Source Separation using Independent Component Analysis

Robert Siwerz
Royal Institute of Technology
Stockholm
rsiwerz@kth.se

Christopher Dahlén
Royal Institute of Technology
Stockholm
cdahlen@kth.se

Abstract

Blind source separation (BSS) is a well researched area and multiple techniques, such as Deep Neural Network, Beamforming and PCA, has shown promising result in trying to separate mixed signals. This paper used the Fast-ICA technique which is related to the PCA but further enhances the independence between the signals. Three different verification metrics were used in order to assess the performance of the technique. By using visual inspection, in order to verify the shape similarities, combined with the Euclidean distance and an auditive verification the results show that Fast-ICA can be used for BSS.

1 Introduction

Blind source separation (BSS) is a common problem and can be exemplified by the *cocktail party problem* where one wishes to separate a speaker in a group of people, that are talking to each other, by using multiple microphones. A well working solution to the cocktail party problem is of importance in many situation where one want to record isolated sounds in a noisy environment. In recent years, Deep Neural Networks (DNN) is accomplishing this by classifying different source spectra [GSE14]. Combined with an Expectation Maximization (EM) algorithm, the DNN has shown promising results in single channel source separation and more recently multi-channel source separation [NLV16].

Other approaches, such as Beamforming has also shown promising result. The idea of Beamforming is to filter the mixed signals, followed by combining and extracting the searched signal and to reject the others [Ade+12]. However, this approach relies on previous information about the microphones, noise signals and the way the source signals are mixed. Further, by design the Beamforming algorithm lies on the assumption that the microphones are omnidirectional and may cause limitation in some experiments [11].

Another statistical approach is the Principal Component Analysis (PCA) which has shown to be a candidante for BSS [WK09]. In particular, PCA was shown to be a useful approach to the case of separation "machine rotating signals" which separates signals that are corrupted with noise [SF05] Although not directly in speech domain, the theory behind it is similar.

A related approach to PCA is the Independent Component Analysis (ICA) in which independence between signals are assumed. This assumption has resulted in a greater performance of

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

finding the true sources, rather than the PCA approach [WK09]. ICA has shown good results in numerous domains such as BSS, feature extraction of digital signals and electrical recordings of brain activity [HO00]. With these techniques in mind, this report aims to investigate how to separate a set of source signals from a set of mixed signals using ICA. Instead of relying on training data, this is an unsupervised approach and does not rely on any previous knowledge. In particlar Fast-ICA invented by Hyvärinen [HO00] is a fast and well documented method for BSS.[HO00][ADD11]

2 Background

This section contains a brief explanation of the ICA method and how it can be applied to separate different sources from a mixed signal. First a general introduction to the ICA is presented thereafter the Fast-ICA algorithm used in this study is described.

2.1 Independent Component Analysis

In ICA a decomposition of a multivariate signal into independent non-Gaussian signals are retrieved in order to find the independent components by maximizing the statistical independence of the estimated components. The basis of ICA dwells upon the assumptions that the source signals are independent of each other and that the values in each source signal have non-Gaussian distributions. Furthermore, a mixture of signals have a higher complexity than any subset of its simplest source signal. From this it follows that any signals drawn from a set of mixtures that are independent and are non-Gaussians histograms or have a low complexity are likely to be source signals [HO00]. When these assumption are correct, blind source separation using ICA has shown good results [MIZ01]. The fundamental steps of source separation using ICA is presented in Figure 1.

