MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Separation of Mixed Audio Sources by Independent Subspace Analysis

Michael A. Casey and Alex Westner

TR2001-31 September 2001

Abstract

We propose the method of independent subspace analysis (ISA) for separating individual audio sources from a single-channel mixture. ISA is based on independent component analysis (ICA) but relaxes the constraint that requires at least as many mixture observation signals as sources. A second extension to ICA is the use of dynamic components to represent non-stationary signals. Sources are tracked by similarity of dynamic components over small time steps. We propose a method for grouping components by partitioning a matrix of independent component cross-entropies that we call an ixegram. The ixegram measures the mutual similarities of components in an audio segment and clustering the ixegram yields the source subspaces and time trajectories. To demonstrate the techniques we give examples of ISA applied to separation of musical and speech sources from single-channel mixtures.

Proceedings of the International Computer Music Conference, Berlin, August 2000

MERL – A MITSUBISHI ELECTRIC RESEARCH LABORATORY
http://www.merl.com

Separation of Mixed Audio Sources by Independent Subspace Analysis

Michael A. Casey

TR-2001-31 September 2001

Abstract

We propose the method of independent subspace analysis (ISA) for separating individual audio sources from a single-channel mixture. ISA is based on independent component analysis (ICA) but relaxes the constraint that requires at least as many mixture observation signals as sources. A second extension to ICA is the use of dynamic components to represent non-stationary signals. Sources are tracked by similarity of dynamic components over small time steps. We propose a method for grouping components by partitioning a matrix of independent component cross-entropies that we call an *ixegram*. The ixegram measures the mutual similarities of components in an audio segment and clustering the ixegram yields the source subspaces and time trajectories. To demonstrate the techniques we give examples of ISA applied to separation of musical and speech sources from single-channel mixtures.

Proceedings of the International Computer Music Conference, Berlin, August 2000.

Separation of Mixed Audio Sources By Independent Subspace Analysis

Michael A. Casey and Alex Westner
Mitsubishi Electric Research Labs
Cambridge, MA, USA
mkc@merl.com

ABSTRACT

We propose the method of independent subspace analysis (ISA) for separating individual audio sources from a single-channel mixture. ISA is based on independent component analysis (ICA), a widely used method for array signal processing and feature extraction, but it extends ICA in several ways. The first extension is a method to extract statistically independent subspaces from the projection of a one-dimensional observation signal onto a manifold, such as the short-time Fourier transform or a constant-Q spectrogram. The projection relaxes the constraint of ICA-based separation systems that requires there to be at least as many mixture observation signals as there are sources. A second extension to ICA is the use of dynamic independent components to represent non-stationary signals. Sources are tracked by similarity of dynamic components over small time steps. We propose a method for grouping components by partitioning a matrix of independent component cross-entropies that we call an *ixegram*. The ixegram measures the mutual similarities of components in an audio segment and clustering the ixegram yields the source subspaces and time trajectories. To demonstrate the techniques we give examples of ISA applied to separation of musical and speech sources from single-channel mixtures.

INTRODUCTION

1.1 Auditory scene analysis

One goal of auditory scene analysis is to extract individual audio sources from a mixture of sources. Example scenarios are separating speech from interfering background sounds and separating individual musical instruments from a polyphonic ensemble, [1].

A number of computational methods for auditory scene analysis have been proposed that use combinations of signal descriptions, such as sinusoidal tracks, correlograms and wide-band noise models, to represent low-level audio elements, [2][3][4]. Once a signal has been decomposed into fundamental representations, further stages attempt to identify groups of related components that form auditory streams. The stream formation algorithms use perceptually-motivated heuristics, such as common onset of harmonically-related components and amplitude or frequency co-modulation of components, [5][6][7].

Due to the parametric representation of signal elements and the heuristic nature of psycho-acoustic grouping rules, the task of designing robust systems for automatic scene analysis is rather complicated. Consequently, the performance of these systems has been limited and difficult to measure in practice.

Recently, however, blind signal separation systems have been proposed that use independent component analysis for separating unknown sources from a mixture. ICA techniques make no explicit assumptions about the composition and grouping of signal components but instead rely on the statistical properties of the latent sources for their identification. These methods enable source separation without parameter fitting and thus show much promise for robust automatic scene analysis implementations.

1.2 ICA and BSS

The problem of identifying unknown sources in a mixture is called blind signal separation (BSS) and has found utility in many signal processing applications. A number of methods have been proposed for solving various forms of BSS, and among them is independent component analysis, [8][9][10][11].

ICA assumes that the individual source components in an unknown mixture have the property of mutual statistical independence, and this property is exploited in order to algorithmically identify the latent sources. It has been shown that many of the proposed algorithms for ICA share a common mathematical foundation, this has been used to demonstrate the equivalence of many ICA approaches, [12].

