*

The LAR measure is also based on dissimilarity of LP coefficients between original and processed speech signals. The LAR parameters are attained from the pth order LP reflection coefficients for the original $r_\Phi(j)$ and processed $r_d(j)$ signals from frame j. The LAR objective measure is shown as below:

$$d_{\mathrm{LAR}} = \left|\frac{1}{M}\sum_{i=1}^{M}\left[\log\frac{1 + r_\phi(j)}{1 - r_\phi(j)} - \log\frac{1 + \hat{r}_d(j)}{1 - \hat{r}_d(j)}\right]\right| \tag{9}$$

2.4.4. *LLR measure*

The LLR measure is also referred to as the Itakura distance (note that the ISD measure emphasises differences in general spectral shape versus an overall gain offset). The LLR measure is found as follows:

$$d_{\mathrm{LLR}} = (\vec{a}_d, \vec{a}_\phi) = \log\left(\frac{\vec{a}_d R_\phi \vec{a}_d^T}{\vec{a}_\phi R_\phi \vec{a}_\phi^T}\right) \tag{10}$$

where $\vec{a}_\phi$ is the LP coefficient vector, $\vec{a}_d$ is a processed speech coefficient vector and R_ϕ is the auto correlation matrix of the estimated signal.

2.4.5. *WSS measure*

The WSS measure by Klatt (1982) is based on an auditory model, in which 36 overlapping filters of progressively larger bandwidth are used to estimate the smoothed short-time speech spectrum. The measure finds a weighted difference between the spectral slopes in each band. The magnitude of each weight reflects whether the band is near a spectral peak or valley, and whether the peak is the largest in the spectrum. A per-frame measure in decibel is found as

$$d_{\mathrm{WSS}}(j) = K_{\mathrm{spl}}(K - \hat{K}) + \sum_{k=1}^{36} w_a(k)(S(k) - \hat{S}(k))^2 \tag{11}$$

where K and $\hat{K}$ are related to overall sound pressure level of the original and enhanced utterances and K_{spl} is a parameter which can be varied to increase the overall performance.

3. Methods

The performance of the proposed method is tested using various datasets under following conditions:

- The first step is mixing the various audio sources with sub-Gaussian and super-Gaussian mixing matrices and to separate them using ICA method.
- The second step is using these separated signals to measure the quality of separation using different speech quality measures.
- The third step is to estimate the Gaussianity and PDF of the sources using the Kurtosis values.
- The last step is to test the quality of the real audio mixtures using the above test conditions.

The methodology along with the results is as follows: as a first step, computer simulations are conducted to perform the source separation in sub-Gaussian and super-Gaussian mixing conditions. To test this condition, four different sub-Gaussian (polyphonic music) signals and super-Gaussian (speech signals) signals are chosen. They are shown in Figure 3. These signals are mixed using different combinations of

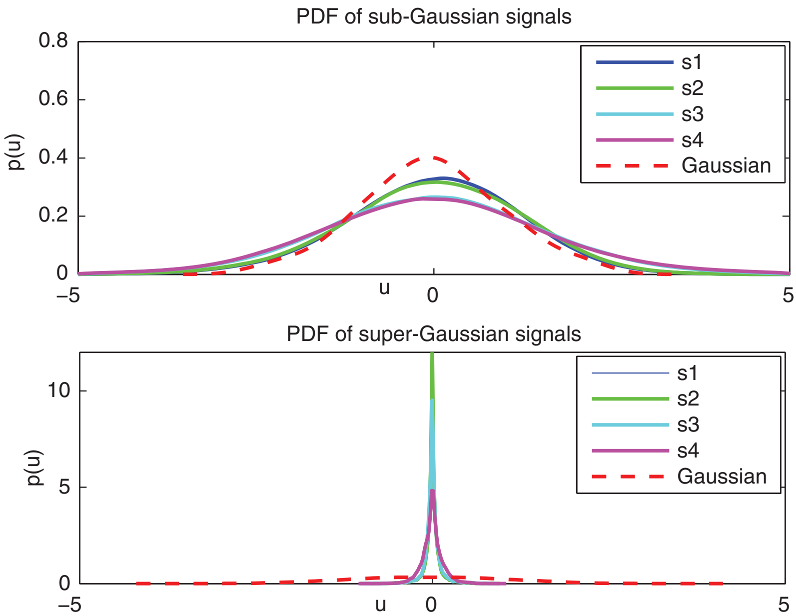

Figure 3. PDF of sub- and super-Gaussian sources.

sub-Gaussian and super-Gaussian mixing matrices. One of the examples of mixing matrices are shown below.