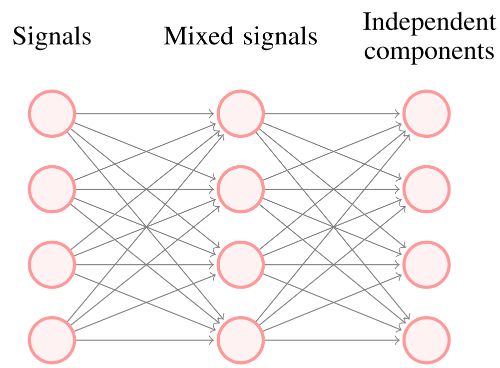


Figure 1:
$S \rightarrow X$: Transformation matrix
$X \rightarrow R$: Un-mixing matrix

Although the ICA method is not able to uniquely reconstruct the signal, the waveform of the signal is preserved. It is possible, albeit in theory, to reconstruct the signal by scaling and permute the ICA reconstruction. It is a powerful technique that has no prior knowledge of the signal [Cho+04].

Several variations of ICA algorithms have been proposed for the purpose of speech signal reconstruction, noise free Fast-ICA being an efficient and commonly used approach [HO00]. However, in many practical cases the use of a model with additive Gaussian noise is utilized. By taking noise into consideration, the signals would resemble the reality to a greater extent than the noise free approach[Zha+]. The Fast-ICA algorithm relies on the input data being *prewhitened* which can be done by an eigenvalue decomposition on the covariance matrix of the data after it has been centered, leaving the components of the signal uncorrelated and with unit variance.

The goal of the component estimation is to iterativly obtain multiple linearly independent projection vectors of the prewhitened data. Independence is in this case defined as non-Gaussianity of the estimated components and is maximized by considering a nonquadratic and nonlinear function

$f(u)$, its first derivative $g(u)$, and its second derivative $g'(u)$. In this study $f(u) = cosh(u)$ as suggested by [Hyv99]. The Fast-ICA is algorithm is described in 1.

Algorithm 1 Fast-ICA.
Input: C Number of desired components
Input: $\mathbf{X} \in \mathbb{R}^{N \times M}$ Prewhitened matrix, where each column represents an N-dimensional sample s, where $C <= N$
Output: $\mathbf{W} \in \mathbb{R}^{N \times C}$ Un-mixing matrix where each column projects $\mathbf{X}$ onto an independent component.
Output: $\mathbf{R} \in \mathbb{R}^{C \times M}$ Independent components matrix, with M columns representing a recovered sample r with C dimensions.

1: **for** p **in** 1 to C **do**
2: $\quad \mathbf{w_p} \leftarrow$ Random vector of length N
3: $\quad$ **while** $\mathbf{w_p}$ changes **do**
4: $\quad\quad \mathbf{w_p} \leftarrow \frac{1}{M}\mathbf{X}g(\mathbf{w_p}^T\mathbf{X})^T - \frac{1}{M}g'(\mathbf{w_p}^T\mathbf{X})\mathbf{1}\mathbf{w_p}$
5: $\quad\quad \mathbf{w_p} \leftarrow \mathbf{w_p} - (\sum_{j=1}^{p-1} \mathbf{w_p}^T\mathbf{w_j}\mathbf{w_j}^T)^T$
6: $\quad\quad \mathbf{w_p} \leftarrow \frac{\mathbf{w_p}}{\|\mathbf{w_p}\|}$
7: $\mathbf{W} = [\mathbf{w_1}, \ldots, \mathbf{w_C}]$
8: $\mathbf{R} = \mathbf{W^T X}$

---

3 Method

3.1 Data

In order to test the BSS using Fast-ICA we artificially mix samples from the TIDIGTS database [LD]. We decided to use four digits from four different speakers. These are selected randomly from the dataset. The process of mixing is done by linearly transform multiple samples in terms of each other. We use a linear transformation matrix defined in Figure 1. adopted from Vincent[VGF06] to mix the signals.

$$A = \begin{bmatrix} 1 & 0.5 & 1-\sqrt{0.5} & 0.5 \\ 0.5 & 1 & 0.5 & 1-\sqrt{0.5} \\ 1-\sqrt{0.5} & 0.5 & 1 & 0.5 \\ 0.5 & 1-\sqrt{.5} & 0.5 & 1 \end{bmatrix} \tag{1}$$

Figure 2: Linear transformation matrix

Furthermore, we have conducted two experiments where an additive Gaussian noise with zero mean and unit variance is used for one of the experiment. The noise is meant to represent the distortion that might occur in a room and is referred to as a *white noise*[Ste99]

The transformation is done to simulate a room with 4 sound sources placed in each corner which is shown is Figure 2. The amount of sound from each speaker picked up by the microphones depends on where they are placed and the distance from the sound sources. The recording is for simplicity simulated to be at the exact point of each speaker indicating ones in the transformation matrix. Subsequently the numbers $\neq 1$ in Figure 2 indicate the scaled distance from the other sound sources.