Recall that statistical independence for a vector random variable is defined by the property that the joint

probability density function factors into the product of the marginal densities such that,

$$P_{\mathbf{u}}(\hat{\mathbf{u}}) = \prod_{i=1}^{N} P_{u_i}(\hat{u_i}). \quad (1)$$

The canonical BSS/ICA model expresses the observation signal as the product of a mixing matrix and vector of statistically independent signals,

$$\mathbf{x} = A\mathbf{s}, \quad (2)$$

where $A = [\mathbf{a}_1, \ldots, \mathbf{a}_\rho]$ is a $n \times \rho$ invertible mixing matrix with linearly independent columns, $\mathbf{s} = [s_1\ s_2\ \ldots\ s_\rho]^T$ is a random vector with ρ statistically independent source signals and $\mathbf{x} = [x_1\ x_2\ \ldots\ x_n]^T$ is an n-dimensional observable random vector with $n \geq \rho$.

The BSS problem is solved by finding an *unmixing* matrix $W \approx A^{-1}$ using only the observed signals $\mathbf{x}$ and assuming statistical independence of the sources $\mathbf{s}$. A is assumed full rank. W is chosen so that the output signals $\mathbf{u}$ are as statistically independent as possible,

$$\mathbf{u} = W\mathbf{x} = WA\mathbf{s}. \quad (3)$$

One method of solving ICA is to derive an update rule for W that minimizes the mutual information between the output joint density function and the marginal densities, [10]:

$$I(\mathbf{u}) = \int P_{\mathbf{u}}(\mathbf{u}) \log \left(\frac{P_{\mathbf{u}}(\mathbf{u})}{\prod_{i=1}^{N} P_{u_i}(u_i)} \right) d\mathbf{u}. \quad (4)$$

This approach can be reduced to a contrast function on the outputs, $\Phi(\mathbf{u})$, for which cumulants up to fourth order are maximized. The extremum of this function corresponds to maximum contrast from Gaussian in each component. Therefore, independent component analysis yields a higher-order decorrelation between components. Higher order decorrelation is stronger than that of principal component analysis (PCA) which produces decorrelation of distributions up to second order. Decorrelation is equivalent to independence only in the case of complete characterization of the distribution by second-order moments as in the Gaussian case.

As stated above, an equivalence class of algorithms is known that performs the factorization of a mixture vector into a vector of statistically independent non-Gaussian sources. In general, the latent signals are identifiable up to a permutation and scaling, thus W is indeterminate and uniqueness constraints on the ICA solution must be applied.

1.3 ICA and auditory scene analysis

The ability to factor a signal into statistically independent components makes ICA an attractive prospect for computational auditory scene analysis. However, BSS algorithms are based on assumptions that are not practical for auditory scene analysis systems. The most limiting of these assumptions is that there must be at least as many observable mixture signals as source signals and that the mixing matrix be full rank.

However, auditory scene analysis canonically assumes that there are fewer sensors than sources and typically reduces to a single sensor problem. Thus the BSS assumption on the dimensionality of $\mathbf{x}$ does not hold, so the BSS form of ICA is not adaptable to the problem of single-channel extraction.

To use the separation properties of ICA for auditory scene analysis we propose a non-BSS method that extracts maximally contrasting features from a single mixture. The features are invertible so good approximations to source signals can be estimated. Our proposed method expresses single-channel auditory scene analysis as independent subspace separation in a manifold.

INDEPENDENT SUBSPACE ANALYSIS

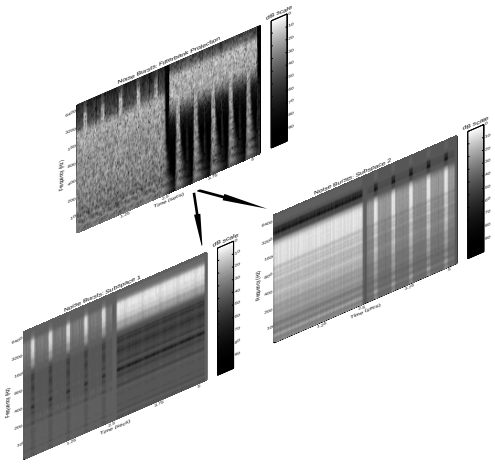

Figure 1: Decomposition of a spectrogram into separate spectrogram subspaces. Two streams from a psychoacoustic noise-burst sequence are correctly separated by independent subspace analysis.

ISA extends ICA by identifying independent multi-component source subspaces of an input vector. Hyvarinen [13], discovered emergent complex cell properties for vision by independent subspace analysis on images. ISA-related methods have also been explored

