Source Dependency Modelling in Frequency Domain Source Separation

Jonathon Chambers
Director, LSSC Consortium

Advanced Signal Processing Group
Loughborough University

Engineering and Physical Sciences Research Council

Acknowledgments

QinetiQ

THALES

Selex ES
A Finmeccanica Company

Texas Instruments

PrismTech

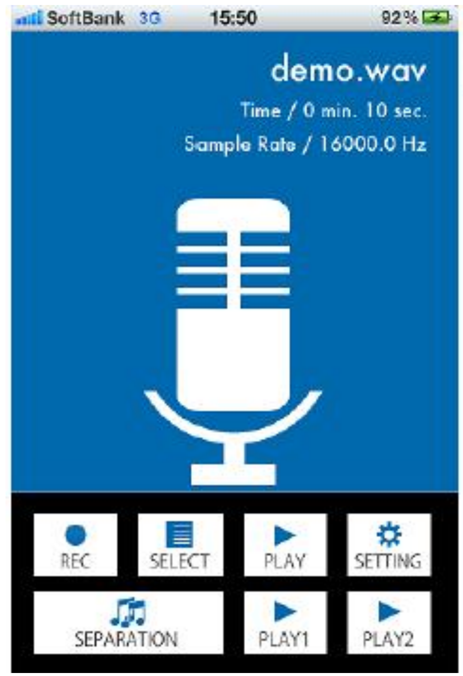
steepest ascent

And, I wish to acknowledge Yanfeng Liang, my PhD researcher, who executed this research.

Loughborough University

Independent Vector Analysis

- IVA application on iPhone



[1] N. Ono, "Blind source separation on iPhone in real environment," in EUSIPCO 2013, Marrakech, Morocco, 2013.

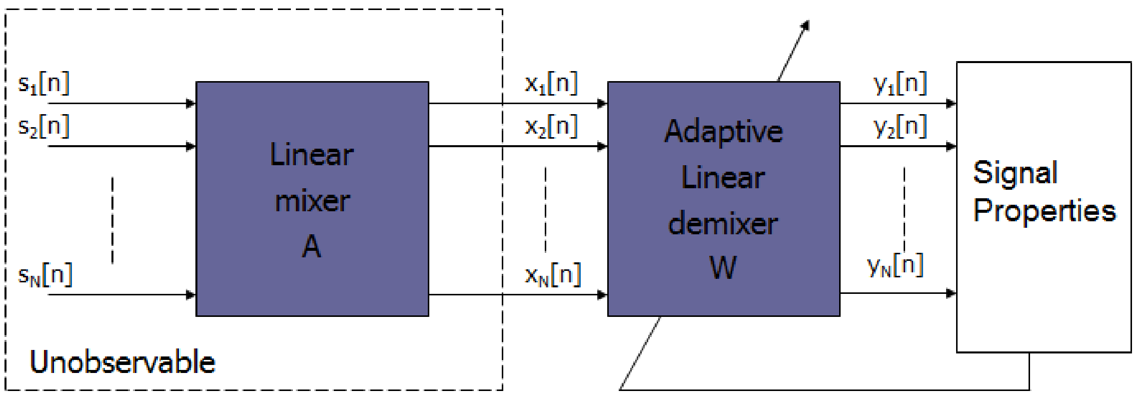
Outline

1. Blind Source Separation (BSS)
2. Independent Vector Analysis (IVA)
3. Copula Based Dependency Model
4. IVA with Multivariate Student’s t Source Prior
5. Experimental Results
6. Conclusions

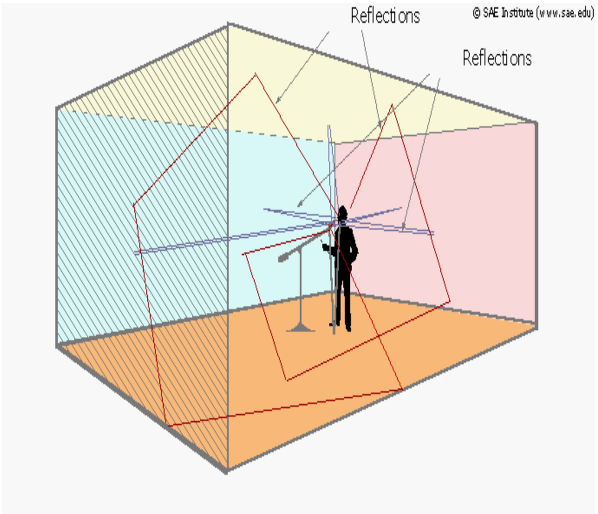
Blind Source Separation

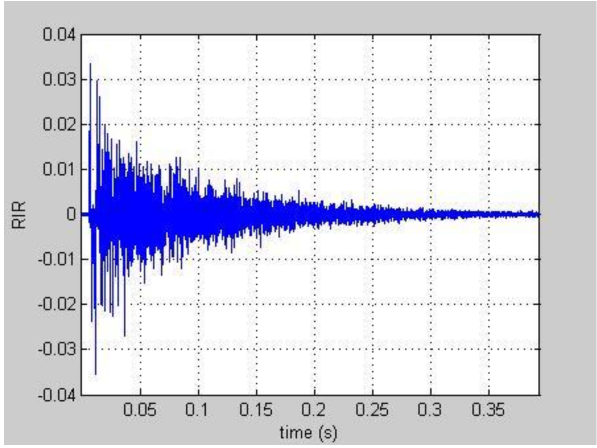
"One of our most important faculties is our ability to listen to, and follow, one speaker in the presence of others. This is such a common experience that we may call it 'the cocktail party problem.' No machine has been constructed to do just this, to filter out one conversation from a number jumbled together…".