$$\text{Sub-Gaussian matrix} = \begin{pmatrix} -1.0000 & -1.0000 & 1.0000 & -1.0000 \\ -0.5250 & -3.4057 & 0.1582 & -2.6094 \\ -1.0000 & -1.0000 & 1.0000 & 1.0000 \\ 2.2949 & -1.1210 & 0.8096 & -3.9043 \end{pmatrix}$$

$$\text{Super-Gaussian matrix} = \begin{pmatrix} 0.0790 & 0.0309 & 0.0049 & 0.0012 \\ 0.0111 & 0.0444 & 0.1000 & 0.1778 \\ 0.0309 & 0.2086 & 0.0000 & 0.0790 \\ 0.3568 & 0.2420 & 0.1494 & 0.0790 \end{pmatrix}$$

Each time the PDF of the signals and mixing matrices were plotted to ensure the validity of the sub- and super-Gaussian sources. The PDF distribution of the sub- and super-Gaussian mixing matrices are shown in Figure 4.

Four independent audio recordings s_1, s_2, s_3 and s_4 were considered. The sub- and super-Gaussian sources were mixed in the following way:

- sub-Gaussian source + sub-Gaussian mixing matrix
- sub-Gaussian source + super-Gaussian mixing matrix
- super-Gaussian source + sub-Gaussian mixing matrix
- super-Gaussian source + super-Gaussian mixing matrix

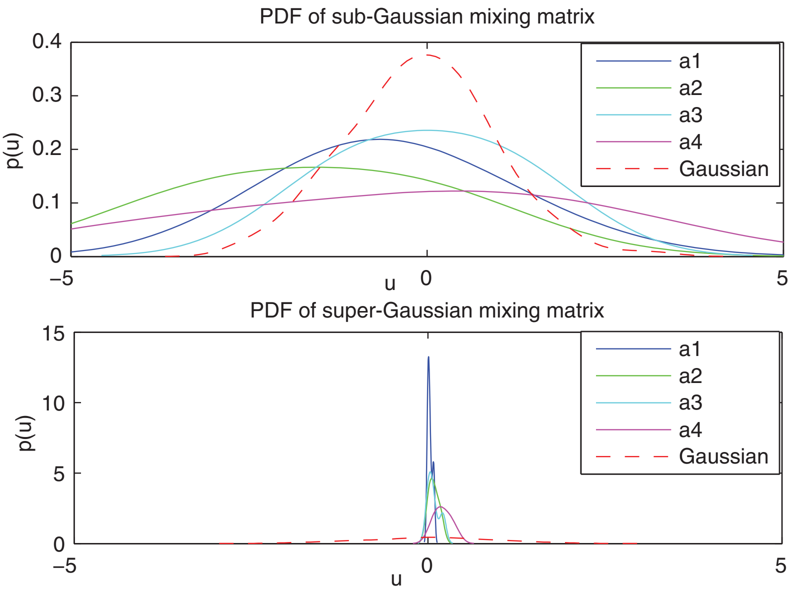

Figure 4. PDF of sub- and super-Gaussian mixing matrices.

The entire mixing process can be expressed in the vector and matrix form as:

$$\begin{aligned} x_1 &= a_{11}s_1 + a_{12}s_2 + a_{13}s_3 + a_{14}s_4 \\ x_2 &= a_{21}s_1 + a_{22}s_2 + a_{23}s_3 + a_{24}s_4 \\ x_3 &= a_{31}s_1 + a_{32}s_2 + a_{33}s_3 + a_{34}s_4 \\ x_4 &= a_{41}s_1 + a_{42}s_2 + a_{43}s_3 + a_{44}s_4 \end{aligned} \tag{12}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{pmatrix}$$

An example of the mixed signal in sub- and super-Gaussian environment are shown in Figure 5. Similar experiments were repeated for all the four above-mentioned mixing processes.

4. Data analysis

At a first step, to investigate the potential of the proposed ICA-based method for testing the quality of the source separation in sub- and super-Gaussian mixing conditions, a few examples of mixture streams of four audio signals are used. The mixed signals in sub- and super-Gaussian environment are separated using Fast ICA algorithm (Hyvärinen 1999). The ICA separated sources are initially evaluated using the Kurtosis parameters.