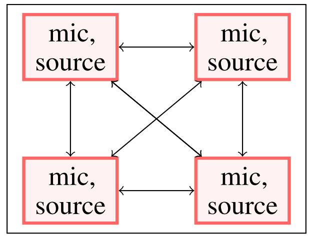


Figure 3: Simulated auditive environment

3.2 Verification

In order to proper verify and quantify the performance of the ICA algorithm this report aims to use three different verification methods. A quantifier is used to properly operationalize the result of a signal. As the ICA is described to produce signals that share a similar shape to its original this report also use a visual inspection by comparing the reconstruction with its corresponding original. Finally, a auditive inspection is performed as an another empirical measure.

To quantify the result of the reconstructed signal vector $\mathbf{r}$ and the normalized original source $\mathbf{s}$ we use the *Euclidean distance* which is defined as:

$$d(\mathbf{r}, \mathbf{s}) = \sqrt{(r_1 - s_1)^2 + (r_2 - s_2)^2 ... (r_n - s_n)^2} \qquad n = |s|$$

A lower distance should in theory indicate that the reconstructed signal resembles the source signal. The similarity is measured as the distance between the liftered Mel-Frequency Cepstral Coefficients (lMFCC) of the reconstructed signals and source signals. One thing to note is that due to the nature of the algorithm the reconstruction of the signals is often permuted meaning that the indexing of the source signal int the original matrix does not often align with the indexing of the reconstructed signals in the resulting matrix, even if the source separation was successful.

After utilizing Fast-ICA the reconstructed signals are re-scaled to be able to plug the reconstructed signals into an audio program. Thereby a small listening experiment with 10 subjects are conducted where the subjects answer whether they think the reconstruction sounds like the source signal and if there are still multiple sound sources in the reconstructed signal. This is done in order to clarify if the signal is perfectly reconstructed or not.

4 Result

This section is divided into three different sections containing each of the verification methods used in this paper. It starts with the visual verification, containing plots of different signals and its corresponding reconstructions. Following is the quantified measure in the form of two tables. Finally an auditive verification is provided in the form of histogram plots, where each plots represents a reconstructed signal.

4.1 Visual verification

Figure 4 and 5 shows the reconstructed signal and its corresponding original, without and with noise respectively. Visual inspection of the plots shows similarities in their shape and the algorithm seems to have separated each source correctly. However, a closer inspection indicates that there could be reconstructions that has not managed to completely separate the sources and therefore shares similariteis with two true signals.

