Chapter III: Statistical Basis Decomposition of Time-Frequency Distributions

3.1 Introduction

In the previous chapters we outlined the need for a method for decomposing an input time-frequency distribution (TFD) into independently controllable features that can be used for re-synthesis. In this chapter we describe a suite of techniques, related to principal component analysis (PCA), that decompose a TFD into statistically independent features. As we shall show in this chapter, statistically-independent decomposition of a Gaussian distributed TFD is performed by a singular value decomposition (SVD). For non-Gaussian TFDs we develop an independent component analysis (ICA) algorithm.

We first introduce the concept of PCA and the necessary mathematical background. We then consider the computation of a robust PCA with SVD and develop the theory for SVD in Section 3.3.7. In Section 3.3.8 we give an example of the application of SVD to sound-structure modeling which demonstrates the potential merits of the technique. We then consider some important limitations of SVD in Section 3.3.15 which are due to the implicit dependence on second-order statistics only. In Section 3.3.16 we consider extensions to SVD to include higher-order statistical measures and, in Section 3.3.18, we consider an information-theoretic interpretation of PCA which provides the framework for developing a higher-order independent component analysis (ICA) algorithm for feature decomposition.

3.2 Time Frequency Distributions (TFDs)

As with any signal characterization scheme, there must be a front-end which decomposes the signal into low-level mathematical objects for further treatment. In this section we shall outline several representations which could be used for a front-end analysis, and we make our choice for further development based on several design criteria; i) efficiency of the transform, ii) data preservation and invertibility, iii) ease of implementation.

Most of the salient characteristics of audio signals exist in the short-time spectro-temporal domain. That is the domain of representation of a signal in which time-varying spectral features can be represented directly without need for further transformation. An example of such an analysis is the well-known short-time Fourier transform (STFT).

3.2.1 Desirable Properties of the STFT as a TFD

Although the short-time Fourier transform is limited in its characterization abilities it does have several very desirable properties. Firstly it can be implemented extremely efficiently using the fast Fourier transform (FFT). For most sound analysis applications an FFT-based analysis will run in real time on standard microcomputer hardware. Secondly, since the Fourier transform can be thought of as a linear operator, there are well-defined signal-processing operations which produce stable, invertible results that are easily implemented without the call for compensating machinery as is the case for many other TFD representations. Thus, for the purposes of sound modeling, we consider that the STFT is a reasonable time-frequency representation.

The main problem of interest with the STFT, as with most TFD representations, is in the redundancy of spectral information. The STFT with an appropriately selected analysis frequency errs on the side of inclusion rather than on the side of omittence of important information. Therefore it is with little loss in generality that we choose the STFT as a front-end frequency analysis method in the following sections. It should be emphasized, however, that all of the statistical basis reduction techniques presented can be applied to any TFD; we shall give examples of the application of statistical basis reduction methods to alternate TFDs in Chapter IV.

3.2.2 Short-Time Fourier Transform Magnitude

It was Ohm who first postulated in the early nineteenth century that the ear was, in general, phase deaf, Risset and Mathews (1969). Helmholtz validated Ohm's claim in psycho-acoustic experiments and noted that, in general, the phase of partials within a complex tone (of three or so sinusoids) had little or no effect upon the perceived result. Many criticisms of this view ensued based on faulting the mechanical acoustic equipment used, but Ohm's and Helmholtz' observations have been corroborated by many later psycho-acoustic studies, Cassirer (1944).

The implications of Ohm's acoustical law for sound analysis are that the representation of Fourier spectral components as complex-valued elements possessing both a magnitude and phase component in polar form is largely unnecessary, and that most of the relevant features in a sound are represented in the magnitude spectrum. This view is, of course, a gross simplification. There are many instances in which phase plays an extremely important role in the perception of sound stimuli. In fact, it was Helmholtz who noted that Ohm's law didn't hold for simple combinations of pure tones. However, for non-simple tones Ohm's law seems to be well supported by psycho-acoustic literature.

Consideration of Ohm's acoustical law has lead many researchers in the speech and musical analysis/synthesis community to simplify Fourier-based representations by using the magnitude-only

spectrum. In the case of the STFT this results in a TFD known as the short-time Fourier transform magnitude (STFTM), Griffin and Lim (1989). The downside in using the STFTM representation appears at the re-synthesis stage. Because the phases have been eliminated a phase-model must be estimated for a given STFTM, the phase must be constrained in such a manner as to produce the correct magnitude response under the operation of an inverse Fourier transform and Fourier transform pair. This property of a phase model is expressed by the following relation:

$$\left|\hat{\mathbf{Y}}\right| \approx \left|\mathrm{FT}\{\mathrm{FT}^{-1}\{|\mathbf{Y}|e^{-j\hat{\phi}}\}\}\right| \qquad \textbf{[82]}$$

where $|\mathbf{Y}|$ is a specified STFTM data matrix, $\hat{\phi}$ is a phase-model matrix and $|\hat{\mathbf{Y}}|$ is the approximated magnitude response matrix for the given magnitude specification and phase model. Since there is a discrepancy between $|\hat{\mathbf{Y}}|$ and $|\mathbf{Y}|$ for most values of $\hat{\phi}$ a least-squares iterative phase estimation technique is used to derive the phase model, Griffin and Lim (1984). We discuss this technique further in the next chapter.