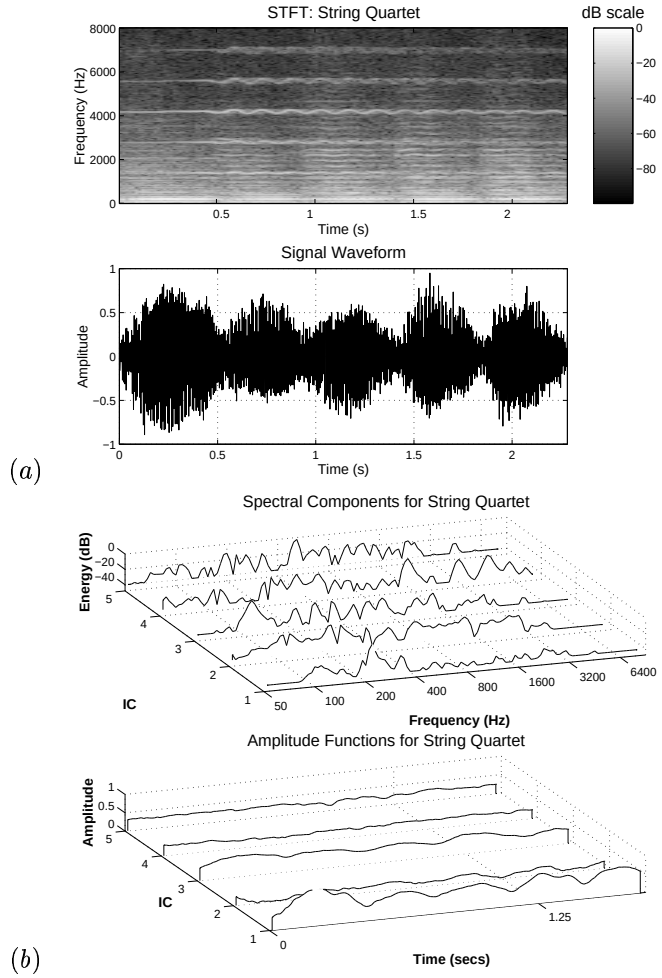

Figure 2: (a) Constant-Q spectrogram of 1 bar from a Beethoven string quartet. (b) 5 independent components extracted from the string quartet. The upper components are the spectral basis $\mathbf{z}_i$, and the lower components are decimated temporal functions, $\mathbf{y}_i$.

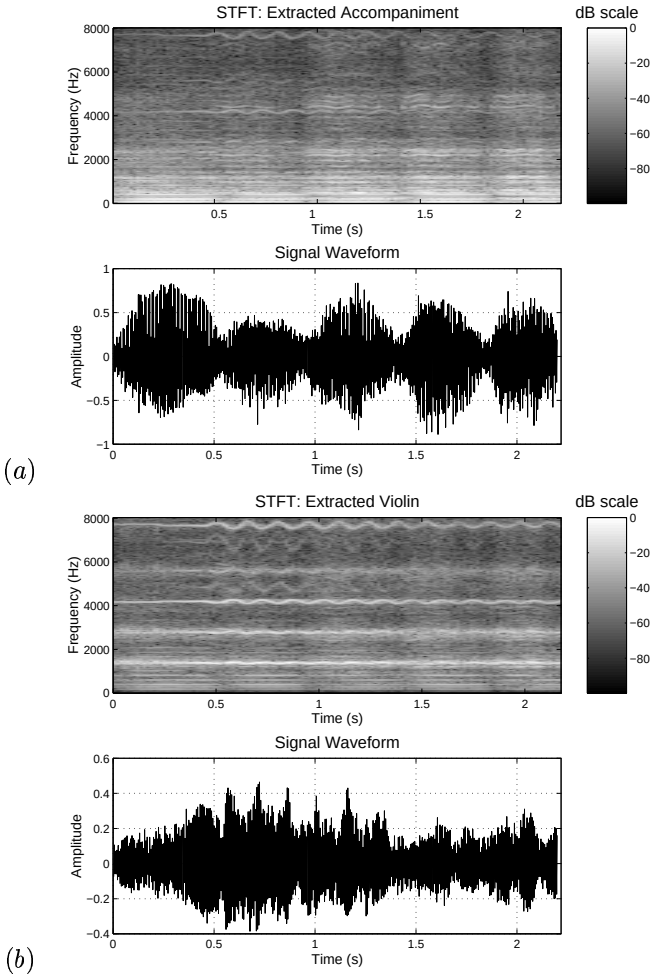

Figure 3: Short-time Fourier transform spectrogram subspaces and recovered signals for (a) extracted accompaniment chords composed of violin-II, viola and 'cello. (b) extracted violin-I melody note with vibrato.

by Lathauwer *et al.* [14], who proposed a subspace version of ICA, and Cardoso [15], who proposed multi-dimensional ICA (MICA) for subspace identification. Both of these works investigated the application of ISA techniques to the analysis of fetal ECG recordings.

In these studies, the source signals are multi-dimensional since they are observed by $N > 1$ sensors. However, in contrast to the one-dimensional sources in Equation 3, the ISA/MICA approach assumes that each estimated source component is an n-tuple composed of $k \geq 1$ signals, therefore a mixture signal is decomposed into multi-dimensional source signals.

In the domain of sound, independent component analysis has been used for decomposition of natural sounds into independent controllable features, [16]. The method operates on a single-channel input signal that is projected onto a frequency basis using the short-time Fourier transform. The extracted components are invertible and the resynthesis of a sound is therefore controllable along independent subspace dimensions due to the independence of the features. The current work extends this previous work by decomposing a spectrogram into independent source subspaces and inverting them to yield source separation.

2.1 Spectrogram subspace separation

Figure 1 shows spectrogram subspace separation of a noise-burst sequence, used for psycho-acoustic experiments, into the correct individual perceptual streams, [1]. Figures 2 and 3 show decomposition of one bar from a Beethoven string quartet into separate subspaces using independent component analysis. By inspection, the first violin subspace is represented by component number 4 shown in Figure 2(b), and the accompaniment chords on second violin, viola and 'cello occupy the subspace spanned by the remaining 4 com-

ponents. The extracted components form a spectrogram for each subspace that is inverted to yield a separated signal. Figures 3(a) and 3(b) show the results of extraction applied to the Beethoven string quartet excerpt. Interestingly, the sources correspond to different functions in the string quartet; the first violin holds a melody note, and the remaining instruments play a chord in a repeating rhythm that accompanies the first violin. This type of separation has many potential applications in automatic music analysis and machine listening.

