BLIND SEPARATION OF DISJOINT ORTHOGONAL SIGNALS: DEMIXING N SOURCES FROM 2 MIXTURES

Alexander Jourjine *Scott Rickard* *Özgür Yılmaz*

Siemens Corporate Research
755 College Road East
Princeton, NJ 08540, USA
{jourjine,rickard}@scr.siemens.com

Princeton University
Fine Hall, Washington Road
Princeton, NJ 08540, USA
{srickard,oyilmaz}@princeton.edu

ABSTRACT

We present a novel method for blind separation of any number of sources using only two mixtures. The method applies when sources are (W-)disjoint orthogonal, that is, when the supports of the (windowed) Fourier transform of any two signals in the mixture are disjoint sets.

We show that, for anechoic mixtures of attenuated and delayed sources, the method allows one to estimate the mixing parameters by clustering ratios of the time-frequency representations of the mixtures. The estimates of the mixing parameters are then used to partition the time-frequency representation of one mixture to recover the original sources. The technique is valid even in the case when the number of sources is larger than the number of mixtures. The general results are verified on both speech and wireless signals. Sample sound files can be found here:
`http://www.princeton.edu/~srickard/bss.html`

1. INTRODUCTION

Demixing noisy mixtures has been a goal of long standing in the field of blind source separation(BSS). One area where BSS methods are important is wireless communications where receiving antennas measure a linear mixture of delayed and attenuated EM radiation of the source signals. Another example lies in the acoustic domain where it is desirable to separate a voice of interest from background noise and interfering speakers.

Assumptions on the statistical properties of the sources usually provide a basis for a demixing algorithm. Some common assumptions are that the sources are statistically independent[1, 2], are statistically orthogonal[3], are non-stationary[4], or can be generated by finite dimensional model spaces[5]. The independence/orthogonality assumption can be verified experimentally for speech signals and is also valid for many wireless communications schemes. Some of these methods work well for instantaneous demixing, but fail if propagation delays are present. Additionally, many algorithms are computationally intensive as they require the computation of higher-order statistical moments or the optimization of a non-linear cost function.

One area of research in blind source separation that is relatively untouched is when there are less mixtures than sources. We refer to such a case as *degenerate blind source separation*. Degenerate blind source separation poses a challenge because the mixing matrix is not invertible and the traditional method of demixing by estimating the inverse mixing matrix does not work. As a result, most of research in channel estimation and BSS has been done for the square non-degenerate case. In the related areas of wireless communication where channel estimation is important, the number of receivers is typically more then the number of emitters. For example, subspace channel estimation methods require at least one more mixture than sources to estimate the sources and the noise[6, 7].

One example of degenerate blind source separation estimates an arbitrary number of sources from a single mixture by modeling the signals as AR processes[8]. However, this is achieved at a price of approximating signals by AR stochastic processes, which can be too restrictive. Another example of degenerate separation uses higher order statistics to demix three sources from two mixtures[9]. This approach is not feasible however for a large number of sources since the use of higher order statistics of mixtures leads to an explosion in computational complexity.

Similar in spirit to this paper, van Hulle employs a clustering method for relative amplitude parameter estimation and degenerate demixing[10]. The assumptions used by van Hulle were that only one signal at a given time is non-zero and that mixing is instantaneous, that is, there is only a relative amplitude mixing parameter associated with each source. In real world acoustic environments or wireless communication domains, these assumptions are not valid.

The results of this paper are derived for anechoic time delay mixtures. We prove that for such a mixing model, estimation of the mixing parameters and complete separation of any number of disjoint orthogonal sources is possible from only two mixtures. The results can be extended to the noisy echoic case[11].

In Section 2 we define the time delay mixing model, introduce the concept of disjoint orthogonality, and describe the mixing parameter estimation. In Section 3 we describe a solution for degenerate demixing.

2. MIXING PARAMETER ESTIMATION

2.1. Source mixing

Consider measurements of a pair of sensors where only the direct path is present. In this case, without loss of general-

Proc. ICASSP2000, Volume 5, Pages 2985-2988, June 5-9, 2000, Istanbul, Turkey

ity, we can absorb the attenuation and delay parameters of the first mixture, $x_1(t)$, into the definition of the sources. The two mixtures can thus be expressed as,

$$x_1(t) = \sum_{j=1}^{N} s_j(t) + n_1(t), \quad (1)$$

$$x_2(t) = \sum_{j=1}^{N} a_j s_j(t-\delta_j) + n_2(t), \quad (2)$$

where δ_j is the arrival delay between the sensors resulting from the angle of arrival, a_j is a relative attenuation factor corresponding to the ratio of the attenuations of the paths between source and sensors, and $n_1(t)$ and $n_2(t)$ are independent white Gaussian noise signals. We use Δ to denote the maximal possible delay between sensors, and thus, $|\delta_j| \leq \Delta, \forall j$.