Without loss of generality, then, we will use the STFTM representation in the examples given in this chapter. The algorithms are defined for complex-valued spectra but work on magnitude-only spectra without the need for modification.

3.2.3 Matrix Representation of TFDs

We represent an arbitrary TFD by a matrix $\mathbf{X}$ which we refer to as the data matrix. In the case of the STFT the data matrix is:

$$\mathbf{X}_{mn} = \mathrm{X}[l, k] \qquad \textbf{[83]}$$

where m and n are the row and column indices of a matrix $\mathbf{X}$. Thus the data matrix can be thought of as a two-dimensional plane with points (m, n). This interpretation of the data matrix will be useful when we discuss applications of auditory group transforms to TFDs.

The statistical basis reduction techniques discussed later in this chapter are sensitive to the orientation of the data matrix. The is due largely to the consideration of *variates* in vector form for which measures of a particular variable occupy the columns of a matrix. Thus a data matrix has the *variates* in the columns and the *observations* in the rows.

3.2.4 Spectral Orientation

As defined above the data matrix is in spectral orientation. That is, the variates are functions of the frequency variable $\omega_k = \frac{2\pi k}{N}$. There are N columns such that each column represents the complex spectral value of a signal at a particular frequency $\frac{2\pi n}{N}$ where n is the column index of the data matrix. Thus in spectral orientation the observations are the time-varying values of the spectrum at a particular frequency.

$$\mathbf{X} = \begin{bmatrix} \mathrm{X}(1,0) & \mathrm{X}(1,1) & \ldots & \mathrm{X}(1,N-1) \\ \mathrm{X}(2,0) & \mathrm{X}(2,1) & \ldots & \mathrm{X}(2,N-1) \\ \ldots & \ldots & \ldots & \ldots \\ \mathrm{X}(M,0) & \mathrm{X}(M,1) & \ldots & \mathrm{X}(M,N-1) \end{bmatrix} \quad \textbf{[84]}$$

The corresponding covariance matrix is $N \times N$ and is defined by:

$$\Phi_{\mathbf{X}} = \mathrm{E}[\mathbf{X}^{\mathrm{T}}\mathbf{X}] - \mathbf{m}^{\mathrm{T}}\mathbf{m} \quad \textbf{[85]}$$

where $\mathbf{m}$ is a vector of column means for the data matrix.

3.2.5 Temporal Orientation

An alternative method of representing a TFD using matrix notation is to orient the matrix temporally. The variates are functions of the time-frame variable l and the observations operate through frequency.

$$\mathbf{X} = \begin{bmatrix} \mathrm{X}(1,0) & \mathrm{X}(2,0) & \ldots & \mathrm{X}(M,0) \\ \mathrm{X}(1,1) & \mathrm{X}(2,1) & \ldots & \mathrm{X}(M,1) \\ \ldots & \ldots & \ldots & \ldots \\ \mathrm{X}(1,N-1) & \mathrm{X}(2,N-1) & \ldots & \mathrm{X}(M,N-1) \end{bmatrix} \quad \textbf{[86]}$$

In spectral orientation the covariance matrix is $M \times M$. In general the choice of orientation of a data matrix is determined by the desirable characterization properties of any subsequent analysis. If the matrix is in temporal orientation then a covariant statistical analysis, one that relies upon the covariance matrix, will yield results that are sensitive to the time-frame variates. However, since for most sound analysis purposes $M \gg N$, the cost of computation of the covariance and subsequent decomposition can be prohibitively great, or at least many orders of magnitude greater than computing the covariance in spectral orientation, see Sandell and Martins (1995).

3.2.6 Vector Spaces and TFD Matrices

For a given TFD in spectral orientation the frequency variates span the column space and the observations span the row space of the data matrix. The row vector space is generally much larger than the column vector space in spectral orientation, and the converse is true of temporal orientation.

1. Column Space of a TFD

The column space $\Re(\mathbf{X})$ of an $m \times n$ TFD matrix is a subspace of the full m-dimensional space which is $\mathbf{R}^m$ in the case of a spectrally-oriented STFTM representation, that is the m-dimensional vector space spanning the field of reals. The dimension of the column space is of interest to us here. It is defined as the rank $\mathbf{r}$ of the matrix which is the number of linearly independent columns.

2. Row Space of a TFD

Conversely, the row space $\Re(\mathbf{X}^T)$ is a subspace of $\mathbf{R}^n$. Thus the co-ordinates represented by a set of observations can be thought of as a linear combination of the column vectors, which are the *basis*, and conversely the basis functions themselves can be thought of as a linear combination of observations. The dimension of the row space is the rank $\mathbf{r}$ which is also the number of linearly independent rows.