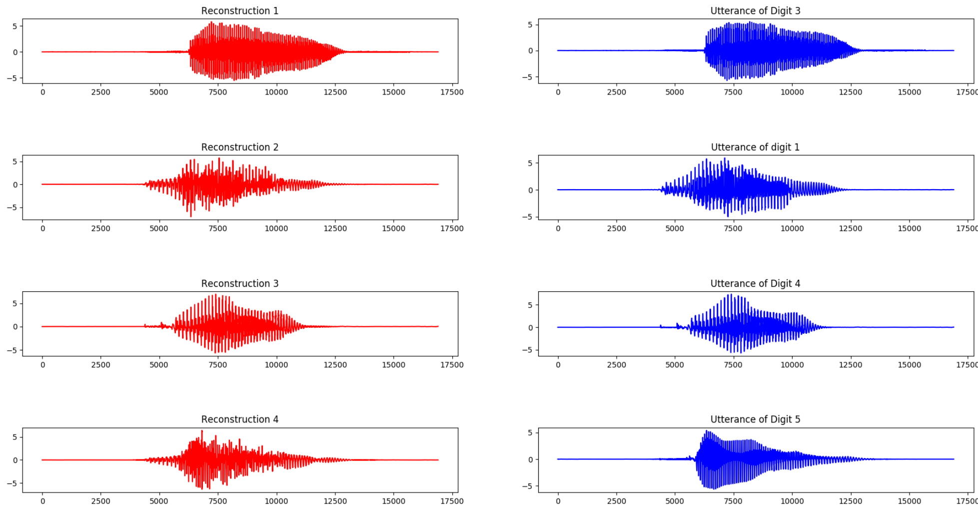


Figure 4: Left column: Reconstructed signals. Right column: Original source signals. Not that the order of the signals is not preserved

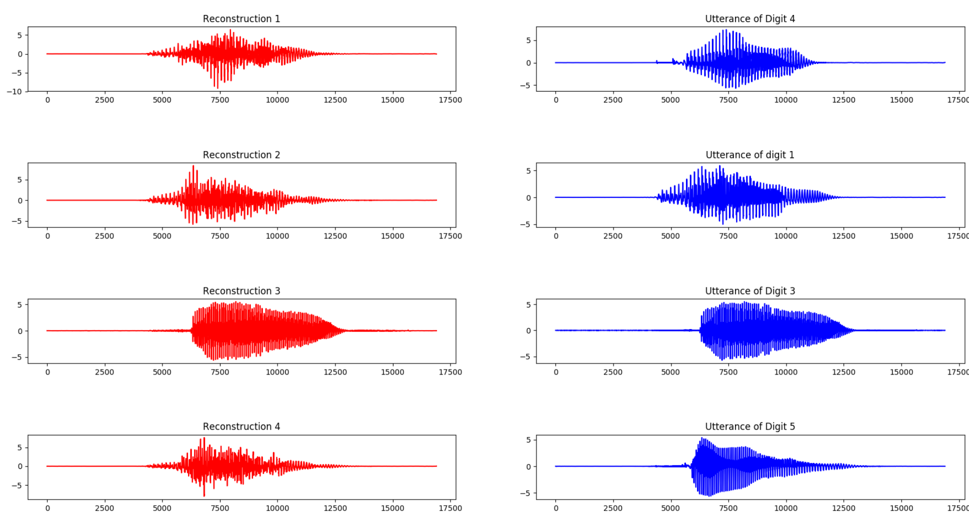


Figure 5: Left column: Reconstructed signals. Right column: Original source signals. Gaussian noise added. Note that the order of the signals is not preserved

4.2 Numerical verification

The tables below show the euclidean distances between each reconstruction to the true signal. The least distance is highlighted. The first table shows the distance between the signals, without any noise added and the second table shows the result when an additive zero mean, unit variance noise is added. The subscripts in the column header indicates the digit which is pronounced in the source signal. Due to the nature of the Fast-ICA algorithm which does not preserve the ordering of the input signals the subscript of the reconstructed signals is just a way of distinguish them from one another.

Table 1: Euclidean distance for each reconstruction against each original source signal (without noise). Source separation for different utterances. The reconstructed signal with the least distance to each source signal is highlighted.

Reconstructed / True	$T_1(_{three})$	$T_2(_{one})$	$T_3(_{four})$	$T_4(_{five})$
R_1	**2.57**	39.11	40.63	35.82
R_2	34.65	**20.97**	34.89	26.99
R_3	36.81	36.42	**12.67**	38.30
R_4	35.63	30.69	37.52	**16.91**

Table 2: Euclidean distance for each reconstruction against each original source signal (with noise). Source separation for different utterances. The reconstructed signal with the least distance to each source signal is highlighted.