2.2. Source Assumptions

Given a windowing function $W(t)$, we call two functions $s_i(t)$ and $s_j(t)$ **W-disjoint orthogonal** if the supports of the windowed Fourier transforms of $s_i(t)$ and $s_j(t)$ are disjoint. The windowed Fourier transform of $s_i(t)$ is defined,

$$\mathcal{F}^W(s_i(\cdot))(\omega,\tau) = \int_{-\infty}^{\infty} W(t-\tau)s_i(t)e^{-\imath\omega t}dt, \quad (3)$$

which we will refer to as $S_i^W(\omega,\tau)$ where appropriate. The W-disjoint orthogonality assumption can be stated concisely,

$$S_i^W(\omega,\tau)S_j^W(\omega,\tau) = 0, \forall i \neq j, \forall \omega, \tau. \quad (4)$$

Note that, if $W(t) = 1$, $S_i^W(\omega,\tau)$ becomes the Fourier transform of $s_i(t)$, which we will denote $S_i(\omega)$. In this case, W-disjoint orthogonality can be expressed,

$$S_i(\omega)S_j(\omega) = 0, \forall i \neq j, \forall \omega \quad (5)$$

which we call **disjoint orthogonality**. In addition, when $W(t) = 1$, we use the Fourier transform theorem,

$$\mathcal{F}^W(s_i(\cdot-\delta))(\omega,\tau) = e^{-\imath\omega\delta}\mathcal{F}^W(s_i(\cdot))(\omega,\tau). \quad (6)$$

In the case where $W(t)$ has finite support, we will assume, as is common in array processing literature, the physical separation of the sensors is small enough relative to the carrier and bandwidth of the signal such that the relative delay between the sensors can be expressed as a phase shift of the signal[12]. This is known as the **narrowband assumption** in array processing and it implies, for our purposes, that Equation 6 holds for all δ, $|\delta| \leq \Delta$, even when $W(t)$ has finite support.

2.3. Amplitude-Delay Estimation

Consider the no noise case with $W(t) = 1$. We can rewrite the model from Equations 1 and 2 for the case with two array elements as,

$$\begin{bmatrix} X_1(\omega) \\ X_2(\omega) \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \\ a_1 e^{-\imath\omega\delta_1} & \dots & a_N e^{-\imath\omega\delta_N} \end{bmatrix} \begin{bmatrix} S_1(\omega) \\ \vdots \\ S_N(\omega) \end{bmatrix} \quad (7)$$

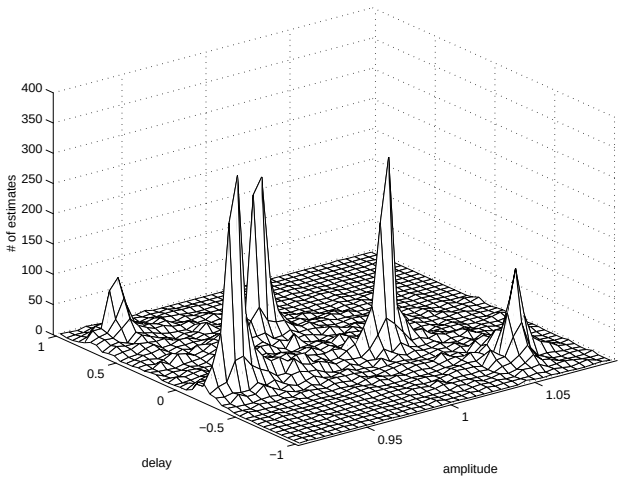

Figure 1: 2D Histogram of amplitude-delay estimates from two mixtures of five sources. The amplitude parameters were $(.98, 1.02, .93, 1.06, .93)$ and the corresponding delay parameters were $(.3, -.2, .8, -.7, -.2)$.