3. Null Space of a TFD

TFD representations contain a good deal of redundancy. This redundancy manifests itself in the null space $\aleph(A)$ of the data matrix. For an $m \times n$ TFD matrix the null space is of dimension $n - \mathbf{r}$ and is spanned by a set of vectors which are a basis for the null space. For TFD data matrices the null space arises from the correlated behavior between the variates. The correlations between frequency bins in a spectrally-oriented TFD data matrix are expressed as linear dependencies in the vector spaces. Thus information about one of the correlated components is sufficient to specify the other components, therefore the remaining components are not well-defined in terms of a vector space for the TFD matrix. In many cases the dimensionality of the null-space of a TFD is, in fact, larger than the dimensionality of the column and row spaces, both of which are $\mathbf{r}$. From this observation we form a general hypothesis about the vector spaces of TFDs:

$$rank\{\aleph(\mathbf{X})\} \gg rank\{\Re(\mathbf{X})\} \qquad \textbf{[87]}$$

from which it follows that $n \gg 2\mathbf{r}$. Estimation of the rank of the TFD thus provides a measure of the degree of redundancy within a sound with respect to the chosen basis of the TFD.

3.2.7 Redundancy in TFDs

For any given frequency analysis technique, the chosen basis functions for projecting a signal into the time-frequency plane are extremely elemental. In the case of the STFT these basis functions are a set of complex exponentials linearly spaced in frequency. Each analysis bin of an STFT frame is thus a projection of the time-domain signal onto an orthogonal basis spanned by the said exponential functions. In the case of the continuous Fourier transform the basis is infinite thus defining a Hilbert space and the DFT (which is used by the STFT) effectively samples this space at discrete intervals. Such a basis is designed to span all possible complex-valued sequences representing each spectral component as a point in a high-dimensional space. Indeed Fourier's theorem states that *any* infinitely-long sequence can be decomposed into an infinite sum of complex exponentials. Thus each infinitesimal frequency component within a signal gets an independent descriptor.

Clearly natural sounds are not this complex. There is a good deal of redundancy in the signals. Much of the redundancy is due to the grouped nature of physical vibrations. That is, a set of frequencies generated by a system are generally related to a fundamental mode of vibration by some form of statistical dependence. We have seen this behavior in the form of the acoustical equations given in Chapter II. The inter-partial dependence of a sound spectrum may be a linear function, a non-linear function, resulting in harmonic, inharmonic or stochastic components, but in each case there is a non-zero joint probability between many of the marginal components defined for each

frequency of vibration. Such statistical dependence within a sound results in uniform motion of a set of points in the time-frequency plane of a TFD. The motion may be linear or non-linear but, nevertheless, the resulting spectrum is statistically dependent to some degree.

Redundancy is an important value for information in the signal since by eliminating it we are able to see what actually varies during the course of a sound and, by inspecting it, we see what stays essentially the same. In fact, the concept of redundancy has been the subject of some perceptual theories. For example, Barlow (1989) considers the concept of redundancy to be fundamental to learning and argues that it is redundancy that allows the brain to build up its “cognitive maps” or “working models” of the world.

Somewhat less ambitious is the claim that redundancy in the low-level projection of a sound onto a spectral basis is a necessary component to extracting meaningful features from the sound, or at least it is a good point of departure for investigating methods for characterizing the structure of natural sounds. This observation leads quite naturally to an information theoretic interpretation of the task of feature extraction and characterization of natural sound TFDs.

3.3 Statistical Basis Techniques for TFD Decomposition

3.3.1 Introduction

In view of the prevailing redundancy in TFDs we seek methods for identifying the null space and characterizing the row and column spaces in terms of a reduced set of basis vectors. The general hypothesis is that the reduced space will represent salient information in the TFD. A stronger hypothesis is that the redundancy-reduced basis may represent the perceptually most important information in the signal. These are the ideas to be investigated in this section.

3.3.2 Principal Component Analysis (PCA)

Principal component analysis was first proposed in 1933 by Hotelling in order to solve the problem of decorrelating the statistical dependency between variables in multi-variety statistical data derived from exam scores, Hotelling (1933). Since then, PCA has become a widely used tool in statistical analysis for the measurement of correlated data relationships between variables, but it has also found applications in signal processing and pattern recognition for which it is often referred to as the Karhunen-Loeve transform, Therrien (1989). The use of PCA in pattern recognition is born out of its ability to perform an optimal decomposition into a new basis determined by the second-order statistics of the observable data.

3.3.3 Previous Audio Research using PCA

The use of Principal Component Analysis for audio research can be traced back to Kramer and Mathews (1956) in which a PCA is used to encode a set of correlated signals. In the 1960s there was some interest in PCA as a method for finding salient components in speech signals, of particular note is the work of Yilmaz on a theoryof speech perception based on PCA, (Yilmaz 1967a,

1967b, 1968), and the application of PCA to vowel characterization, (Plomp *et al.* 1969; Klein *et al*. 1970; Zahorian and Rothenburg 1981). Yilmaz was concerned with the identification of invariants in speech, thus his work is perhaps the most relevant to the current work. PCA has also been applied in the processing of audio signals for pattern recognition applications by basis reduction of the Short-Time Fourier Transform (STFT), Beyerbach and Nawab (1991), and in modeling Head-Related Transfer Functions for binaural signal modeling, Kistler & Wightman (1992).