Reconstructed / True	$T_1(_{four})$	$T_2(_{one})$	$T_3(_{three})$	$T_4(_{five}$
R_1	**22.30**	29.59	33.63	32.46
R_2	30.78	**25.65**	33.38	24.80
R_3	40.20	37.74	**3.85**	35.44
R_4	29.50	32.42	35.33	**23.11**

In general the lowest distances is obtained when there is no noise in the signals. The tables also indicates the case when ICA did not manage to source separate the mixed signals. In the third and second row of Table 1 and Table 2 respectively the reconstructed signal achieved the lowest distance measure to two signals. By visual inspection of 5 there is a reconstructed signal that appears to have similar shape to two true signals.

4.3 Auditive verification

The final type of verification is shown as bar plots in which each plot represent a reconstruction. They show the result of asking 10 subjects which digit or digits they can distinguish. The result shows that the ICA algorithm completely separated digit 5 (spoken by woman) from a noisy representation and the same digit from a noiseless representation. In all the other reconstructions, noisy or noiseless, the result shows that non of the digits were completely separated and that multiple digits were mixed in the signals. This is further enhanced by the result in the visual and quantitative verification. One interesting finding is that the digit 1 (spoken by man) has not been picked up at all in the noisy environment.

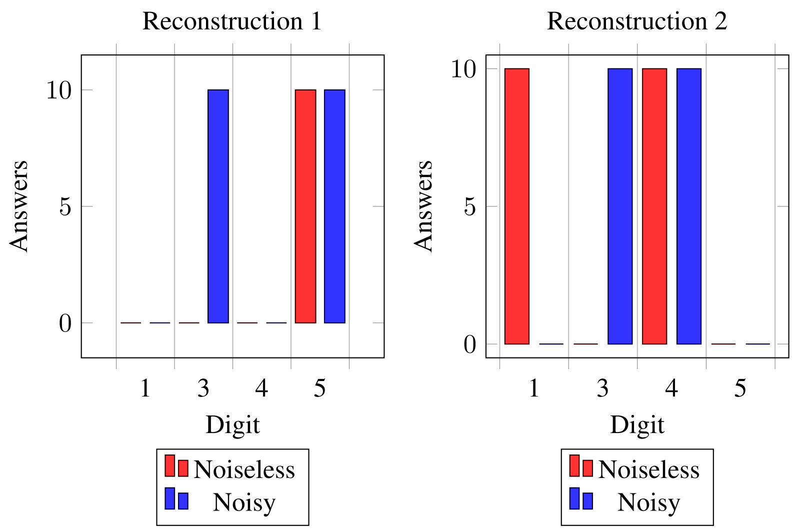
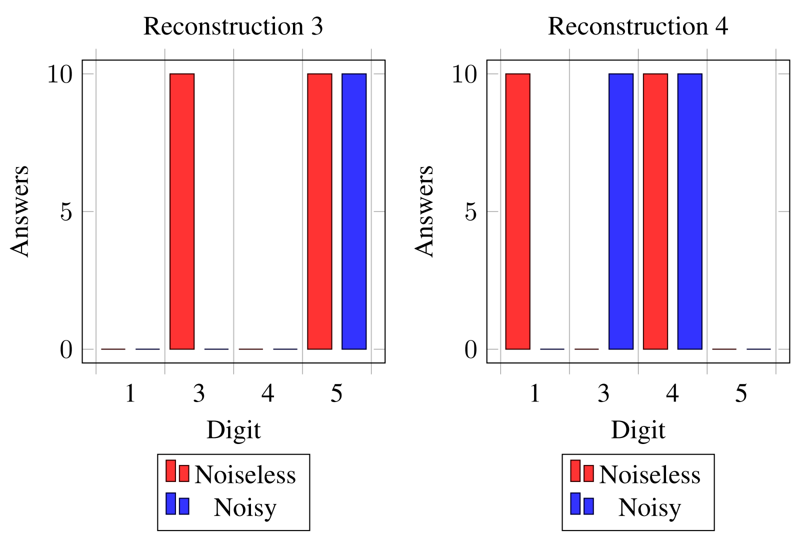