Colin Cherry
1914-1979



Blind Source Separation

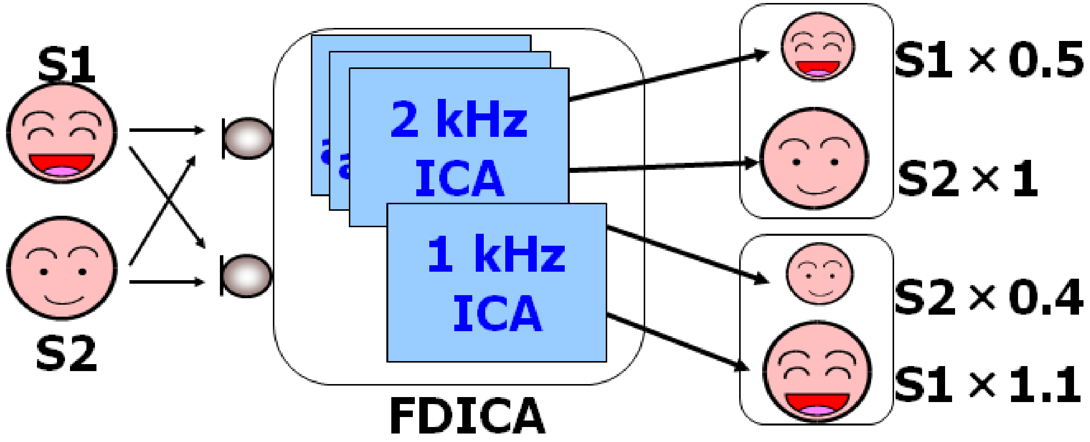




Independent component analysis (ICA) is the central tool for BSS, which is a higher order statistic method.

ICA has two ambiguities

1. The scaling ambiguity
2. The permutation ambiguity

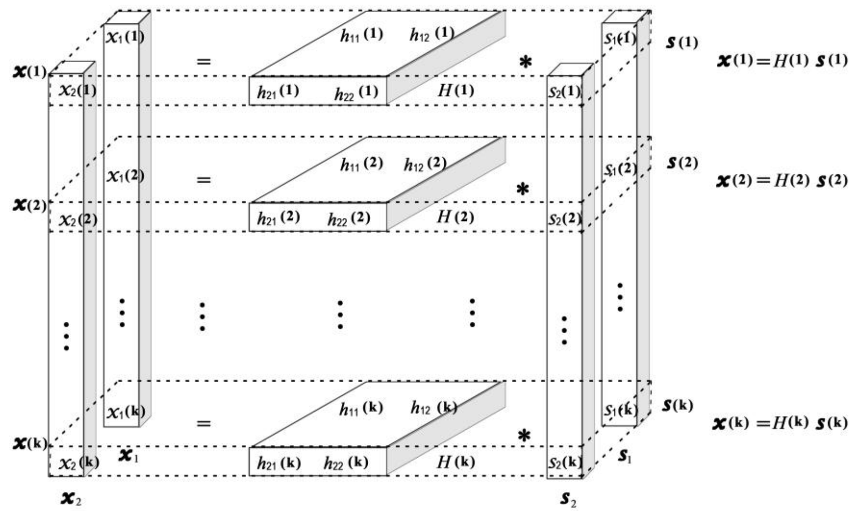


Loughborough University

Independent Vector Analysis

The cost function for IVA is the Kullback-Leibler divergence between the joint probability density function and the product of marginal probability density functions of the individual source vectors [1].

$$
\begin{aligned}
J &= \mathcal{KL}\left(p(\hat{\mathrm{s}}_1 \ldots \hat{\mathrm{s}}_N) || \prod p(\hat{\mathrm{s}}_i)\right) \\
&= \int p(\hat{\mathrm{s}}_1 \cdots \hat{\mathrm{s}}_N) \log \frac{p(\hat{\mathrm{s}}_1 \cdots \hat{\mathrm{s}}_N)}{\prod q(\hat{\mathrm{s}})} d\hat{\mathrm{s}}_1 \cdots d\hat{\mathrm{s}}_N \\
&= const - \sum_{k=1}^{K} \log |det(W(k))| - \sum_{i=1}^{N} E[\log p(\hat{\mathrm{s}}_i)]
\end{aligned}
$$



[2] T. Kim, H. Attias, S. Lee, and T. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Transactions on Audio, Speech and Language processing*, vol. 15, pp. 70–79, 2007.

Independent Vector Analysis

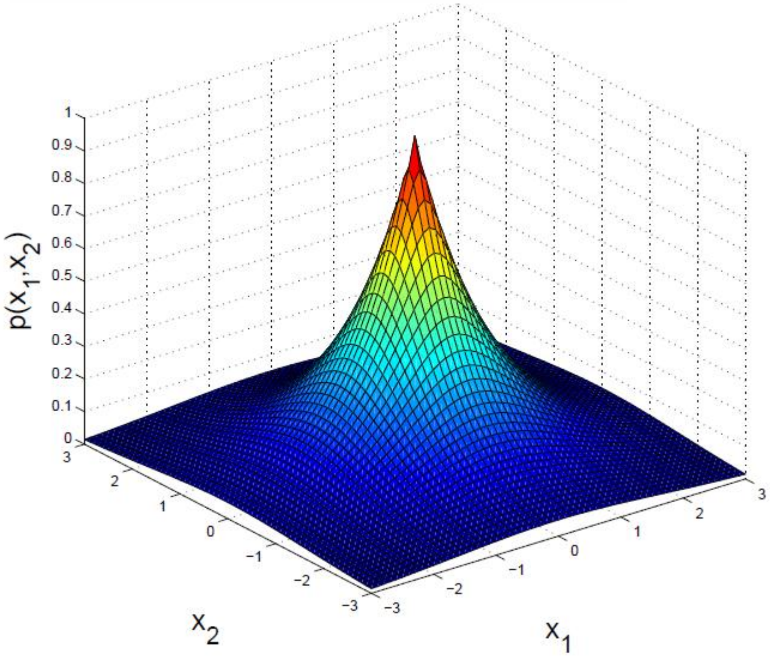
The natural gradient IVA:

$$W(k)_{new} = W(k)_{old} + \eta \Delta W(k)$$

$$\Delta W(k) = \left(\mathrm{I} - E[\varphi^{(k)}(\hat{s})\hat{s}^*(k)]\right)W(k)$$

$$\varphi^{(k)}(\hat{s}) = [\varphi^{(k)}(\hat{s}_1), \cdots, \varphi^{(k)}(\hat{s}_N)]^T$$

$$\varphi^{(k)}(\hat{\mathbf{s}}_i) = -\frac{\partial \log p(\hat{\mathbf{s}}_i)}{\partial \hat{s}_i(k)}$$



Original source prior for IVA: $$q(\mathbf{s}_i) \propto exp\left(-\sqrt{(\mathbf{s}_i - \boldsymbol{\mu}_i)^\dagger \Sigma_i^{-1} (\mathbf{s}_i - \boldsymbol{\mu}_i)}\right)$$

Original score function for IVA: $$\varphi^{(k)}(\hat{\mathbf{s}}_i) = -\frac{\partial log q(\hat{s}_i^{(1)} \cdots \hat{s}_i^{(k)})}{\partial \hat{s}_i^{(k)}} = \frac{\hat{s}_i^{(k)} / (\sigma_i^{(k)})^2}{\sqrt{\sum_{k=1}^{K} \left| \frac{\hat{s}_i^{(k)}}{\sigma_i^{(k)}} \right|^2}}$$

Independent Vector Analysis

- Fast fixed point IVA

The cost function:

$$J_{FastIVA} = \sum_{i=1}^{N} \Big(E[F(\sum_{k=1}^{K} |\hat{s}_i(k)|^2)] - \sum_{k=1}^{K} \lambda_i(k)(\mathbf{w}_i(k)^\dagger \mathbf{w}_i(k) - 1) \Big)$$

The updates rules:

$$\begin{aligned} \mathbf{w}_i(k) \leftarrow & E[F'(\sum_{k'=1}^{K} |\hat{s}_{i,o}(k')|^2) + |\hat{s}_{i,o}(k)|^2 F''(\sum_{k'=1}^{K} |\hat{s}_i(k')|^2))]\mathbf{w}_i(k) \\ & - E[(\hat{s}_{i,o}(k))^* F'(\sum_{k'=1}^{K} |\hat{s}_{i,o}(k')|^2)\mathbf{x}(k)] \end{aligned}$$

$$W(k) \leftarrow (W(k)(W(k))^\dagger)^{-1/2} W(k).$$

- Auxiliary function based IVA

Auxiliary function technique

$$\bar{\Theta} = argmin_{\Theta} J(\Theta)$$

$$J(\Theta) = min_{\tilde{\Theta}} Q(\Theta, \tilde{\Theta})$$

$$\tilde{\Theta}(i+1) = argmin_{\tilde{\Theta}} Q(\Theta(i), \tilde{\Theta})$$

$$\Theta(i+1) = argmin_{\Theta} Q(\Theta, \tilde{\Theta}(i+1))$$

The update rules:

$$r_i = \sqrt{\textstyle\sum_{k=1}^{K} |\mathbf{w}_i^\dagger(k)\mathbf{x}(k)|^2}$$

$$V_i(k) = E[\frac{g'_R(r_i)}{r_i}\mathbf{x}(k)\mathbf{x}(k)^\dagger]$$

$$\mathbf{w}_i(k) = (W(k)V_i(k))^{-1}\mathbf{e}_i$$

$$\mathbf{w}_i(k) = \frac{\mathbf{w}_i(k)}{\sqrt{\mathbf{w}_i^\dagger(k)V_i(k)\mathbf{w}_i(k)}}$$

Copula Based Dependency Model

- Sklar's Theorem:

$$F(z_1, \cdots, z_d) = C(F_1(z_1), \cdots, F_d(z_d))$$

- Copula density function:

$$c(\boldsymbol{u}) = \frac{\partial^d C(u_1, \cdots, u_d)}{\partial u_1 \cdots u_d}$$

- The joint probability density function:

$$\begin{aligned} p(z_1, \cdots, z_d) &= \frac{\partial^d F(z_1, \cdots, z_d)}{\partial z_1 \cdots z_d} \\ &= \frac{\partial^d C(F_1, \cdots, F_d)}{\partial F_1 \cdots F_d} \frac{\partial F_1}{\partial z_1} \cdots \frac{\partial F_d}{\partial z_d} \\ &= c(F_1, \cdots, F_d) \prod_{i=1}^{d} p_i(z_i) \end{aligned}$$