In addition to speech and acoustical encoding, PCA of musical instrument sounds has been researched quite extensively, (Stautner 1983; Stapleton and Bass 1988; Sandell and Martens 1996). The results for musical instrument modeling are reported to be of widely varying quality with little or no explaination of why some sounds are better characterized than others by a PCA. In the following sections we develop an argument that suggests some important limitations with PCA, and with its numerical implementation using SVD. This leads us to a new approach for decomposing time-frequency representations of sound into statistically salient components.

3.3.4 Definition of PCA

PCA has many different definitions but they all have several features in common. These can be summarized as follows:

PCA Theorem: *The k-th principal component of the input vector* $\mathbf{x}$ *is the normalized eigenvector* $\mathbf{v}_k$ *corresponding to the eigenvalue* λ_k *of the covariance matrix* Φ_x*, where the eigenvalues are ordered* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$.

where the covariance matrix is defined in Equation 85. A proof of this theorem may be found in Deco and Obradovic (1996). A PCA is, then, a linear transform matrix $\mathbf{U}$ operating on a TFD matrix $\mathbf{X}$ as follows:

$$\mathbf{Y} = \mathbf{XV} \qquad [88]$$

with $\mathbf{Y}$ representing the linearly-transformed TFD matrix. If the rows of the linear transformation matrix $\mathbf{V}^T$ are the eigenvectors of the covariance matrix Φ_x then it is said to perform a Karhunen-Loeve Transform of the input column space $\Re(\mathbf{X})$. In this case $\mathbf{V}$ is an orthonormal matrix and thus satisfies the following relations:

$$\mathbf{VV}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I} \qquad [89]$$

and the relationship between the input and output covariance can be expressed as:

$$\Phi_y = \mathbf{V}^T\Phi_x\mathbf{V} = \mathbf{V}^T\mathbf{V}\Sigma = \Sigma \qquad [90]$$

where Σ is a diagonal matrix of eigenvalues which correspond to the variances of a set of independent Gaussian random variables which span the input space:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \sigma_N^2 \end{bmatrix}. \quad [91]$$

(For a derivation of diagonalization of the covariance matrix see Appendix II). Under this definition, the PCA essentially linearly decorrelates the output variates in $\mathbf{Y}$ such that each column is statistically independent to second order with respect to the other columns. Traditionally, in statistical texts, the matrix of eigenvectors $\mathbf{V}$ is referred to as the weights matrix and the linearly transformed matrix $\mathbf{Y}$ is referred to as the scores of the PCA. This nomenclature follows Hotelling's original formulation.

3.3.5 Joint Probability Density Functions and Marginal Factorization

We now assume a statistical interpretation of TFD data matrix variates. The probability density function (PDF) of each column of the input is defined as a marginal density in the joint probability density of the column space of the TFD. A definition of statistical independence is derived in the form of the relationship between the joint probability distribution of the columns of a TFD and the individual column marginal distributions. Specifically, the output columns are statistically independent if and only if:

$$p_{\mathbf{Y}}(\mathbf{Y}) = \prod_{i=1}^{N} p_{\mathbf{Y}_i}(\mathbf{Y}_i) \quad [92]$$

that is, the output marginals are independent PDFs if and only if their joint density function can be expressed as a product of the marginals. In the case of Gaussian input densities, PCA decorrelates the input PDF to second order and thus exhibits the marginal factorization property described by Equation 92, see Comon (1994) and Deco and Obradovic (1996).

3.3.6 Dynamic Range, Scaling, Rank, Vector Spaces and PCA

There are several problems with PCA as defined above for the purposes of TFD decomposition. The first is that since PCA is defined as a diagonalization of the input covariance, the system loses sensitivity to lower magnitude components in favor of increasing the sensitivity of higher magnitude components. This is because the input covariance is essentially a *power* representation of the input variates. The result of transforming to a power representation is a loss in dynamic range due to finite word-length effects and numerical precision in floating-point implementations.

This relates to an issue on the usefulness of PCA in general. PCA depends on the scaling of the input coordinates. This is referred to in the literature as the "scaling problem". The problem manifests itself in the solution of the diagonalization using eigenvalues. The pivoting requires scaling of each row in order to yield a Gaussian elimination, the condition number of the TFD matrix deter-

mines the sensitivity of the data to scaling and whether or not the matrix is indeed singular to working precision.

PCA does not define a solution when the columns of the input matrix are linearly dependent. In this case the null space of the matrix is non empty. In fact, for TFDs we have already developed the hypothesis that the null space is in fact much larger than the row and column space of the data matrix, see Equation 87. Equivalently we can interpret the identification of the size of the null space as a rank estimation problem, we can see this in the relation defined in Equation 95. The PCA definition as diagonalization of the covariance does not explicitly provide a method for handling the null space of a matrix. This is because methods involving the identification of eigenvalues rely on full-row rank of the data covariance matrix. Therefore this form of PCA is of little practical use in implementing redundancy reduction techniques for TFDs. However, we shall refer to the canonical PCA form in our discussions in the following sections since the basic theoretical framework is somewhat similar for the null-space case as well as the case of non-Gaussian input distributions.