The decomposition shown above operates on a one-dimensional source mixture signal composed of c independent sources,

$$s(t) = \sum_{j=1}^{c} s_j(t). \tag{5}$$

A spectrogram is obtained by projection to a new basis. The one-dimensional input signal $s(t)$ is split into finite-length random vectors; $\mathbf{s}^{(k)} \in \mathbf{R}^w$ where $1 \leq k \leq m$ is an ordered frame index. The windowed signal is multiplied by a $n \times w$ transformation matrix M, where n is the number of channels of the transformation. Taking the absolute value of the transformed signal produces an observation vector $\mathbf{x}^{(k)} \in \mathbf{R}^n$ for each windowed input frame k. We may represent a magnitude Fourier transform or a smoothed constant-Q filterbank output in this manner,

$$\mathbf{x}^{(k)} = M^T \mathbf{s}^{(k)}. \tag{6}$$

By convention, each column of a spectrogram corresponds to a spectral slice which is a snapshot of the spectrum at time k. The rows contain the spectral channels. Each frame of the input spectrogram is expressed as a weighted sum of ρ independent basis vectors, $\mathbf{z}_i \in \mathbf{R}^n$. These basis vectors are themselves spectral slices that represent *features* of the spectrum that can be separated due to their statistical independence.

The basis vectors are defined to be static but each is weighted by a time-varying scalar coefficient, $y_i^{(k)}$. A weighted sum of ρ basis vectors reconstructs a spectrogram frame from independent features:

$$\mathbf{x}^{(k)} = \sum_{i=1}^{\rho} y_i^{(k)} \mathbf{z}_i. \tag{7}$$

The utility of the subspace method is greatest when the independent spectral features correspond to individual sources in a mixture. Each source is spanned by a subset of such basis vectors that define a subspace. The subspaces are composed of a matrix with basis vectors in the columns,

$$Z_j = \left[\mathbf{z}_1^{(j)},\ \mathbf{z}_2^{(j)},\ \cdots\ \mathbf{z}_{\rho_j}^{(j)}\right],\ Z_j \subseteq \{Z\}. \tag{8}$$

Now, to reconstruct a spectrogram frame, each source subspace is included as a weighted sum of its basis vectors. The weight coefficients are obtained by projection of the input onto each basis component in the subspace. Assuming orthonormal components,

$$\mathbf{y}_j^T = Z_j^T \mathbf{x} \tag{9}$$

which is the projection of vector $\mathbf{x}$ onto the subspace spanned by the basis vectors Z_j. By successively projecting on to each of the c sets of basis vectors, the frames of the input spectrogram are decomposed into sums of independent subspaces,

$$\mathbf{x} = Z_1 \mathbf{y}_1^T + Z_2 \mathbf{y}_2^T + \cdots + Z_c \mathbf{y}_c^T. \tag{10}$$

To extend the method to a block of spectrogram frames, the familiar transposed two-dimensional form of a spectrogram is expressed as a matrix: $X(n,k) \equiv X^T$, which is a matix with n rows and k columns,

$$X^T = \sum_{j=1}^{c} Z_j Y_j^T. \tag{11}$$

This expression partitions a block of spectrogram frames into separate spectrograms formed from static multi-component subspaces of the input manifold. Each spectrogram corresponds to a subspace and is obtained from a set of basis vectors:

$$X_j^T = Z_j Y_j^T \tag{12}$$

The weights, Y_j, corresponding to each subspace are obtained by projection of the input spectrogram against each subspace basis in a similar manner to Equation 9:

$$Y_j = X Z_j. \tag{13}$$

Finally, the separated source signals, s_j, are obtained by inverting the transformation carried out by Equation 6. This step completes the subspace separation method for projections of a one-dimensional signal onto a manifold.

$$\mathbf{s}_j = M^{-1} X_j \tag{14}$$

Extensions of the subspace method to the complex case are well defined with $X^H \in \mathbf{C}^{n \times k}$, where X^H indicates the Hermitian transpose, thus,

$$X^H = Z_j Y_j^H,\ Z_j \in \mathbf{C}^{n \times \rho_j},\ Y_j \in \mathbf{C}^{k \times \rho_j}. \tag{15}$$