Copula Based Dependency Model

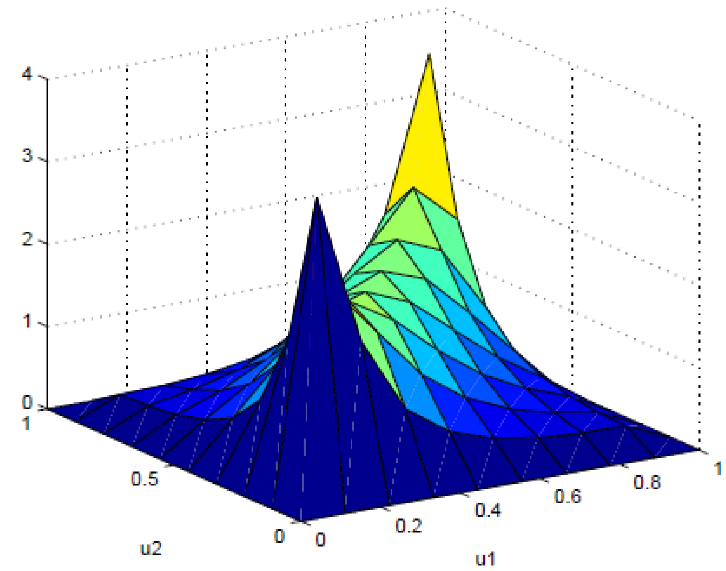
- T copula

$$C(\boldsymbol{u}) = \int_{-\infty}^{F_1^{-1}} \cdots \int_{-\infty}^{F_d^{-1}} \frac{\Gamma(\frac{v+d}{2})}{\Gamma(\frac{v}{2})(\sqrt{\pi v |\Sigma|})}(1 + \frac{\boldsymbol{z}^{\dagger}\Sigma^{-1}\boldsymbol{z}}{v})^{-\frac{v+d}{2}} dz$$

- T copula density function

$$c(u_1, \cdots, u_d) = \frac{\Gamma(\frac{v+d}{2})\Gamma(\frac{v}{2})^{d-1}}{|\Sigma|^{\frac{1}{2}}\Gamma(\frac{v+1}{2})^d} \frac{\prod_{i=1}^{d}(1 + \frac{|z_i|^2}{v})^{\frac{v+1}{2}}}{(1 + \frac{\boldsymbol{z}^{\dagger}\Sigma^{-1}\boldsymbol{z}}{v})^{\frac{v+d}{2}}}$$

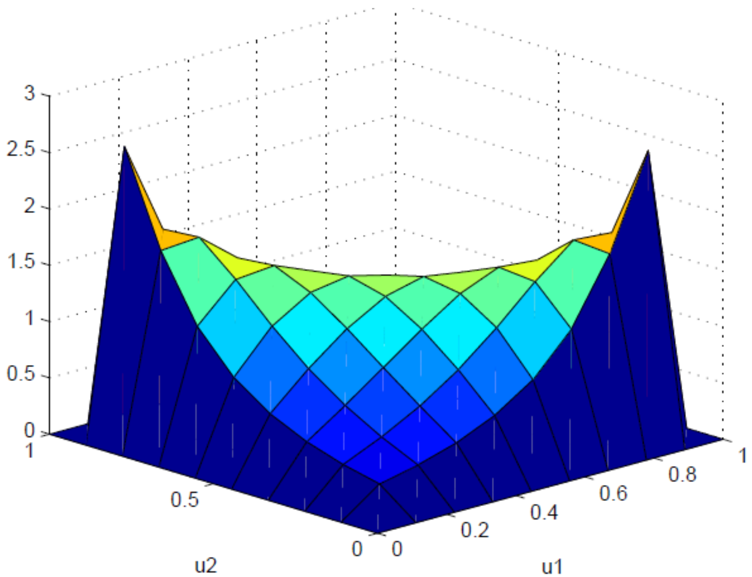
- T copula density function with four degrees freedom and 0.7 correlation coefficient.



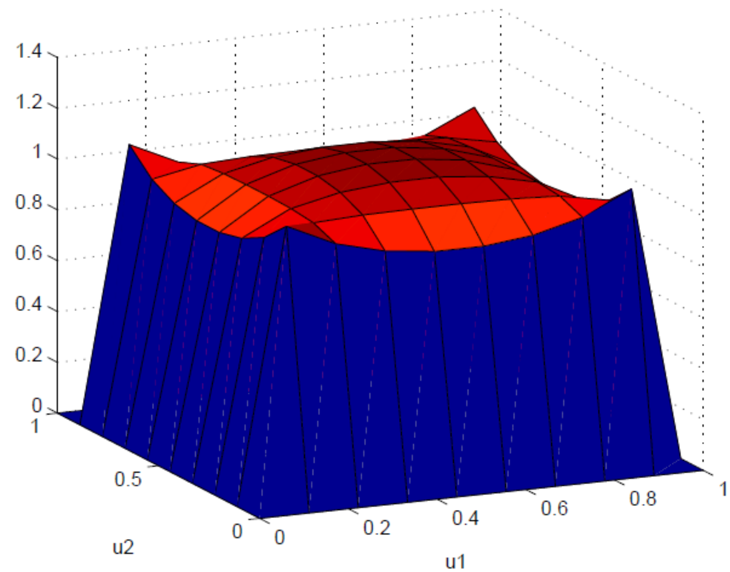
Loughborough University

Copula Based Dependency Model

- T copula density function with four degrees freedom and -0.6 correlation coefficient.



- T copula density function with four degrees freedom and zero correlation coefficient.