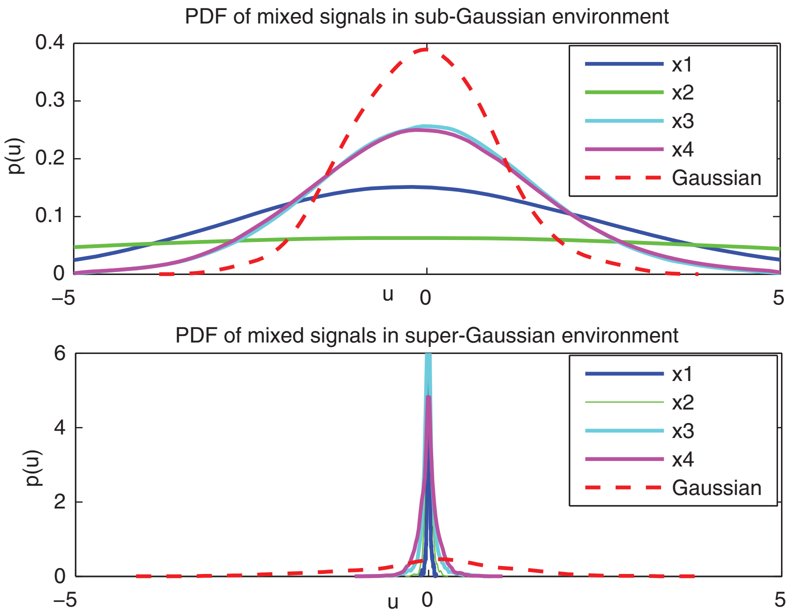

Figure 5. Examples of PDF of sub- and super-Gaussian mixed signals.

The second step is to measure the quality of the separation using objective quality assessment. Objective methods rely on a mathematically based measure between the original and coded/degraded speech signal. The success of these measures rests with their correlation with subjective quality. Objective quality measure results are presented in four areas. It may be noted that there are several ways to obtain overall quality scores. The real audio recordings are analysed using four objective audio measures such as IS, LLR, LAR and WSS at different SNRs. For most measures, finding a mean across a large test set is reasonable. If users want a general measure of performance the mean of the resulting frame-level scores is more useful. Hence, mean estimates of LLR, LAR, IS and WSS are computed for SNR range from 30 to -20 dB for real audio recordings. The real audio mixtures are tested for four categories (super-Gaussian source + super-Gaussian mixing matrix, super-Gaussian source + sub-Gaussian mixing matrix, sub-Gaussian source + super-Gaussian mixing matrix and sub-Gaussian source + sub-Gaussian mixing matrix) using both noisy and clear speech.

5. Results and observations

In this section, the separation performance of the statistically instantaneous signals mixed in sub-Gaussian and super-Gaussian audio conditions using various audio quality measures are examined.

The ICA estimated Gaussian mixtures are plotted in Figures 6 and 7, respectively. The overall kurtosis results for four different categories have been tabulated in Table 1. From the results, it is evident that kurtosis value is always positive. The positive kurtosis

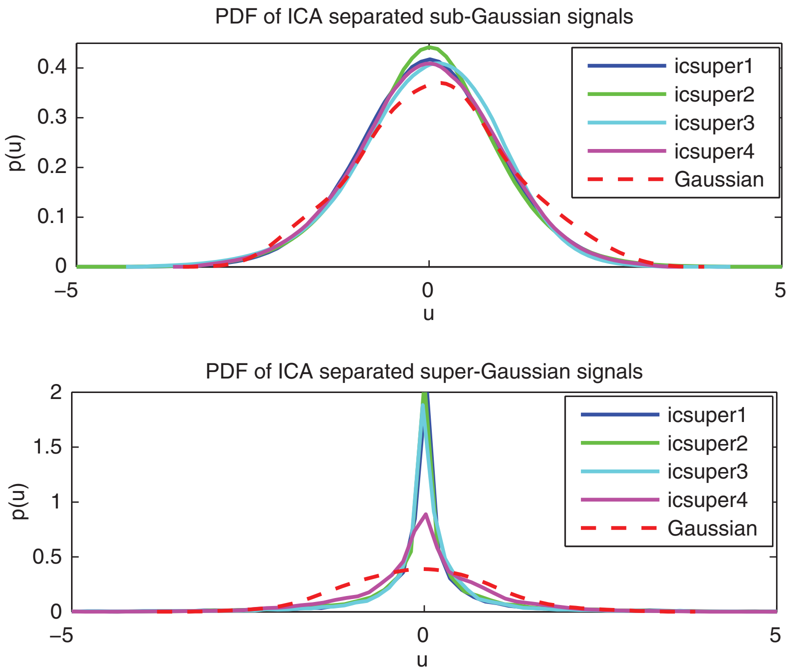

Figure 6. Examples of PDF of ICA separated sub- and super-Gaussian signals.