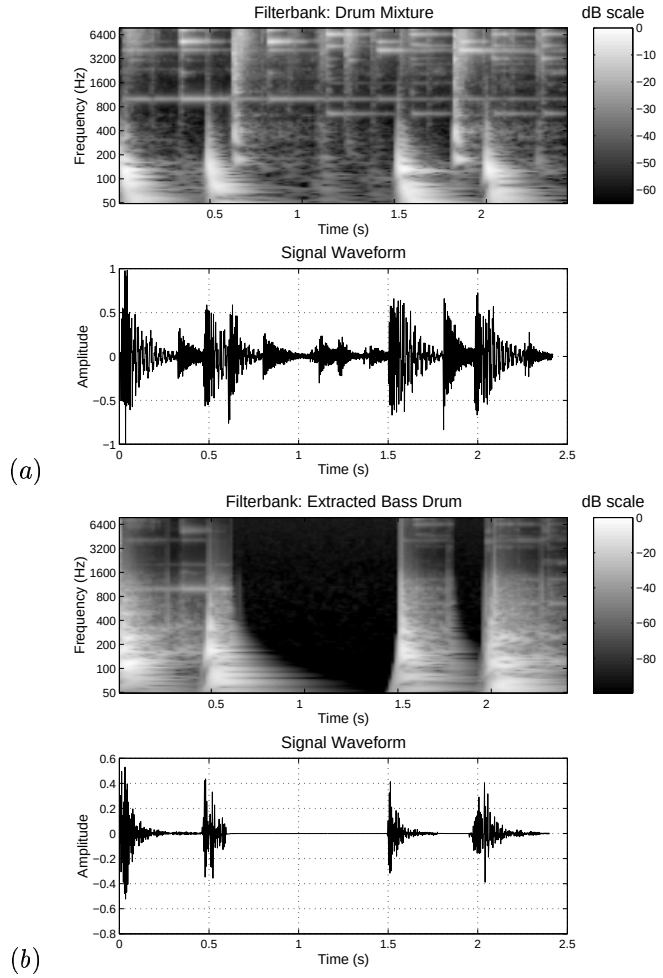

Figure 4: (a) Constant-Q spectrogram of a drum mixture. (b) Extracted bass drum subspace

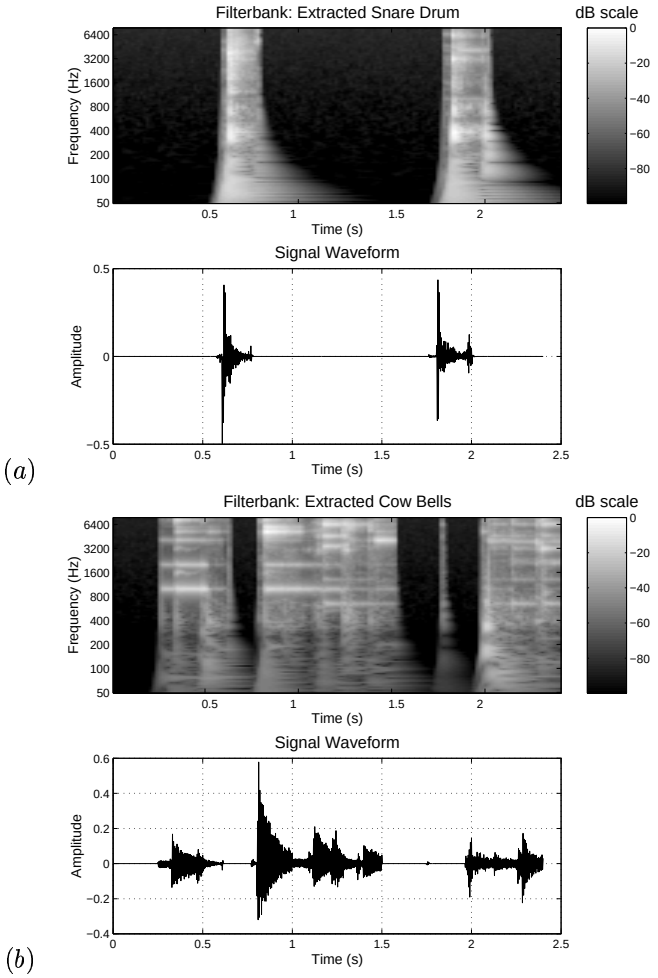

Figure 5: (a) Extracted snare drum subspace. (b) Extracted cowbell subspace. The components were analysed in 0.25s intervals and grouped in time using ixegram clustering.

2.2 Non-stationary sources

The extension of ISA to non-stationary spectrogram components is derived by assuming subspaces to be approximately stationary for an interval of spectrogram frames, δk. The decomposition of a spectrogram is expressed in blocks of frames, each block having a unique subspace decomposition.

We represent a blocked version of ISA by rewriting Equation 7 to include a block index l,

$$\mathbf{x}^{(k,l)} = \sum_{i=1}^{\rho} y_i^{(k,l)} \mathbf{z}_i^{(l)} \tag{16}$$

which admits a unique subspace decomposition for each block.

$$[X^{(l)}]^T = \sum_{j=1}^{c} Z_j^{(l)} [Y_j^{(l)}]^T. \tag{17}$$

Blocks are composed of independent spectrogram subspaces, $[X_j^{(l)}]^T$, spanned by a subset of basis vectors,

$$[X_j^{(l)}]^T = Z_j^{(l)} [Y_j^{(l)}]^T. \tag{18}$$

The blocks may overlap if a temporally smooth subspace decomposition is required. Block lengths are chosen in the range $0.25s \leq \delta k \leq 10s$ and block hop sizes are typically chosen to be half the block length. If each spectrogram frame is $.02s$ then a block length of $0.5s$ produces an observation matrix, $X^{(l)}$, with 25 rows.