For disjoint orthogonal sources, we note that at most one of the N sources will be non-zero for a given ω, thus,

$$\begin{bmatrix} X_1(\omega) \\ X_2(\omega) \end{bmatrix} = \begin{bmatrix} 1 \\ a_i e^{-\imath\omega\delta_i} \end{bmatrix} S_i(\omega), \quad \text{for some } i. \quad (8)$$

Therefore, we can calculate the relative amplitude and delay parameters associated with one source using,

$$(a_i, \delta_i) = \left(\left\| \frac{X_2(\omega)}{X_1(\omega)} \right\|, \Im(\log(\frac{X_1(\omega)}{X_2(\omega)}))/\omega \right), \quad (9)$$

for some i, where $\Im$ denotes taking the imaginary part. When the noise is non-zero and $W(t)$ has finite support, Equation 9 is no longer exact, however, the mixing parameters can be approximated for a given (ω,τ) using,

$$(\hat{a}_i, \hat{\delta}_i) = \left(\left\| \frac{X_2^W(\omega,\tau)}{X_1^W(\omega,\tau)} \right\|, \Im(\log(\frac{X_1^W(\omega,\tau)}{X_2^W(\omega,\tau)}))/\omega \right), \quad (10)$$

for some i. Equation 10 has been shown to yield accurate mixing parameter estimates for appropriate $W(t)$ under a variety of noise and multipath conditions[11].

Using Equation 10, every (ω,t) yields an estimate pair for the relative amplitude-delay parameter associated with one source. For W-disjoint orthogonal signals, if we were to calculate amplitude-delay estimates from a number of time-frequency points, we would expect to see clusters around the true delay mixing parameters for each source. If we were to use a standard clustering technique on the amplitude-delay estimates, the number of clusters found would be the estimate of the number of sources, and the cluster centers would be the amplitude-delay estimates associated with each source.

Sample results of estimation of mixing parameters on mixtures of real speech signals are given in Figure 1, which shows the results of mixing parameter estimation of five sources from two mixtures. Note that two sources with the

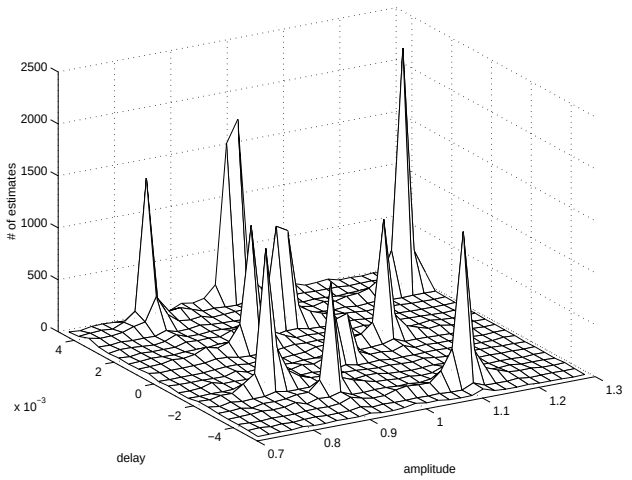

Figure 2: Two-dimensional histogram of number of estimates for delay-amplitude mixing parameters for ten M-ary FSK sources obtained using two mixtures.

same angle of arrival can be differentiated by their amplitudes alone. Mixing parameter estimation for two mixtures of ten M-ary FSK wireless signals is shown in Figure 2.

3. DEMIXING

If the number of sources is equal to the number of mixtures, the non-degenerate case, the standard demixing method is to invert the mixing matrix. We can write the mixing model for two sources as,

$$\begin{bmatrix} X_1(\omega) \\ X_2(\omega) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ a_1 e^{-\imath\omega\delta_1} & a_2 e^{-\imath\omega\delta_2} \end{bmatrix} \begin{bmatrix} S_1(\omega) \\ S_2(\omega) \end{bmatrix}. \quad (11)$$

In the non-degenerate case, the mixing parameter estimation technique described in the previous section can be used for matrix inversion demixing.

When the number of sources is greater than the number of mixtures ($N > M$), the degenerate case, matrix inversion is no longer possible. Nevertheless, in this case we can still demix by partitioning the time-frequency plane using one of the mixtures based on estimates of the mixing parameters between mixtures.

For W-disjoint orthogonal signals, using Equation 4, we know that the value of $X_1^W(\omega,\tau)$ at any frequency ω for a given τ is equal to $S_i^W(\omega,\tau)$ for some i. Moreover, the ratio $X_1^W(\omega,\tau)/X_2^W(\omega,\tau)$ depends only on the mixing parameters associated with one source. Thus, for each time-frequency point, we can determine which of the N peaks in the two-dimensional histogram of amplitude-delay estimates is closest to the $(\hat{a}_i, \hat{\delta}_i)$ estimate for the given (ω,τ). Each peak corresponds to one source, therefore, partitioning $X_1^W(\omega,\tau)$ into N time-frequency signals and converting the resulting partitioned time-frequency signals back into the time domain produces the N original source estimates.