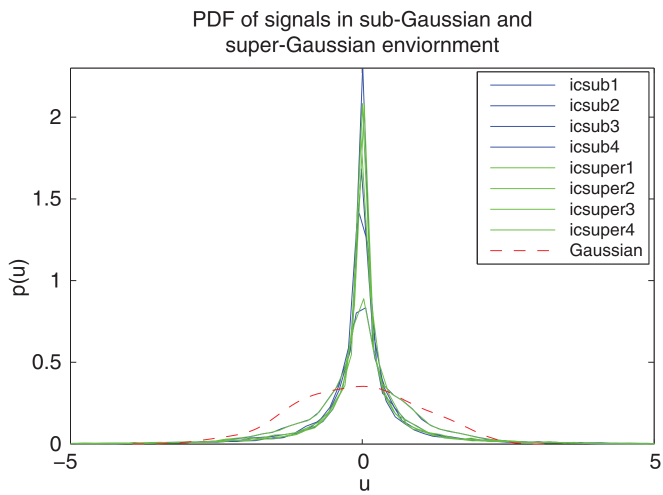

Figure 7. PDF of ICA separated sub- and super-Gaussian signals together.

value demonstrates that after separating different sub- and super-Gaussian audio mixtures, all separated sources remain as super-Gaussian. This explains the conditions where, irrespective of the mixing conditions, all the mixed sources remain as super-Gaussian in nature.

Table 1. Overall kurtosis results for sub-Gaussian and super-Gaussian mixtures.

Sub-Gaussian sources	Super-Gaussian sources	Sub-Gaussian mixing matrix	Super-Gaussian mixing matrix	Kurtosis
X		X		2.0067
X			X	5.0845
	X	X		8.5845
	X		X	12.4867

Table 2. Super-Gaussian signal + super-Gaussian mixing matrix LAR, LLR, WSS and IS.

	WSS estimates		LLR	LAR	IS
SNR in dB	Median	Mean	Mean	Mean	Mean
30	38.09	43.12	0.267	3.45	0.56
25	43.12	49.12	0.398	4.31	1.45
20	51.23	57.81	0.482	5.52	1.71
15	60.12	68.12	0.652	6.56	2.63
10	70.12	79.21	0.786	7.74	3.21
0	93.12	99.67	1.342	8.83	4.32
−5	101.23	105.21	1.867	10.34	6.08
−10	106.23	109.67	2.124	11.82	7.32
−20	110.12	113.08	2.423	13.21	8.56

Table 2 and Figure 8 summarises the objective measures for pure super-Gaussian mixtures across different SNRs ranging from −20 to 30 dB. It can be seen that WSS values are lowest for 30 dB SNR and highest for −20 dB. However there is decreasing value in other three (IS, LAR and LLR) enhancement routines, which provide quality improvement. From the results (Table 2) it is clear that, WSS has proved to be the most useful measure used due to its wider dynamic range as compared to other parameters.

Similarly Tables 3–5 show the result for super-Gaussian source and sub-Gaussian mixing matrix, sub-Gaussian source and super-Gaussian mixing matrix and sub-Gaussian source and sub-Gaussian mixing matrix, respectively. The same results are also plotted in Figures 9–11, respectively. From these results, it is clear that WSS values are showing decreasing trend over the mixtures. As expected, the WSS shows highest values for the pure super-Gaussian mixtures (Table 2 and Figure 8) and the lowest for the pure sub-Gaussian mixtures (Table 5 and Figure 11). The similar trend continues for other parameters (IS, LAR and LLR). However, as compared to WSS, the other three parameters (IS, LAR and LAR) show a small range of changes across different Gaussian mixtures, making the WSS as the most suitable candidate for Gaussian mixture estimation.

The mean and median values of WSS estimates of noisy speech, at different SNR are shown in Figure 12. From the results, it is observed that super + super-Gaussian is having lowest WSS estimate followed by super + sub-Gaussian, sub + super-Gaussian and then highest for sub + sub-Gaussian. This trend is more prominent particularly for low SNRs. The reason could be attributed to the presence of weaker speech units in relatively higher concentration, in the Gaussian mixture with higher WSS estimates compared to others;