Copula Based Dependency Model

- The calculations of Chi and lamda for Chi plot [2]

$$\tilde{H}_i = \sum_{j \neq i} I(x_j \leq x_i, y_j \leq y_i)/(n-1)$$

$$\tilde{F}_i = \sum_{j \neq i} I(x_j \leq x_i)/(n-1)$$

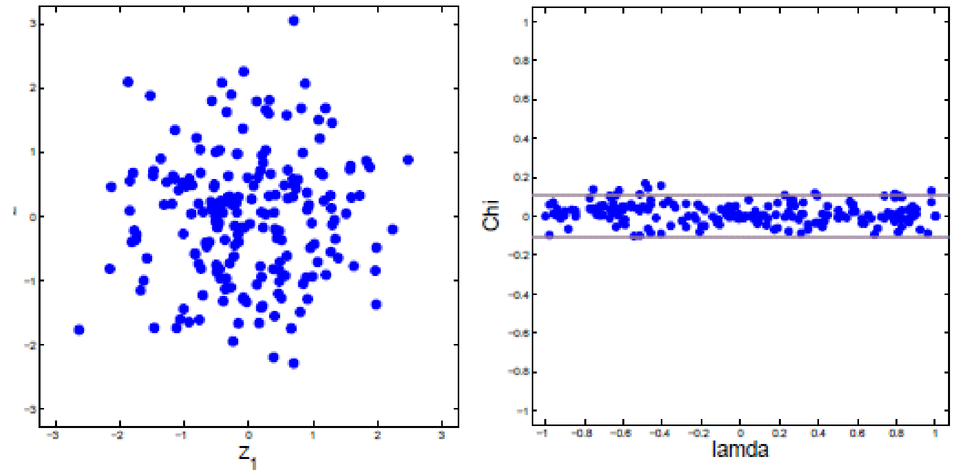
$$\tilde{G}_i = \sum_{j \neq i} I(y_j \leq y_i)/(n-1)$$

$$\tilde{S}_i = sign[(\tilde{F}_i - 0.5)(\tilde{G}_i - 0.5)]$$

$$Chi_i = (\tilde{H}_i - \tilde{F}_i\tilde{G}_i)/\sqrt{\tilde{F}_i(1-\tilde{F}_i)\tilde{G}_i(1-\tilde{G}_i)}$$

$$lamda_i = 4S_i max((\tilde{F}_i - 0.5)^2, (\tilde{G}_i - 0.5)^2)$$

- Scatter plot and Chi plot of two independent random variables

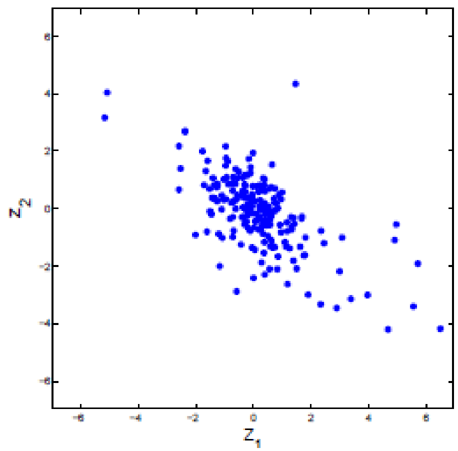


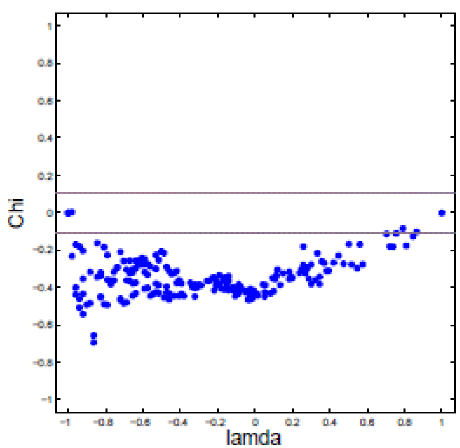
In Chi plot, if all the points are between the two lines, it means independent, otherwise dependent.

[3] N. I. Fisher and P. Switzer, "Chi-plots for assessing dependence," *Biometrika*, vol. 72, pp. 253–265, 1985.

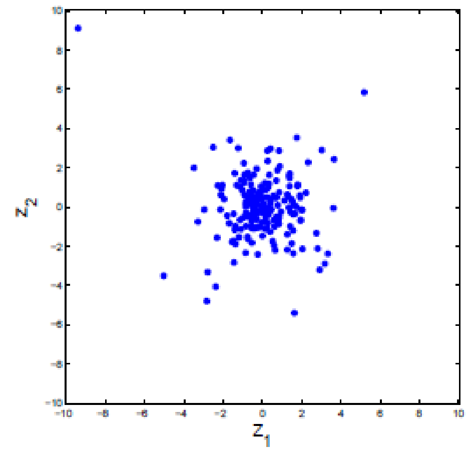
Copula Based Dependency Model

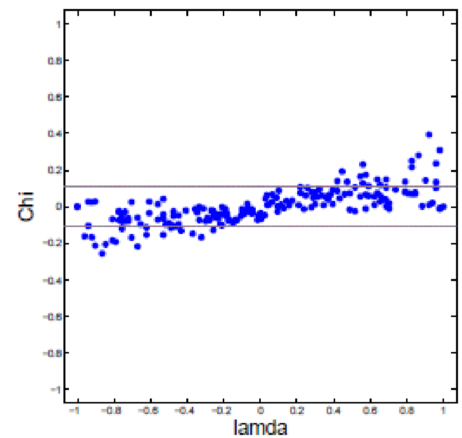
Two correlated variables with a t copula