In detail, we use Ω to denote the support of $X_1^W(\omega,\tau)$, that is, the set of (ω,τ) pairs for which with $X_1^W(\omega,\tau) \neq 0$. Similarly, the support of $S_i^W(\omega,\tau)$ is Ω_i. For W-disjoint orthogonal sources, we have, $\Omega = \cup_i \Omega_i$, and,

$$\Omega_i \cap \Omega_j = \emptyset,\ i \neq j. \quad (12)$$

For a given $(\omega,\tau) \in \Omega$, we can determine the i for which $(\omega,\tau) \in \Omega_i$ by choosing the closest cluster center to the estimate generated using Equation 10. Repeating this for every (ω,τ) in Ω, and assigning,

$$S_i^W(\omega,\tau) = X_1^W(\omega,\tau), \quad (13)$$

whenever $(\omega,\tau) \in \Omega_i$, we get the windowed Fourier transform of s_i for each i. The inverse Fourier transform of $S_i^W(\omega,\tau)$ gives us the individual source functions around $t = \tau$. By repeating this for all t, we reconstruct each source function.

An example of degenerate demixing of five speech sources from two mixtures is given in Figure 3. Tests on both anechoic and echoic degenerate mixtures show that this technique, which we call DUET (Degenerate Unmixing Estimation Technique), is an extremely robust BSS method[13].

4. SUMMARY

In this paper we presented a number of new results on demixing degenerate mixtures, a problem that has been largely unaddressed in the literature. Our approach was to assume that the source signals are W-disjoint orthogonal and then to note that mixing is, for a given time-frequency choice, just a function of one source. For the anechoic mixing model, the ratio of the windowed Fourier transform of the two mixtures for a given time-frequency choice depends only on the mixing attenuation and delay parameter associated with one source. Clustering these ratio estimates reveals the mixing parameters. Using the cluster centers to partition the windowed Fourier transform of one of the mixtures, it is possible to obtain estimates of the original sources. The fact that both estimation and separation can be done when the number of sources is larger than the number of mixtures without significant computational complexity represents a significant advancement in the state of the art.

We have verified that, perhaps surprisingly, speech signals satisfy W-disjoint orthogonality enough to allow for mixing parameter estimation and degenerate separation. Experimental evidence shows that multiple speakers talking simultaneously can be demixed with two microphones with high fidelity of recovered signals. For wireless disjoint orthogonal signals, such as frequency hopped waveforms, blind estimation of mixing parameters and blind separation of signals are achievable.

In the present work we were not able to go in depth on a number of interesting issues. Among them are the exact relationship between statistical orthogonality and disjoint orthogonality, the question of how to select the "best" windowing function, and interplay between the choice of windowing function and violation of the representation of time shifts in the time domain by complex factors in the frequency domain. One could also think of extending our methods to the non-linear media, where mixing models become non-linear and cause delays that might depend on the frequency.