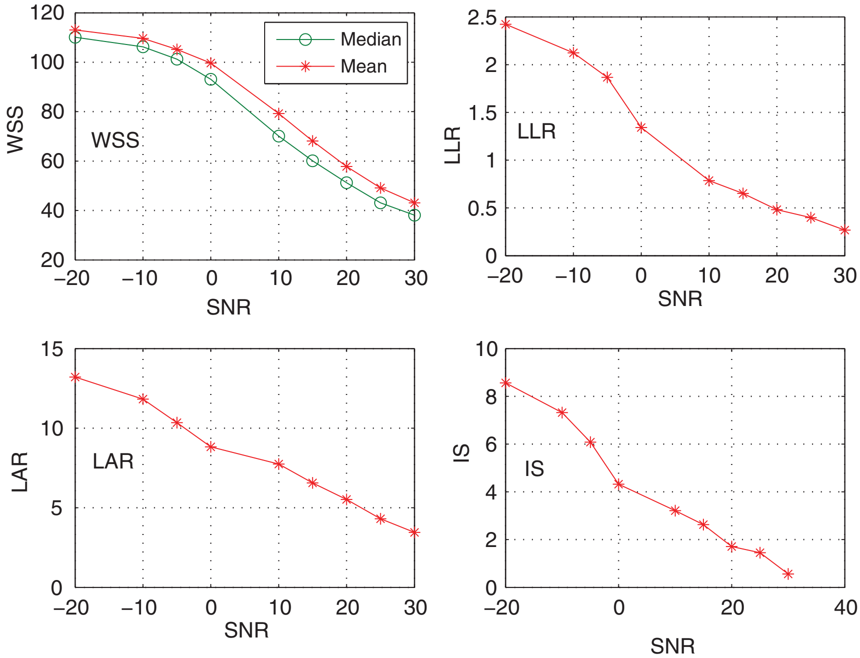

Figure 8. LLR, LAR, IS and WSS estimates versus SNR plots in super-Gaussian source + super-Gaussian mixing matrix.

Table 3. Super-Gaussian signal + sub-Gaussian mixing matrix LAR, LLR, WSS and IS.

	WSS estimates		LLR	LAR	IS
SNR in dB	Median	Mean	Mean	Mean	Mean
30	35.09	42.12	0.397	4.34	0.89
25	43.12	48.12	0.519	5.13	1.95
20	49.31	54.18	0.614	6.45	2.37
15	55.11	61.32	0.721	7.25	3.46
10	62.12	68.21	0.846	8.37	4.21
0	83.42	89.67	1.542	9.45	5.23
−5	90.23	94.21	2.187	11.43	6.98
−10	95.23	100.67	2.424	12.92	8.23
−20	101.12	105.08	2.823	14.56	9.65

as the speech parameters under consideration for them, would undergo higher distortion under the influence of noise.

6. Conclusion

It is concluded that both the super- and sub-Gaussian signals, after separation using ICA are always super-Gaussian in nature. While sub-Gaussian sources can also be separated,

Table 4. Sub-Gaussian signal + super-Gaussian mixing matrix LAR, LLR, WSS and IS.

	WSS estimates		LLR	LAR	IS
SNR in dB	Median	Mean	Mean	Mean	Mean
30	32.09	38.12	0.423	5.34	1.23
25	39.12	45.12	0.621	6.13	2.34
20	46.23	51.81	0.734	7.23	3.21
15	52.12	58.12	0.841	8.23	4.32
10	59.12	66.21	0.986	9.21	5.32
0	83.12	88.67	1.782	10.51	7.84
−5	89.23	94.21	2.418	12.45	8.65
−10	92.23	99.27	2.744	13.83	9.78
−20	100.32	103.48	3.182	15.67	10.97

Table 5. Sub-Gaussian signal + sub-Gaussian mixing matrix LAR, LLR, WSS and IS.

	WSS estimates		LLR	LAR	IS
SNR in dB	Median	Mean	Mean	Mean	Mean
30	32.09	38.12	0.423	5.34	1.23
25	39.12	45.12	0.621	6.13	2.34
20	46.23	51.81	0.734	7.23	3.21
15	52.12	58.12	0.841	8.23	4.32
10	59.12	66.21	0.986	9.21	5.32
0	83.12	88.67	1.782	10.51	7.84
−5	89.23	94.21	2.418	12.45	8.65
−10	92.23	99.27	2.744	13.83	9.78
−20	100.32	103.48	3.182	15.67	10.97

the quality of separation is poorer, and the outcome is always a super-Gaussian approximation of the original signal. Hence, a signal such as polyphonic music would be separated into non-polyphonic music and there would be greater emphasis on aspects of the audio such as voice or other super-Gaussian sources. There is a strong impact of the mixing matrix distribution on the quality of source separation. If the mixing matrix is super-Gaussian in nature, the quality of separation is better than when the mixing matrix is sub-Gaussian.