Another problem with the definition of PCA in the form outlined above is that the resulting basis spans only the column space of the input. Thus it does not generalize to the problem of identifying a basis for the row space. The covariance matrix is necessarily square which renders it invertible under the condition of full column rank. The covariance is also a symmetric matrix which is defined by the relation $\Phi_x = \Phi_x^T$, thus the row space and the column space of the input representation are collapsed to the same space, namely a power representation of the column space of the TFD. In performing a PCA using the covariance method we are thus discarding information about the space of row-wise observations in favor of characterizing the column-wise variates.

In order to address the problems of dynamic range and row/column space basis identification, we seek a representation which does not rely on the covariance method; rather, the sought method should directly decompose the input TFD into a separate basis for the row and column space of the TFD data matrix. We know that the rank of the row and column spaces is equal thus the null space will be the same from both points of view.

We now develop practical techniques for decorrelating input components of a TFD. These techniques are defined so as to address the problems of dynamic range, scaling, vector-space representation and matrix rank that we have discussed in this section.

3.3.7 The Singular Value Decomposition (SVD)

The singular value decomposition has become an important tool in statistical data analysis and signal processing. The existence of SVD was established by the Italian geometer Beltrami in 1873 which was only 20 years after the conception of a matrix as a multiple quantity by Cayley. As we shall see, the singular value decomposition is a well-defined generalization of the PCA that addresses many of the problems cited above.

A singular value decomposition of an $m \times n$ matrix $\mathbf{X}$ is any factorization of the form:

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^{T} \quad [93]$$

where $\mathbf{U}$ is an $m \times m$ orthogonal matrix; i.e. $\mathbf{U}$ has orthonormal columns, $\mathbf{V}$ is an $n \times n$ orthogonal matrix and Σ is an $m \times n$ diagonal matrix of singular values with components $\sigma_{ij} = 0$ if $i \neq j$ and $\sigma_{ii} \geq 0$; (for convenience we refer to the ith singular value $\sigma_i = \sigma_{ii}$). Furthermore it can be shown that there exist non-unique matrices $\mathbf{U}$ and $\mathbf{V}$ such that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_N \geq 0$. The columns of the orthogonal matrices $\mathbf{U}$ and $\mathbf{V}$ are called the left and right singular vectors respectively; an important property of $\mathbf{U}$ and $\mathbf{V}$ is that they mutually orthogonal.

We can see that the SVD is in fact closely related to the PCA. In fact the matrix product $\mathbf{U}\Sigma$ is analogous to the matrix $\mathbf{Y}$ defined for PCA:

$$\mathbf{Y} = \mathbf{XV} = \mathbf{U}\Sigma \quad [94]$$

Because both the singular vectors defined for an SVD are square and have orthonormal columns their inverses are given by their transposes. Thus $\mathbf{V}^{-1} = \mathbf{V}^{T}$. Now the relation in Equation 94 can be expressed $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^{T}$ which is the definition of an SVD.

The first major advantage of an SVD over a PCA is that of rank estimation and null-space identification. $\aleph\{\mathbf{X}\}$ can be identified for both the left and right singular vectors as the space spanned by vectors corresponding to the singular values for which $\sigma_j = 0$, whereas if $\sigma_j \neq 0$ then the corresponding singular vectors $\mathbf{U}_j$ and $\mathbf{V}_j$ are in the range of $\mathbf{X}$ which is spanned by the column space of the left and right singular vectors which, in turn, span the row space and column space of the data matrix $\mathbf{X}$.

The upshot of these observations is that we can construct a basis for each of the vector spaces of $\mathbf{X}$. Recalling the relation between the rank of the null space and the rank of the row and column spaces of a matrix:

$$rank\{\aleph(\mathbf{X})\} = N - rank\{\Re(\mathbf{X})\} \quad [95]$$

the SVD provides a theoretically well-defined method for estimating the rank of the null space, specifically it is the number of zero-valued singular values. This in turn defines the rank of the data matrix $\mathbf{X}$.

The SVD defined thus has implicitly solved the problems inherent in the PCA definition. Firstly, the SVD decomposes a non-square matrix, thus it is possible to directly decompose the TFD representation in either spectral or temporal orientation without the need for a covariance matrix. Furthermore, assuming a full SVD, the decomposition of a transposed data matrix may be derived from the SVD of its complimentary representative by the relation:

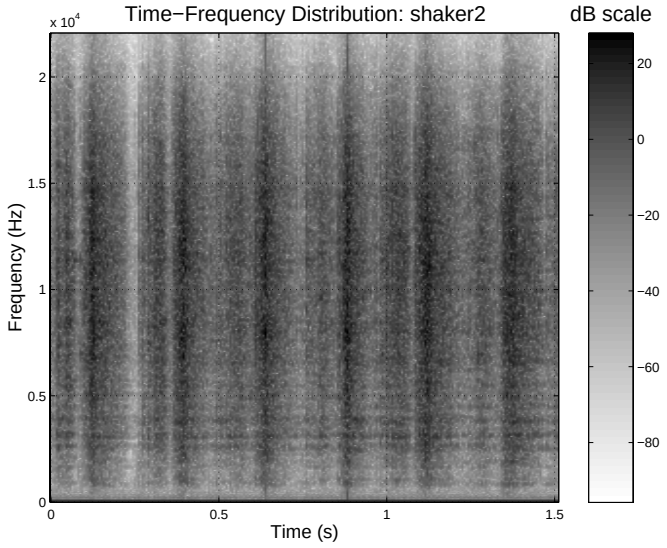

FIGURE 8. Short-time Fourier transform TFD of 1.5 seconds of a percussive shaker. The vertical dark regions are the shake articulations. Analysis parameters are N=1024, W=512, H=256. Sample rate is 44.1kHz.

$$\mathbf{X}^{\mathrm{T}} = \mathbf{V}\Sigma\mathbf{U}^{\mathrm{T}} \qquad \textbf{[96]}$$

which follows from the relation $\Sigma^{\mathrm{T}} = \Sigma$. This means that the full SVD decomposition of a matrix in spectral orientation can be used to specify an SVD decomposition in temporal orientation and *vice-versa*. Thus the direct SVD decomposition keeps all the relevant information about the null, row and column spaces of a data matrix in a compact form.

Since the SVD decomposes a non-square matrix directly without the need for a covariance matrix, the resulting basis is not as susceptible to dynamic range problems as the PCA. Thus, components of a TFD that lye within the working precision of a particular implementation are not corrupted by squaring operations. Theoretically it is not in fact possible to invert a non-square matrix. Thus implementation of a SVD is a compromise between the theoretical definition and practically tractable forms. The machinery of compromise in the SVD is the psuedoinverse of a matrix.

3.3.8 Singular Value Decomposition of Time-Frequency Distributions

3.3.9 A Simple Example: Percussive Shaker

Figure 8 shows the STFTM TFD of a percussive shaker instrument being played in regular rhythm. The observable structure reveals wide-band articulatory components corresponding to the shakes and a horizontal stratification corresponding to the ringing of the metallic shell. What does not show clearly on the spectrogram is that the rhythm has a dual shake structure with an impact

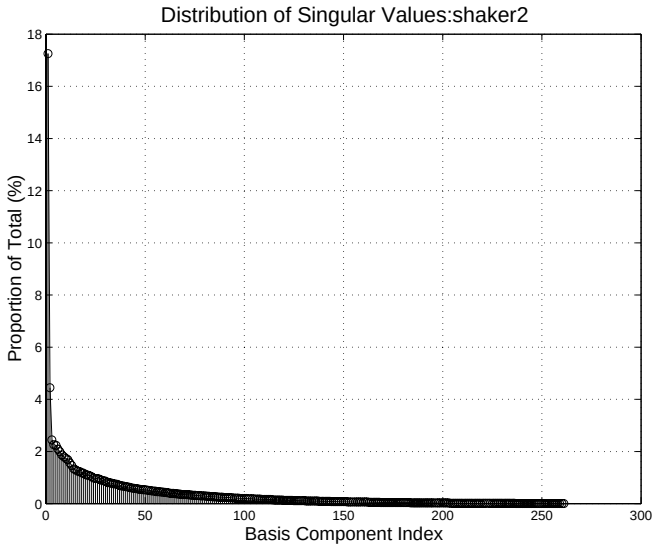

FIGURE 10. The singular values of an SVD of the percussive shaker sound. The first 19 component account for approximately 50% of the total variance in the signal.

occurring at both the up-shake and down-shake of the percussive action. This results in an anacrusis before each main shake. We would like a basis decomposition to reveal this elementary structure using very few components.

From an acoustical perspective the magnitude of the broad-band regions corresponds to the force of the shake. The shaker comprises many small particles which impact the surface of the shell creating a ramped impact and decay which has Gaussian characteristics.

3.3.10 Method

The shaker TFD was treated in spectral orientation which is the transpose of the spectrogram representation show in the figure. A full SVD decomposition was performed on the STFTM for 1.5 seconds of the sound.

3.3.11 Results

The singular values of the SVD are shown in Figure 10. The first three singular vectors decay rapidly from accounting for 17% of the total variance in the signal to accounting for approximately 2.4% of the total variance in the signal. Since the first three components have a much steeper decay than the rest of the components they are considered to hold the most important characteristic information in the TFD.