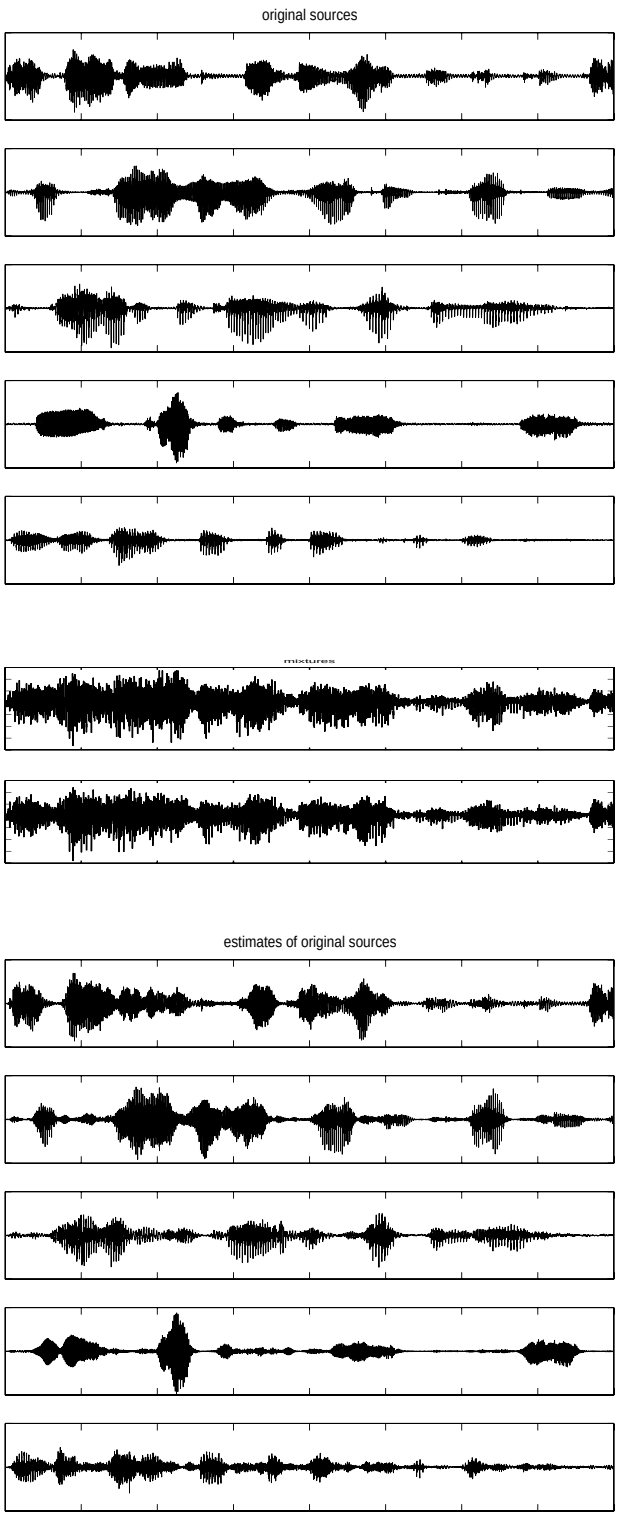

Figure 3: Five original sources, two mixtures, and the five estimates of the original sources. The separated sources average 14.3 dB SNR improvement.

On another front, it is clear that much more work needs to be done to further extend the treatment of the echoic case and in particular to derive better bounds on parameters when we expect our method to work. We plan to address these and other issues in subsequent publications.

5. REFERENCES

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

[2] J.F. Cardoso. Blind signal separation: Statistical principles. *Proceedings of IEEE, Special Issue on Blind System Identification and Estimation*, pages 2009–2025, October 1998.

[3] E. Weinstein, M. Feder, and A. Oppenheim. Multi-channel signal separation by decorrelation. *IEEE Trans. on Speech and Audio Processing*, 1(4):405–413, October 1993.

[4] L. Parra and C. Spence. Convolutive blind source separation based on multiple decorrelation. *IEEE Transactions on Speech and Audio Processing*, March 2000. Accepted for publication.

[5] H. Broman, U. Lindgren, H. Sahlin, and P. Stoica. Source separation: A TITO system identification approach. *Signal Processing*, 73:169–183, 1999.

[6] A.-J. van der Veen, S. Talwar, and A. Paulraj. A subspace approach to space-time signal processing for wireless communication systems. *IEEE Transactions on Signal Processing*, 45(1):173–190, January 1997.

[7] X. Wang and H. V. Poor. Blind multiuser detection: A subspace approach. *IEEE Transactions on Information Theory*, 44(2):677–690, March 1998.

[8] R. Balan, A. Jourjine, and J. Rosca. A particular case of the singular multivariate AR identification and BSS problems. In *1st International Conference on Independent Component Analysis*, Assuis, France, 1999.

[9] P. Comon. Blind channel identification and extraction of more sources than sensors. In *SPIE Conference*, pages 2–13, San Diego, July 19-24 1998.

[10] M. Van Hulle. Clustering approach to square and non-square blind source separation. In *IEEE Workshop on Neural Networks for Signal Processing (NNSP99)*, pages 315–323, August 1999.

[11] A. Jourjine, S. Rickard, and Ö. Yilmaz. Blind Separation of Disjoint Orthogonal Sources. Technical Report SCR-98-TR-657, Siemens Corporate Research, 755 College Road East, Princeton, NJ, Sept. 1999.

[12] H. Krim and M. Viberg. Two Decades of Array Signal Processing Research, The Parametric Approach. *IEEE Signal Processing Magazine*, pages 67–94, July 1996.

[13] `http://www.princeton.edu/~srickard/bss.html`.