In this research, evaluation of super- and sub-Gaussian signals are done using LLR, LAR, IS and WSS as objective measures of speech quality. This is performed by computing estimates of these objective measures for noisy speech with white noise for the above Gaussian measures at SNRs −20 to 30 dB. First, these measures are computed for noisy speech with reference to corresponding clear speech and then for the enhanced speech with reference to the corresponding noisy speech. WSS has proved to be the most useful measure used due to its wider dynamic range. The two estimates of WSS do provide a clue to the type of Gaussian measure in use due to differences in its phonetic content. The discrimination provided is highest at lower SNRs. The estimates being lowest for

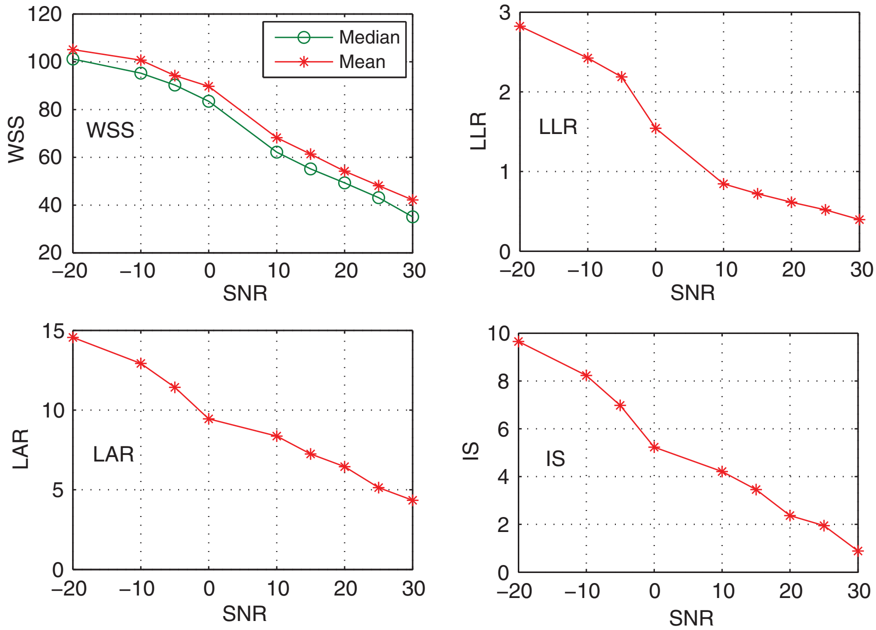

Figure 9. LLR, LAR, IS and WSS estimates versus SNR plots in super-Gaussian source + sub-Gaussian mixing matrix.

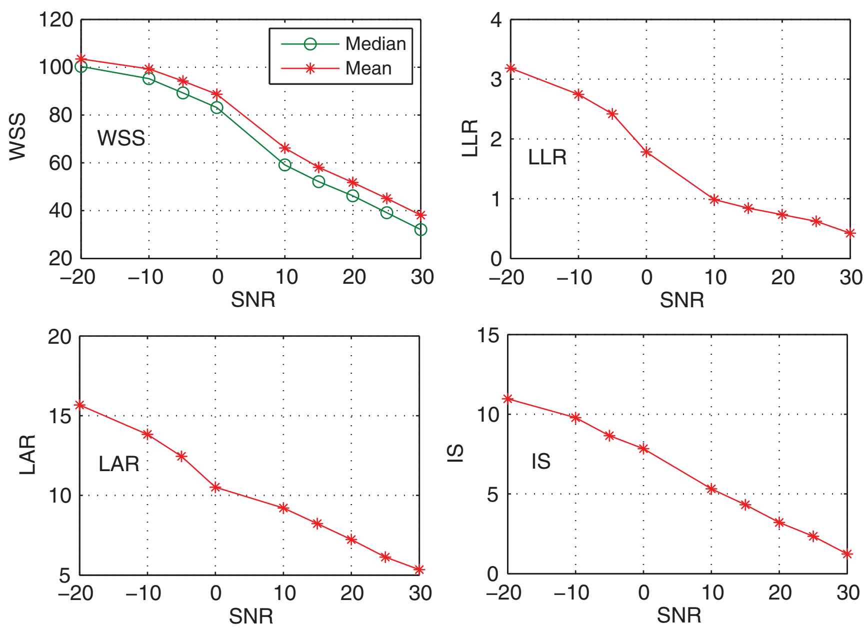

Figure 10. LLR, LAR, IS and WSS estimates versus SNR plots in sub-Gaussian source + super-Gaussian mixing matrix.