Figures 4 and 5 show a non-stationary decomposition of a drum mixture into separate source subspaces. The spectrograms are reconstructed from $0.25s$ blocks of basis vectors. Blocks corresponding to similar sources are identified by component grouping which is discussed in Section 3.1.

2.3 Maximally informative subspace

In the preceeding section, a set of basis vectors defines a subspace of the input that is subjected to ICA decomposition into source subspaces. The size of the set, ρ, is chosen to be smaller than the number of variates, n. An estimate of ρ is given by the non-zero singular values of a singular value decomposition (SVD) of the input matrix,

$$X = U\Sigma V^T. \tag{19}$$

The choice of ρ determines how many of the right singular vectors, $\mathbf{v}_i \in V$, will be passed to ICA extraction. These vectors define a subspace of the input, and this subspace is chosen to be maximally informative.

An SVD orders basis vectors by the size of their singular values, σ_i, which are the diagonal elements of Σ. A threshold, $0 < \phi \leq 1$, is defined such that $\frac{1}{\sum_{i=1}^{n}\sigma_i}\sum_{i=1}^{\rho}\sigma_i \geq \phi$ which gives a value for ρ.

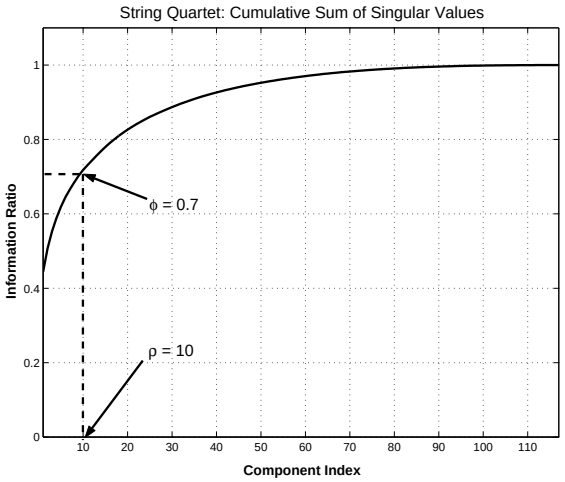

Figure 6: Cumulative sum of basis component singular values for the string quartet extract. Singular values are normalized such that $\sum_{i=1}^{n}\sigma_i = 1$. An optimal number of basis components, ρ, is chosen to maximize the information contained in the chosen basis subspace.

By Gibbs second theorem, a Gaussian distribution maximizes the entropy over all other distributions with the same mean and variance. The threshold, ϕ, is thus interpreted as an approximation to the proportion of information to retain in the chosen subspace, see Figure 6.

When $\phi = 1$, the full basis is retained from an SVD for ICA decomposition. The resulting basis vectors have a Gabor filter structure with each component having a compact region of support in frequency. Whilst this is a significant result from the perspective of data-derived basis functions, such bases are not useful as recognizable source features.

However, when $\phi \ll 1$ the basis components appear as spectral features with support across the entire frequency range. This indicates that their is a tradeoff between the amount of detail to keep in the ICA subspace, and the recognizability of the resulting features. Complete information support yields general orthogonal basis functions, and deflation to a maximally informative subspace yields distinct features of the input sources.

INDEPENDENT SUBSPACE GROUPING

To identify the components belonging to a multi-component subspace some type of grouping must be performed. We introduce a method for calculating the similarities of components that enables partitioning into subspaces using the pair-wise dissimilarities of independent components.

3.1 The IXEGRAM

The similarity of components is represented in an *ixegram*, the independent component cross-entropy matrix. The ixegram is computed by exhaustive pair-wise similarity measures over the set of independent components using an approximation to the symmetric Kullback-Leibler divergence. The Kullback-Leibler divergence takes a scalar random variable, $\hat{u}$, and produces a measure of the distance between two probability functions, $P_a(\hat{u})$ and $P_b(\hat{u})$. The ixegram entries are defined by the following expressions,

$$D(i,j) = \delta_{KL}(\mathbf{z}_i, \mathbf{z}_j),\ i,j \in \{1 \cdots n\}, \tag{20}$$

$$\delta_{KL}(\mathbf{z}_i, \mathbf{z}_j) \equiv KL(P_{\mathbf{z}_i}(\hat{u}), P_{\mathbf{z}_j}(\hat{u})). \tag{21}$$

The symmetric Kullback-Leibler divergence between two probability density functions, p and q, defined on a random variable, $\hat{u}$, is given by:

$$\begin{aligned} KL(p(\hat{u}), q(\hat{u})) &= \frac{1}{2}\int p(\hat{u}) \log\left(\frac{p(\hat{u})}{q(\hat{u})}\right) d\hat{u} \\ &+ \frac{1}{2}\int q(\hat{u}) \log\left(\frac{q(\hat{u})}{p(\hat{u})}\right) d\hat{u} \end{aligned}$$

So the independent component vectors, $\mathbf{z}_i$ and $\mathbf{z}_j$, are used to calculate probability functions in the range of $\hat{u}$ and the ixegram calculates the pair-wise Kullback-Leibler distance between all probability functions in the set of components.