Figure 6: Bar plots showing the result of what digit the ten test subjects could distinguish from the reconstructed signal, with and without noise. Note that multiple digits occurred in one reconstruction. The different utterances are 1 (man), 3 (man), 4 (woman), 5 (woman)

The



,

Figure 7: Bar plots showing the result of what digit the ten test subjects could distinguish from the reconstructed signal, with and without noise. Note that multiple digits occurred in one reconstruction. The different utterances are 1 (man), 3 (man), 4 (woman), 5 (woman)

5 Discussion

The visual inspection showed that the approach works well in reconstructing of the shape similarity. Only looking at this result one might draw the conclusion that the algorithm successfully reconstructed the signal but due to similarities in some utterances it actually seems as some reconstructions has mixed signals in them. This is also confirmed when looking at the auditive verification. In most cases, there were multiple utterances in a single reconstruction. As a result, the algorithm does not output the expected result and was not able to completely source separate the digits. However, there were reconstructions with multiple signals in them and the auditive verification shows consensus between the subjects on which sources contained in each reconstruction. And in that regard, the ICA did manage to distinguish digits but again, not completely.

Moreover, the distance measures in Table 1 and Table 2 aligns with the visual inspection since the least distance corresponds to the most similar waveform shape. As expected the noisy environment resulted in larger distances in general compared to the non-noisy environment. However, looking at the auditive verification in Figure 6 and Figure 7 the source separation in a noisy environment was as good as in a non-noisy environment, indicating that Fast-ICA could handle noisy data to some extent.

Further, the auditive verification does not agree with the waveforms and the distance measures since the perceived digits in Figure 6 and Figure 7 don't correspond to neither the least distances nor the most similar waveforms. Hence, only looking at the waveforms or looking at the distances of the reconstructed signals is in this case not sufficient for telling if source signals are successfully separated from a mixed signal.

Another interesting aspect of the result is that the algorithm successfully distinguished at least two of the four source signals and in two cases managed to fully separate the source. However, in a noisy environment the utterance of the digit 1 could not be found at all, which implies that noise reduces the ability of reconstructing multiple signals.

A limitation of this study is that the mixed signal is artificially created by a linear combination of the given source signals which not generally the case in a reality where the the mixed signal can be complexly convoluted and distorted by echo. It would be interesting to try this approach using non-artificially mixed data to explore the behaviour of Fast-ICA in a real world environment.

6 Conclusion

Depending on the context of use Fast-ICA could be used for BSS. Our results are somewhat ambiguous but the source signals could be recovered to some extent taking into account that two sources could be comprised in one reconstructed signal. However the reconstructed signals containing multiple sources could easily be identified by auditive verification. Further research could be done on what impact noise has on the performance of Fast-ICA and how it behaves using non-artificially mixed signals.

7 Appendix

- **Literature study** As suggested by the peer reviewer, one or two sentences are added for explaning the withening of the data and also emphasizing that the result of Fast-ICA algorithm can permute the signals in the source signal matrix.
- **Correctness** In the introduction we added a little bit to why a solution to the problem is of importance. Stating the problem formulation a little bit clearer as suggested byt the reviewer. We also clearified the euclidean distance calculation which was asked for.
  The prewhitening explanation is extended stated that is a way of making the data uncorrelated.
- **Clarity of presentation** Here we decided not to follow the advise of the peer reviewer of refactoring the sections since chopping up the report in that way would compromise of the flow of the text.
  Described how we selected the different digits from different speakers
  Clarified how the auditive experiment was conducted
  Added more information about the lack of order preserveness(?) in the visual verification plots
  Clarified in the auditive result histograms what digits are heard from the experiment.