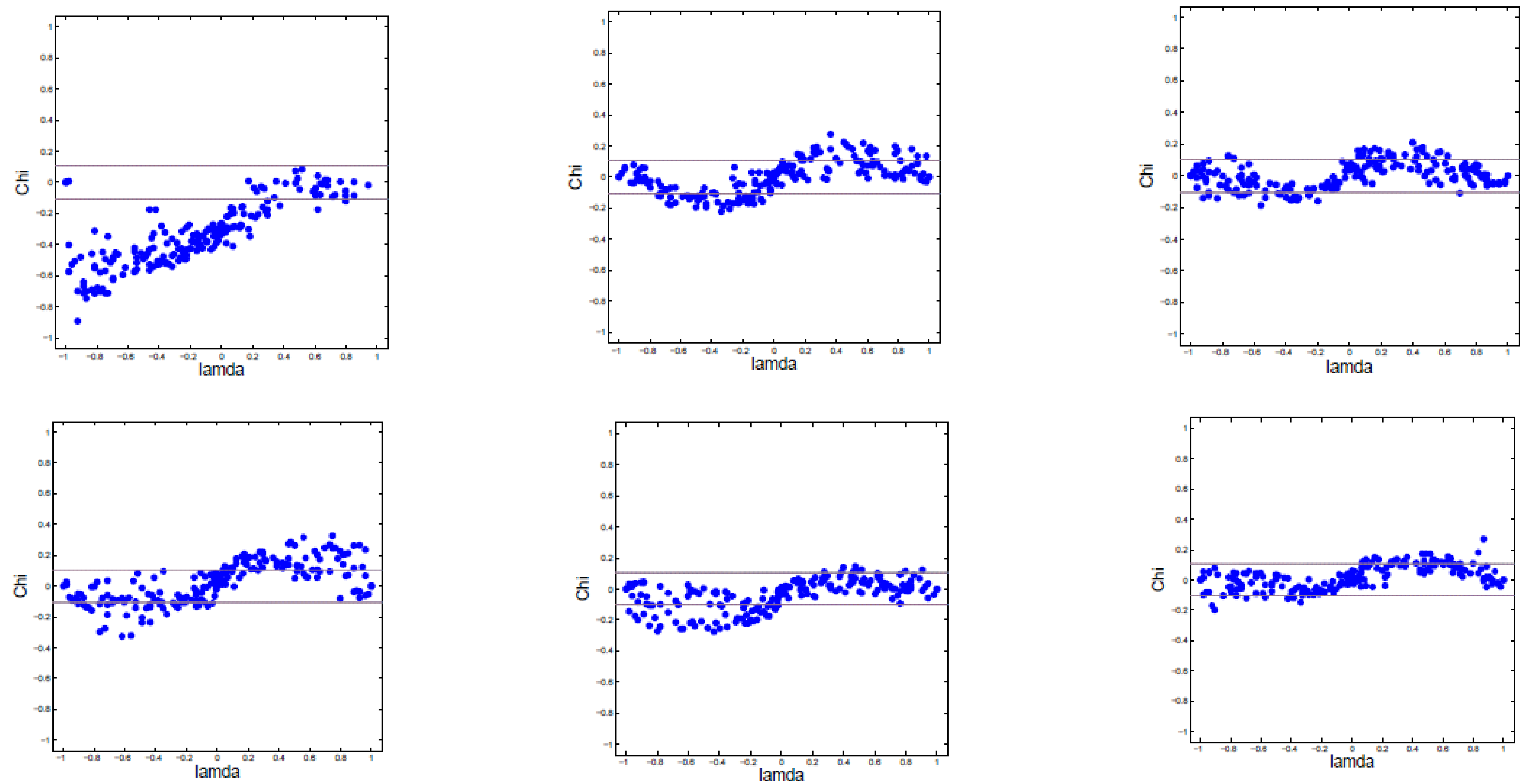
Two uncorrelated variables with a t copula





Copula Based Dependency Model

The Chi plots of two frequency bins of a speech signal, i.e. 50th and 51th frequency bins, 50th and 55th frequency bins, 50th and 60th frequency bins, 50th and 100th frequency bins, 50th and 200th frequency bins, 50th and 500th frequency bins.



IVA with Multivariate Student’s t Source Prior

- Assuming the marginal distribution for the t copula is a student’s t distribution:

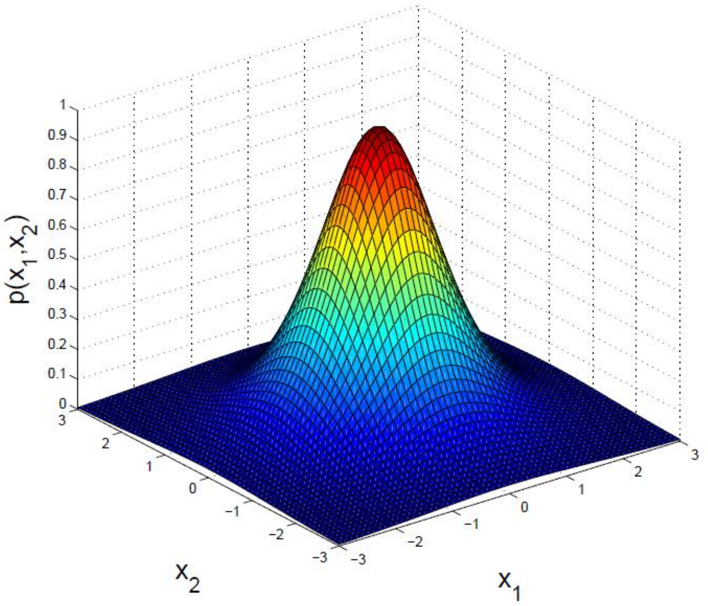
$$p(s_i(k)) = \frac{\Gamma(\frac{v+K}{2})}{\sqrt{v\pi}\Gamma(\frac{v}{2})}(1 + \frac{|s_i(k)|^2}{v})^{-\frac{v+1}{2}}$$

- The multivariate student’s t source prior:

$$p(\mathbf{s}_i) \propto (1 + \frac{\sum_{k=1}^{K} |s_i(k)|^2}{v})^{-\frac{v+K}{2}}$$

- The new nonlinear score function:

$$\varphi^{(k)}(\hat{s}_i(1)\cdots\hat{s}_i(k)) = \frac{v+K}{v}\frac{\hat{s}_i(k)}{1+\frac{1}{v}\sum|\hat{s}_i(k)|^2}$$



Loughborough University

Experimental Results

Different speech signals are selected from the TIMIT dataset, and convolved into mixtures. The reverberation time RT60 is set to be 200ms.