Since the symmetric Kullback-Leibler divergence is a distance measure, it has the following useful properties: $KL(P_{\mathbf{z}_i}(\hat{u}), P_{\mathbf{z}_j}(\hat{u})) \geq 0$ and, furthermore, $KL(P_{\mathbf{z}_i}(\hat{u}), P_{\mathbf{z}_j}(\hat{u})) = 0$ iff $P_{\mathbf{z}_i}(\hat{u}) = P_{\mathbf{z}_j}(\hat{u})$.

The ixegram matrix has the following structure:

$$D = \begin{bmatrix} \delta_{KL}(\mathbf{z}_1, \mathbf{z}_1) & \delta_{KL}(\mathbf{z}_1, \mathbf{z}_2) & \cdots & \delta_{KL}(\mathbf{z}_1, \mathbf{z}_n) \\ \delta_{KL}(\mathbf{z}_2, \mathbf{z}_1) & \delta_{KL}(\mathbf{z}_2, \mathbf{z}_2) & \cdots & \delta_{KL}(\mathbf{z}_2, \mathbf{z}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{KL}(\mathbf{z}_n, \mathbf{z}_1) & \delta_{KL}(\mathbf{z}_n, \mathbf{z}_2) & \cdots & \delta_{KL}(\mathbf{z}_n, \mathbf{z}_n) \end{bmatrix}$$

The symmetric Kullback-Leibler divergence produces a matrix that is square symmetric therefore $D^T = D$. All the entries in D are non-negative and the diagonal terms are necessarily all 0.

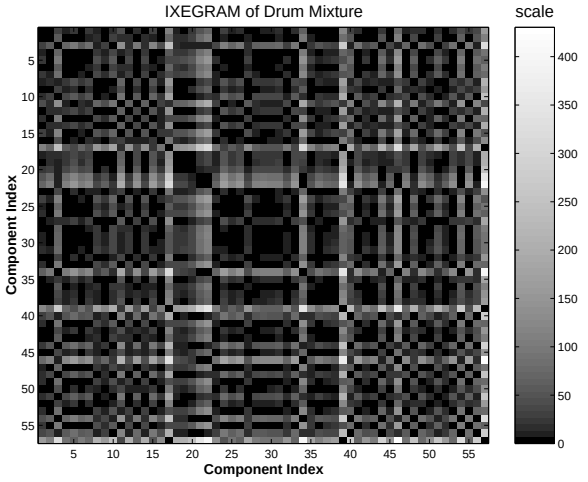

Figure 7: An *ixegram* of time-varying independent components extracted from the drum mixture shown in Figure 4. Dark regions indicate a high degree of similarity based on the Kullback-Leibler entropy.

An ixegram of time-varying independent components for the drum mixture is shown in Figure 7. The dark regions show components that are most similar. Components are identified by their row and column indices, as in $D(i, j)$.

3.2 Clustering the IXEGRAM

Due to the form of the ixegram being a symmetric, non-negative, dissimilarity matrix, it lends itself to grouping with a diadic data clustering algorithm such as that proposed by Hofmann, [17].

To find an optimal partitioning of the ixegram into k classes, a cost function that measures within-cluster compactness and between-cluster homogeneity is defined,

$$H(M; D) = \sum_{c=1}^{k} \frac{1}{\sum_{j=1}^{n} M_{jc}} \sum_{i=1}^{n} \sum_{k=i}^{n} M_{ic} M_{kc} D_{ik}$$

where D is the ixegram matrix and M is a $n \times k$ assignment matrix, each entry of which is the probability of assigning component $\mathbf{z}_i$ in class c_j,

$$M = \begin{bmatrix} P(\mathbf{z}_1|c_1) & P(\mathbf{z}_1|c_2) & \cdots & P(\mathbf{z}_1|c_k) \\ P(\mathbf{z}_2|c_1) & P(\mathbf{z}_2|c_2) & \cdots & P(\mathbf{z}_2|c_k) \\ \vdots & \vdots & \ddots & \vdots \\ P(\mathbf{z}_n|c_1) & P(\mathbf{z}_n|c_2) & \cdots & P(\mathbf{z}_n|c_k) \end{bmatrix}$$

A deterministic annealing algorithm finds the optimal M that minimizes the cost in Equation 3.2 given an ixegram, D. This probabilistic clustering yields groups of components, assigned by M, both vertically, into multi-component subspaces, and horizontally, into dynamic component trajectories.

Figure 8 shows the results of clustering dynamic independent components into a source subspace trajectory for a speaker in the context of continuous waterfall noise. The extracted speech subspace clearly identifies when the speech is present and significantly attenuates the background noise. This example demonstrates the application of the subspace separation method to speech signals and concludes our discussion of dynamic independent subspace separation.

SUMMARY

In this paper we have introduced methods for independent subspace analysis for single-channel audio mixtures. It was shown that independent component analysis can be expressed as source separation from the projection of a single-channel mixture onto a high-dimensional manifold. The method was shown to identify multi-component source subspaces of the manifold that contain independent source elements of the input mixture.

Independent subspace analysis operates on subsets of basis components spanning the input manifold. These subsets are chosen to be maximally informative using a cumulative sum of singular-values threshold criterion.

We introduced the *ixegram* as a measure space for grouping independent basis components based on the Kullback-Leibler differential entropy. The results of separation and grouping experiments on a number of contrasting examples suggest that the technique can perform separation of source signals without parametric model fitting or prior knowledge of the composition of input data.