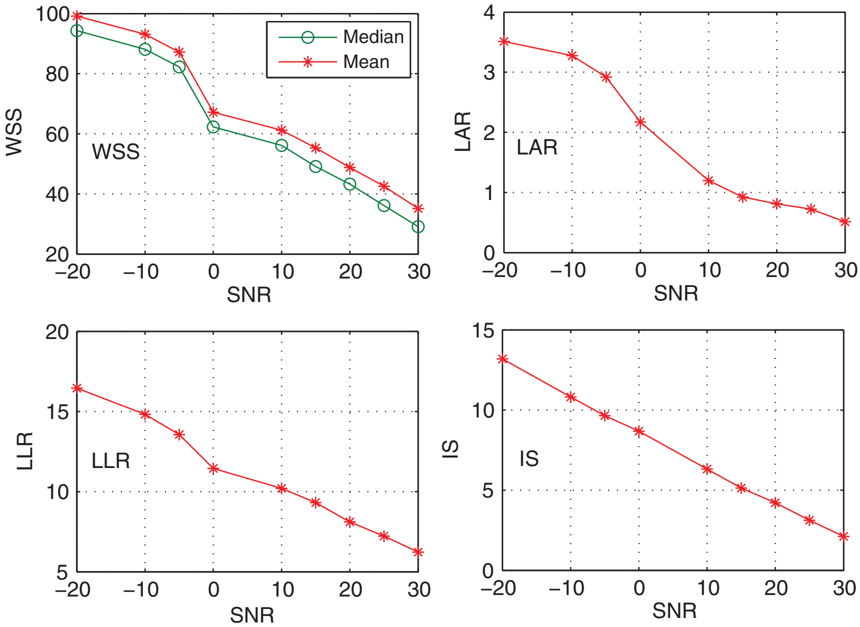

Figure 11. LLR, LAR, IS and WSS estimates versus SNR plots in sub-Gaussian source + sub-Gaussian mixing matrix.

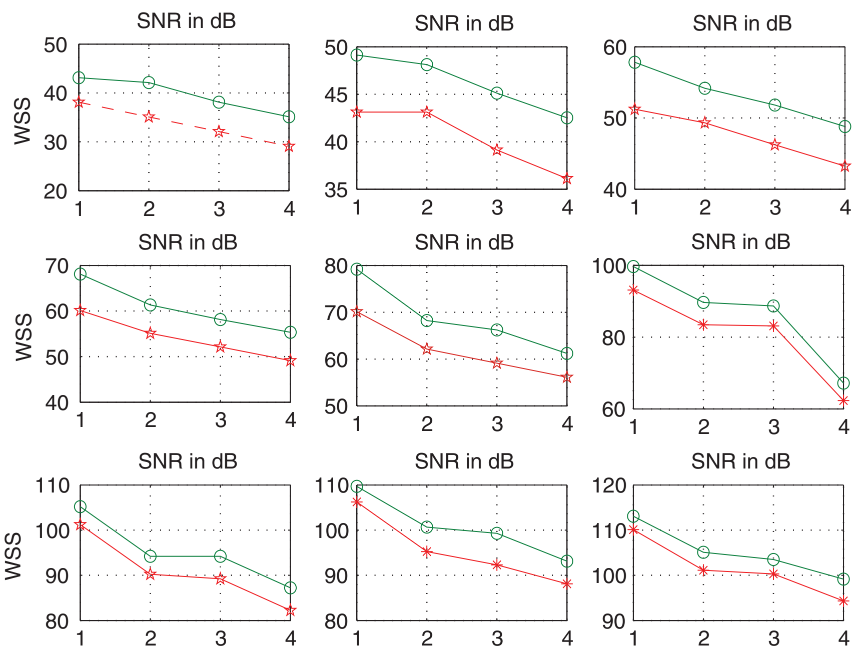

Figure 12. Mean (*) and Median (o) plots of WSS estimates (y axis) for different SNRs in dB, i.e. 30, 25, −20, 15, 10, 0, −5, −10 and −20 (Left to right, top to bottom) for *super source + super mixture, super source + sub mixture, sub source + super mixture and sub source + sub mixture* (denoted in x axis by 1, 2, 3, 4, respectively).

sub-Gaussian mixtures and highest for super-Gaussian mixtures. The reason could be attributed to the existence of weaker speech units in comparatively higher concentration, in the language with higher WSS estimates compared to others; as the speech parameters under consideration for them, would undergo higher distortion under the influence of noise.

References

Bell, A.J., and Sejnowski, T.J. (1997), 'The ''independent Components'' of Natural Scenes are Edge Filters', *Vision Research*, 37, 3327–3338.

Benaroya, L., Bimbot, F., and Gribonval, R. (2006), 'Audio Source Separation with a Single sensor', *IEEE Transactions on, Audio, Speech, and Language Processing*, 14, 191–199.

Blaschke, T., and Wiskott, L. (2003), 'CuBICA: Independent Component Analysis by Simultaneous Third- and Fourth-Order Cumulant Diagonalization', *IEEE Transactions on Signal Processing*, 52, 1250–1256.