References

[Hyv99] A. Hyvarinen. "Fast and robust fixed-point algorithms for independent component analysis". In: *IEEE Transactions on Neural Networks* 10.3 (May 1999), pp. 626–634. DOI: 10.1109/72.761722. URL: https://doi.org/10.1109%2F72.761722.

[Ste99] Michael L. Stein. *Interpolation of Spatial Data*. Springer New York, 1999. DOI: 10.1007/978-1-4612-1494-6. URL: https://doi.org/10.1007%2F978-1-4612-1494-6.

[HO00] A. Hyvärinen and E. Oja. "Independent component analysis: algorithms and applications". In: *Neural Networks* 13.4-5 (June 2000), pp. 411–430. DOI: 10.1016/s0893-6080(00)00026-5. URL: https://doi.org/10.1016%2Fs0893-6080%2800%2900026-5.

[MIZ01] Noboru Murata, Shiro Ikeda, and Andreas Ziehe. "An approach to blind source separation based on temporal structure of speech signals". In: *Neurocomputing* 41.1-4 (Oct. 2001), pp. 1–24. DOI: 10.1016/s0925-2312(00)00345-3. URL: https://doi.org/10.1016%2Fs0925-2312%2800%2900345-3.

[Cho+04] Seungjin Choi et al. *Blind Source Separation and Independent Component Analysis: A Review*. 2004.

[SF05] C. Servière and P. Fabry. "Principal component analysis and blind source separation of modulated sources for electro-mechanical systems diagnostic". In: *Mechanical Systems and Signal Processing* 19.6 (Nov. 2005), pp. 1293–1311. DOI: 10.1016/j.ymssp.2005.08.001. URL: https://doi.org/10.1016%2Fj.ymssp.2005.08.001.

[VGF06] E. Vincent, R. Gribonval, and C. Fevotte. "Performance measurement in blind audio source separation". In: *IEEE Transactions on Audio, Speech and Language Processing* 14.4 (July 2006), pp. 1462–1469. DOI: 10.1109/tsa.2005.858005. URL: https://doi.org/10.1109%2Ftsa.2005.858005.

[WK09] F. Westad and M. Kermit. "Independent Component Analysis". In: *Comprehensive Chemometrics*. Elsevier, 2009, pp. 227–248. DOI: 10.1016/b978-044452701-1.00045-4. URL: https://doi.org/10.1016%2Fb978-044452701-1.00045-4.

[11] "Acoustic Beamforming using a TDS3230 DSK". In: (2011).

[ADD11] Vivek Anand, Dr. R. S. Anand. Dr, and M. L. Dewal. *Implementation of blind source separation of speech signals using independent component analysis*. 2011.

[Ade+12] Hidri Adel et al. "Beamforming Techniques for Multichannel audio Signal Separation". In: *International Journal of Digital Content Technology and its Applications* 6.20 (Nov. 2012), pp. 659–667. DOI: 10.4156/jdcta.vol6.issue20.72. URL: https://doi.org/10.4156%2Fjdcta.vol6.issue20.72.

[GSE14] Emad M. Grais, Mehmet Umut Sen, and Hakan Erdogan. "Deep neural networks for single channel source separation". In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, May 2014. DOI: 10.1109/icassp.2014.6854299. URL: https://doi.org/10.1109%2Ficassp.2014.6854299.

[NLV16] Aditya Arie Nugraha, Antoine Liutkus, and Emmanuel Vincent. "Multichannel Audio Source Separation With Deep Neural Networks". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24.9 (Sept. 2016), pp. 1652–1664. DOI: 10.1109/taslp.2016.2580946. URL: https://doi.org/10.1109%2Ftaslp.2016.2580946.

[LD] R. Gary Leonard and George R. Doddington. *TIDIGITS*.

[Zha+] Caihua Zhao et al. "An effective method on blind speech separation in strong noisy environment". In: *Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology, 2005*. IEEE. DOI: 10.1109/iwvdvt.2005.1504588. URL: https://doi.org/10.1109%2Fiwvdvt.2005.1504588.