The first three left and right singular vectors are shown in Figure 11. The third singular vectors demonstrate the intended structure of an anacrusis followed by a down-beat for each shake. The left temporal vectors articulate this structure clearly. The right spectral vectors reveal the broad-band nature of the shake impacts. The remaining singular vectors account for the temporal pattern-

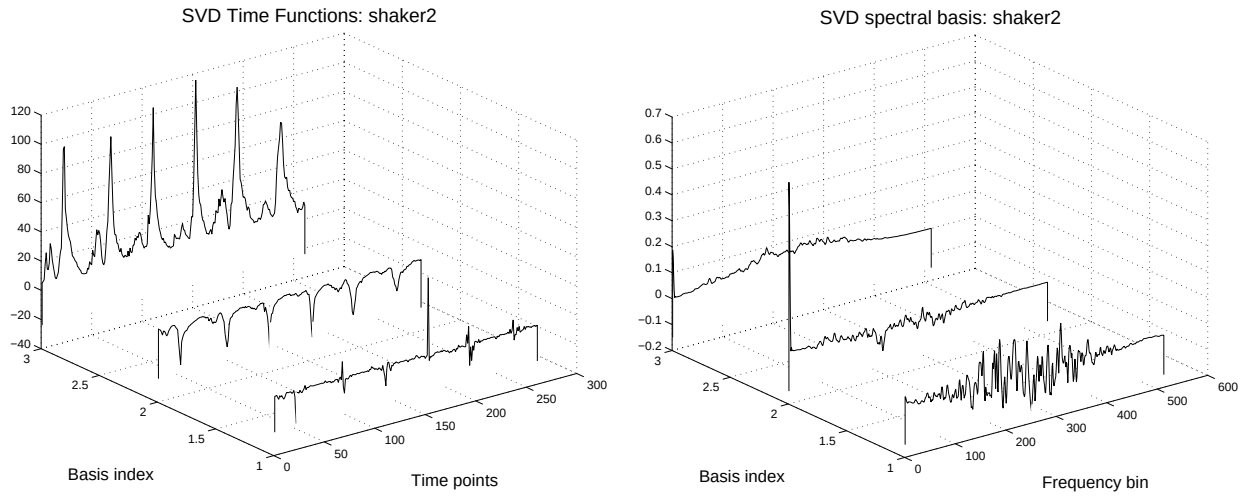

FIGURE 11. First three left and right singular vectors of the shaker sound. The left singular vectors correspond to a time-function for each of the right-singular vector spectral basis components. The outer-product of each pair of basis components forms an independent spectral presentation for that basis

ing and broad-band excitation of the particulate components of the shaker sound as well as the spectral structure of the metallic shell which is exhibited by the narrow-band spectral structure of the first right singular vector. From these results we conclude that the SVD has done a remarkably efficient job of representing the structure of the shaker sound.

3.3.12 A More Complicated Example: Glass Smash

Figure 12 shows 1.00 second of a glass smash sound. We can see from the figure that a number of discernible features are visible in this spectral representation; namely a low-frequency decaying impact noise component, a wide-band onset component and a series of high-frequency scattered particulate components which correspond to the broken glass shards. Ultimately, we would like a basis decomposition to represent these elements as separate basis functions with independent temporal characteristics.

From an ecological acoustics perspective, the bandwidth of the onset click, and the rate of decay of the low-frequency impact noise as well as the number of high-frequency particles serves to specify the nature of the event. In this case the glass-smash is relatively violent given the density of particles and the bandwidth and decay-times of the noise components.

From a signal perspective it is reasonable to treat this sound as a sum of independent noisy components since the individual particles corresponding to sharding are generated by numerous independent impacts. Each particle, however, contains formant structures as is indicated by the wide-band synchrony of onsets in the particulate scattering. This synchrony is manifest as correlations in the

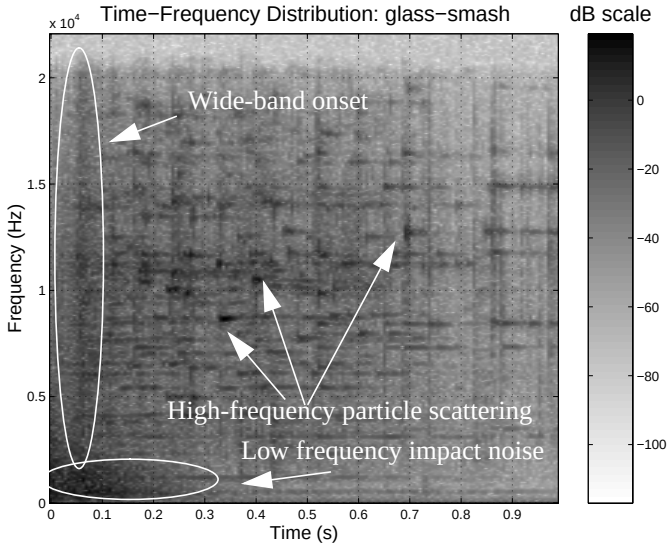

FIGURE 12. Short-time Fourier transform TFD of a glass-smash sound. Analysis parameters are N=1024, W=512, H=256. Sample rate is 44.1kHz.

underlying pdfs of the marginal distributions of the glass-smash TFD. From these observations it seems reasonable that an SVD might reveal quite a lot about the structure assuming that the nature of the statistical independence of spectral components is roughly Gaussian. For such a noisy sequence this assumption seems like a reasonable first-order approximation.