TABLE I

SEPARATION PERFORMANCE COMPARISON IN SDR

mixtures	original(dB)	proposed(dB)	improvement(dB)
mixture1	18.81	20.12	1.31
mixture2	15.94	17.26	1.32
mixture3	9.97	11.73	1.76
mixture4	11.68	12.40	0.72
mixture5	18.80	19.91	1.11
mixture6	12.27	18.74	6.47
mixture7	8.88	11.10	2.22
mixture8	15.57	17.09	1.52
mixture9	18.10	19.50	1.4
mixture10	16.84	19.65	2.81

TABLE II

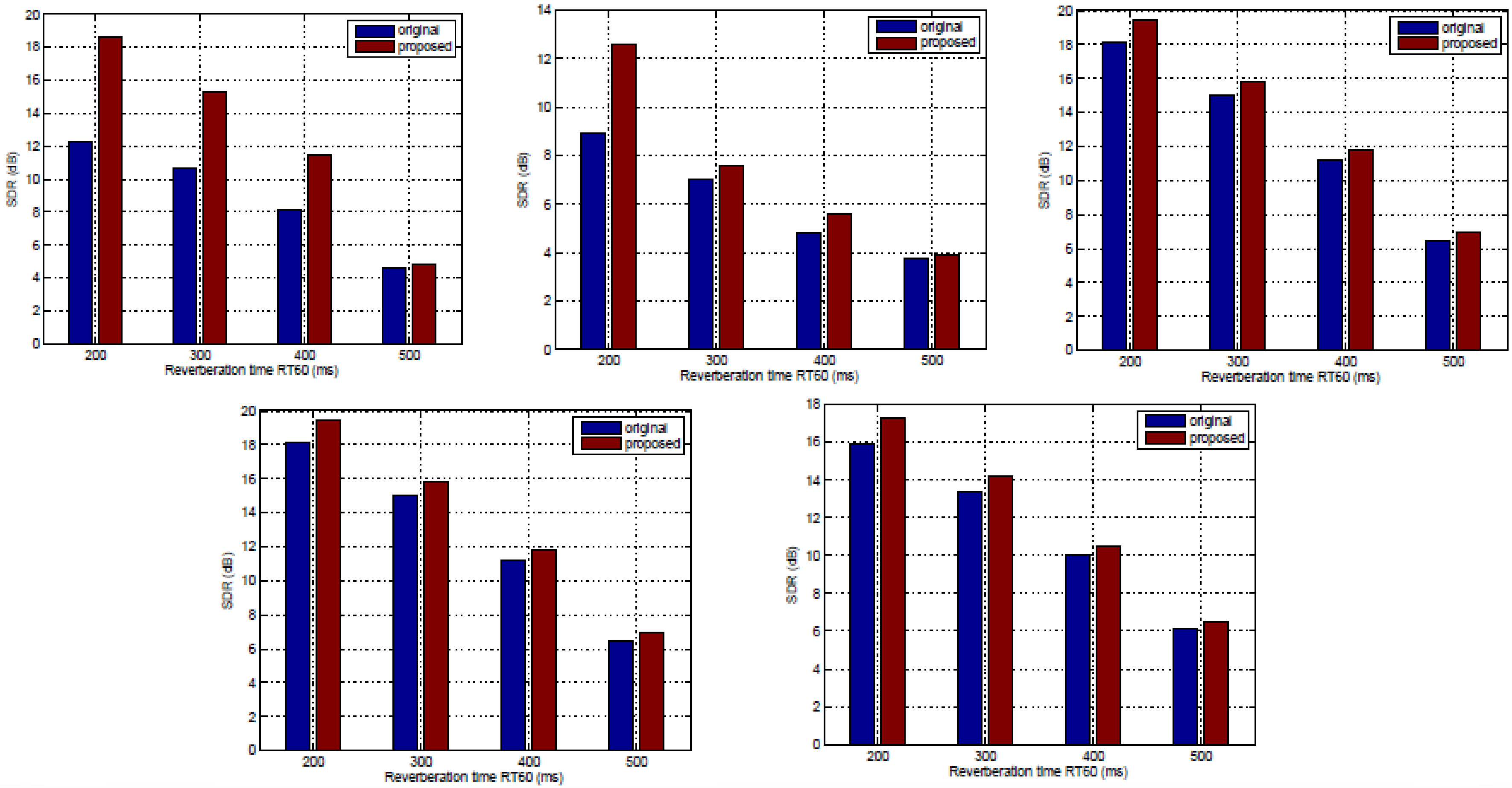
SEPARATION PERFORMANCE COMPARISON IN SIR

mixtures	original(dB)	proposed(dB)	improvement(dB)
mixture1	20.30	21.43	1.13
mixture2	17.88	19.00	1.12
mixture3	12.08	12.77	0.69
mixture4	14.42	14.97	0.55
mixture5	20.28	20.95	0.67
mixture6	14.08	20.94	6.86
mixture7	10.72	12.57	1.85
mixture8	16.98	18.77	1.79
mixture9	20.14	20.80	0.66
mixture10	19.53	21.54	2.01

50 different speech mixtures are tested, and the average SDR and SIR improvements are 1.3 dB and 1.1 dB respectively.