Campbell, D., Jones, E., and Glavin, M. (2009), 'Audio Quality Assessment Techniques – A Review, and Recent Developments', *Signal Processing*, 89, 1489–1500.

Cardoso, J. (1998), 'Blind Signal Separation: Statistical Principles', in *Proceedings of the IEEE 86*, 9, 2009–2025.

Cichocki, A., and Amari, S.I. (2002), *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, New York, NY, USA: John Wiley & Sons, Inc.

Cristescu, R., Ristaniemi, T., Joutsensalo, J. and Karhunen, J. (2000), 'CDMA Delay Estimation Using Fast ICA Algorithm', in, *The 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications 2000, PIMRC 2000*, vol. 2, pp.1117–1120.

De Martino, F., Gentile, F., Esposito, F., Balsi, M., Di Salle, F., Goebel, R., and Formisano, E. (2007), 'Classification of fMRI Independent Components Using IC-fingerprints and Support Vector Machine Classifiers', *NeuroImage*, 34, 177–194.

Dubnov, S., Tabrikian, J., and Targan, M.A. (2006), 'Speech Source Separation in Convolutive Environments Using Space-time-frequency Analysis', *EURASIP Journal of Advances in Signal Processing*, 2006, 172, 2006.

Eriksson, J., and Koivunen, V. (2004), 'Identifiability, Separability, and Uniqueness of linear ICA Models', *IEEE, Signal Processing Letters*, 11, 601–604.

Foote, J. (1999), 'An Overview of Audio Information Retrieval', *Multimedia Systems*, 7, 2–10.

Hansen, J.H.L., and Pellom, B.L. (1998). 'An Effective Quality Evaluation Protocol for Speech Enhancement Algorithms', in, *Proceedings of International Conference on Spoken Language Processing (ICSLP)*, Sydney, Australia, pp. 2819–2822.

He, T., Clifford, G., and Tarassenko, L. (2006), 'Application of Independent Component Analysis in Removing Artefacts from the Electrocardiogram', *Neural Computing & Applications*, 15, 105–116.

Hyvärinen, A. (1999), 'Fast and Robust Fixed-Point Algorithms for Independent Component Analysis', *IEEE Transactions on Neural Networks*, 10, 626–634.

Hyvarinen, A., Karhunen, J., and Oja, E. (2001), *Independent Component Analysis*, New York, NY, USA: Wiley Interscience.

Jang, G.J., and Lee, T.W. (2003), 'A Maximum Likelihood Approach to Singlechannel Source Separation', *Journal of Machine Learning Research*, 4, 1365–1392.

Klatt, D. (1982), 'Text-to-Speech Conversion System', in, *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Cambridge, MA, pp. 1589–1592.

Lee, T.W. (1998), *Independent Component Analysis: Theory and Applications*, MA, USA: Kluwer Academic Publishers.

Lee, T.W., Girolami, M., Bell, A.J., and Sejnowski, T.J. (2000), 'A Unifying Information-theoretic Framework for Independent Component Analysis', *Computers & Mathematics with Applications*, 39, 1–21.

Lee, T.W., Lewicki, M.S., and Sejnowski, T.J. (1999), 'Unsupervised Classification with Non-Gaussian Mixture Models Using ICA', in *Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems*, Cambridge, MA, USA: MIT Press, pp. 508–514.

Morita, S., and Nanri, Y. (2006), 'Sound Source Separation of Trio using Stereo Musig Sound Signal Based on Independent Component Analysis', in *2006 IEEE International Conference on, Multimedia and Expo*, pp. 185–188.

Pham, D., Garrat, P., and Jutten, C. (1992), 'Separation of a Mixture of Independent Sources Through a Maximum Likelihood Approach', in *European Signal Processing Conference*, pp. 771–774.

Stone, J.V. (2002), 'Independent Component Analysis: an Introduction', *Trends in Cognitive Sciences*, 6, 59–64.

Vincent, E., Gribonval, R., and Fevotte, C. (2006), 'Performance Measurement in Blind audio Source Separation', *IEEE Transactions on, Audio, Speech, and Language Processing*, 14, 1462–1469.

Wilson, K. (2007), 'Speech Source Separation by Combining Localization Cues with Mixture Models of Speech Spectra', in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 33–36.

Zarzoso, V., Murillo-Fuentes, J.J., Boloix-Tortosa, R., and Nandi, A.K. (2006), 'Optimal Pairwise Fourth-order Independent Component Analysis', *IEEE Transactions on, Signal Processing*, 54, 3049–3063.

Zarzoso, V., and Nandi, A.K. (2002), 'Blind Separation of Independent Sources for Virtually Any Source Probability Density Function', *IEEE Transactions on, Signal Processing*, 47, 2419–2432.