Future work will explore the use of ISA for identifying sources in context for the purpose of multi-media indexing and annotation.

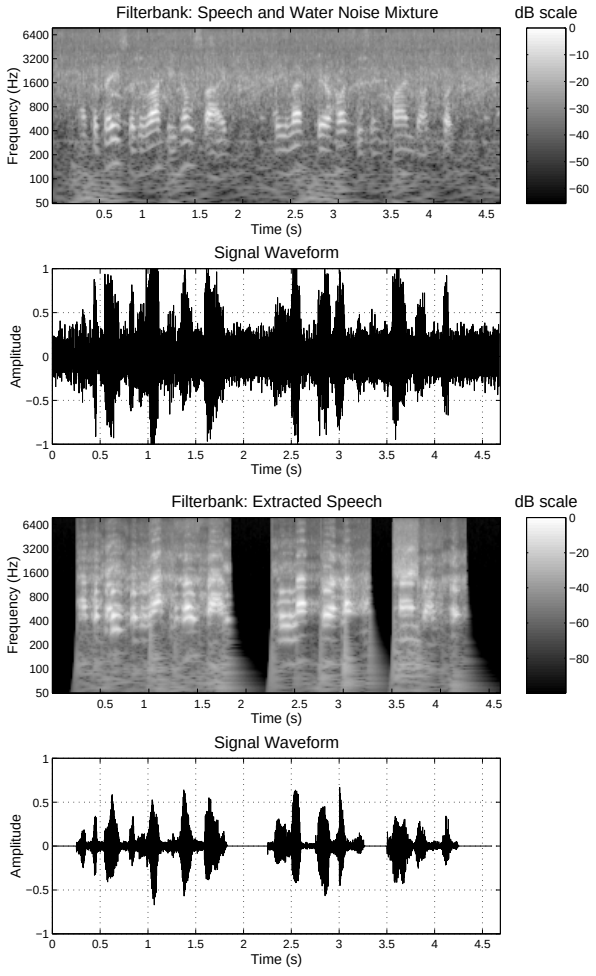

Figure 8: (a) Constant-Q spectrogram representation of a speech and waterfall noise mixture. (b) Extracted time-varying speech subspaces grouped into time trajectories using an ixegram. Components were analysed in $0.25s$ blocks and clustered into 2 groups.

References

[1] A.S. Bregman. *Auditory Scene Analysis.* MIT Press, 1990.

[2] D. Ellis. Hierarchic models of hearing for sound separation and reconstruction. In *Proc. IEEE Workshop on Apps. of Sig. Proc. to Acous. and Audio*, Mohonk, 1993.

[3] M.P. Cooke. *Modeling auditory processing and organisation.* PhD thesis, CS dept., Univ. of Sheffield, 1991.

[4] G.J. Brown. *Computational auditory scene analysis: A representational approach.* PhD thesis, CS dept., Univ. of Sheffield, 1992.

[5] D. Ellis. *Prediction-Driven Computational Auditory Scene Analysis.* PhD thesis, Masaschusetts Institute of Technology, Media Laboratory, 1996.

[6] D. F. Rosenthal and H. Okuno, editors. *Computational auditory scene analysis*, chapter Mid-level representations for computational auditory scene analysis: the weft element, pages 257–272. Lawrence Erlbaum, Mahwah, NJ, 1998.

[7] E.D. Scheirer. Sound scene segmentation by dynamic detection of correlogram comodulation. Technical Report TR491, Massachusetts Institute of Technology, April 1999.

[8] C. Jutten and J. Herault. Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture. *Signal processing*, 24:1–10, 1991.

[9] J-F. Cardoso and Antoine Souloumiac. Blind beamforming for non gaussian signals. *IEE Proceedings*, 140(6):362–370, December 1993.

[10] P. Comon. Independent component analysis, a new concept? *Signal Processing, Elsevier*, 36(3):287–314, April 1994. Special issue on Higher-Order Statistics.

[11] A. Bell and T. Sejnowski. An information-maximisation approach to blind separation and blind deconvolution. *Neural Comp.*, 7:1129–1159, 1995.

[12] T. Lee, M. Girolami, A. Bell, and T. Sejnowski. A unifying information-theoretic framework for independent component analysis. *International Journal on Mathematical and Computer Modeling*, 1998.

[13] A. Hyvarinen and P. Hoyer. Independent subspace analysis shows emergence of phase and shift invariant features from natural images. In *Proc. Int. Joint Conf. on Neural Networks*, Washington, D.C., 1999.

[14] L. Lathauwer, D. Callaerts, B. Moor, and J. Vandewalle. Fetal electrocardiogram extraction by source subspace separation. In *Proc. IEEE SP Workshop on Stat. Signal Array Proc*, pages 356–359, 1996.

[15] J-F. Cardoso. Multidimensional independent component analysis. In *Proc. ICASSP '98. Seattle*, 1998.

[16] M.A. Casey. *Auditory Group Theory: with application to statistical basis methods for structured audio.* PhD thesis, Massachusetts Institute of Technology, Media Laboratory, February 1998.

[17] Thomas Hofmann and Joachim M. Buhmann. Pairwise data clustering by deterministic annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(1):1–14, 1997.