Experimental Results

- Performance in different room environments



Loughborough University

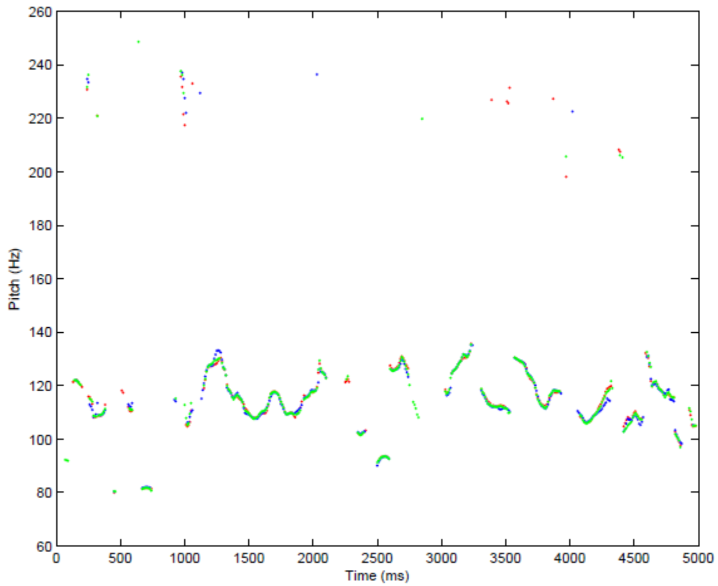
Experimental Results

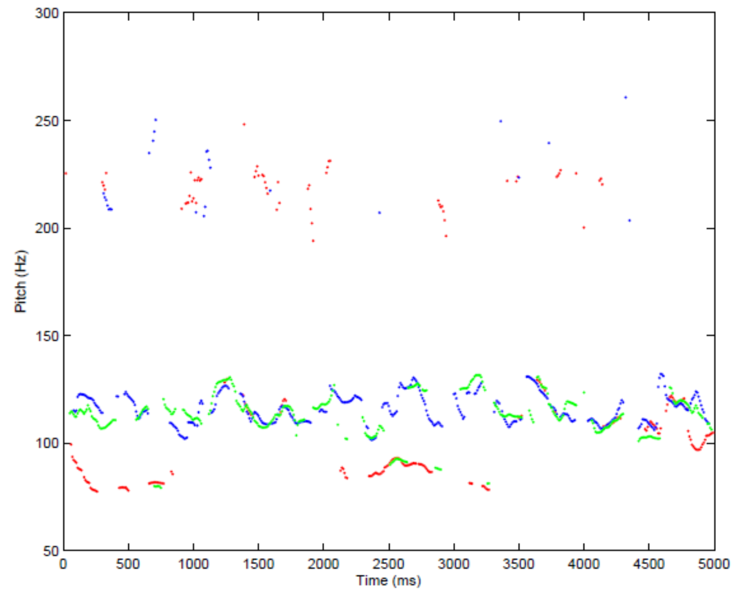
- Experiments by using real room recordings.

TABLE III

SEPARATION RATE COMPARISON WHEN USING REAL ROOM RECORDINGS

	mixtures	original	proposed
separation rate	0.0379	0.2515	0.2794





Conclusions

- The dependency structure within the frequency domain speech signals was exploited by introducing copula theory.
- The t copula was found suitable to model the inter-frequency dependency, which was also confirmed by observing the Chi-plot between two frequency bins of a real speech signal.
- A multivariate student's t distribution was constructed by using the t copula density function and univariate student's t marginal distribution, which was adopted as the new source prior for the NG-IVA algorithm.
- The separation performance was tested in different reverberant room environments and also by using real room recordings. The average SDR and SIR improvement are approximately 1.3dB and 1.1dB respectively.

References

[1] N. Ono, “Blind source separation on iPhone in real environment”, EUSIPCO 2013, Marrakech, Morocco, 2013.

[2] T. Kim, H. Attias, S. Lee and T. Lee, “Blind source separation exploiting higher-order frequency dependencies”, IEEE Transactions on Audio, Speech and Language processing, vol. 15, pp. 70-79, 2007.

[3] N. I. Fisher and P. Switzer, “Chi-plots for assessing dependence”, Biometrika, vol. 72, pp. 253-265, 1985.

[4] I. Lee, T. Kim and T.-W. Lee, “Fast fixed-point independent vector analysis algorithms for convolutive blind source separation”, Signal Processing, vol. 87, pp. 1859-1871, 2007

[5] S. Demarta and A. J. McNeil, “The t copula and related copulas”, International Statistical Review, vol. 73,pp. 111-129, 2005

[6] **Y. Liang**, G. Chen, S. M. Naqvi and **J. A. Chambers**, “Independent vector analysis with a multivariate student’s t distribution source prior for speech separation”, Electronics Letters, vol. 49, pp. 1-2, 2013

[7] **Y. Liang**, G. Chen, S. M. Naqvi and **J. A. Chambers**, “Copula based independent vector analysis with multivariate student’s t source prior for frequency domain blind source separation”, in revision for IEEE Transactions on Audio, Speech and Language Processing.

Thanks for your attention!

Questions?