3.3.13 Method

The data matrix $\mathbf{x}$ is first decomposed using a full SVD as described in Section 3.3.7. This yields a set of orthonormal basis functions for both the row space and column space of $\mathbf{x}$ as well as a diagonal matrix Σ of singular values. In this example we chose to represent the matrix in spectral orientation which is essentially the transpose of the spectrogram orientation shown in Figure 12.

3.3.14 Results

The singular values for the glass smash sound are distributed across much of the basis thus suggesting a relatively noisy spectrum in terms of the distribution of Gaussian variances in the orthogonal basis space, see Figure 13. The left and right singular vectors of the spectrally-oriented SVD are given in Figure 8. The 5th left singular basis vector shows a pattern of decaying amplitude through time which corresponds to the low-pass spectral-basis component of the 5th right singular vector.

Other discernible features in the left singular vectors are the time patterns of the glass shards, bases 1-4, which are iterated with a decaying envelope through time. The narrow-band nature of the

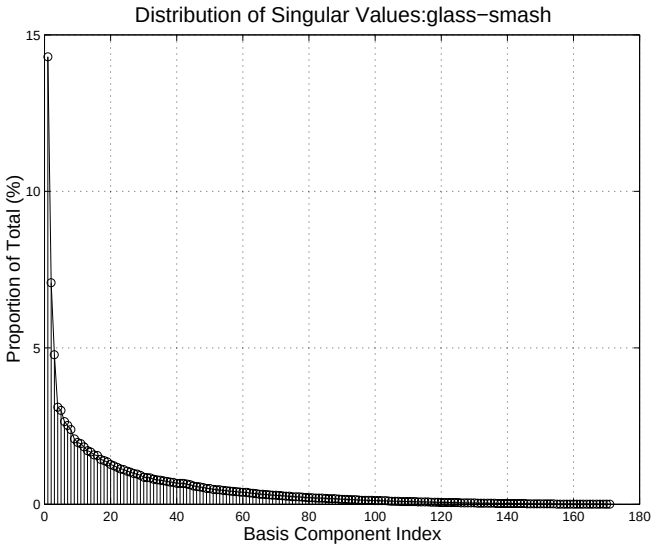

FIGURE 13. Distribution of singular values for the glass-smash sound. The first 14 singular values account for approximately 50% of the variance of the original signal.

peaks in the first 4 right singular vectors suggest high-Q filter characteristics which are due to the ringing of glass particles in the spectrum.

It was our goal in applying the SVD to the glass-smash sound to reveal elements of the complex structure in the noisy TFD shown in Figure 12. The coarse structure of the sound is indeed revealed by the decomposition but it does not appear that the signal has been characterized as successfully as the shaker example. We now discuss possible causes for inadequacy in a PCA of a TFD using the SVD.

3.3.15 Limitations of the Singular Value Decomposition

As we have discussed previously, an SVD decorrelates the input covariance by factoring the marginals of the second order statistics. This has the effect of rotating the basis space onto the directions that look most Gaussian. Whilst this assumption is valid for TFDs whose independent components comprise Gaussian-distributed magnitudes we conjecture that this assumption is too limiting for the case of most sounds. Speech and music sounds have been shown to have probability density functions which are non-Gaussian, therefore their PDFs are characterized by cumulants above second order, see [Bell&Sejnowski86] [Sejnowski88].

As an illustration of this point consider the scatter plot shown in Figure 14. The input distribution is a 2-dimensional uniform random variable which is evenly distributed in both dimensions. An SVD produces a basis which generates the basis rotation shown by the scatter plot in Figure 15. The SVD essentially creates a basis for the most Gaussian directions in the data without sensitivity to alternate distributions.

Thus we seek an alternate formulation of the SVD which is sensitive to higher-order statistical measures on the input data. We interpret this as necessitating a dependency on cumulants at higher than second order. The hypothesis is that such a decomposition will enable a more accurate statistically independent decomposition of data that is not Gaussian distributed.

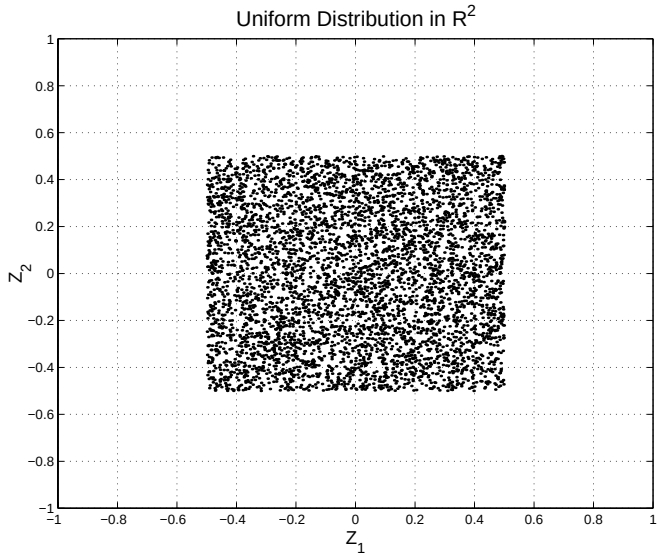

FIGURE 14. Scatter plot of a uniformly distributed random variable Z in 2-space.

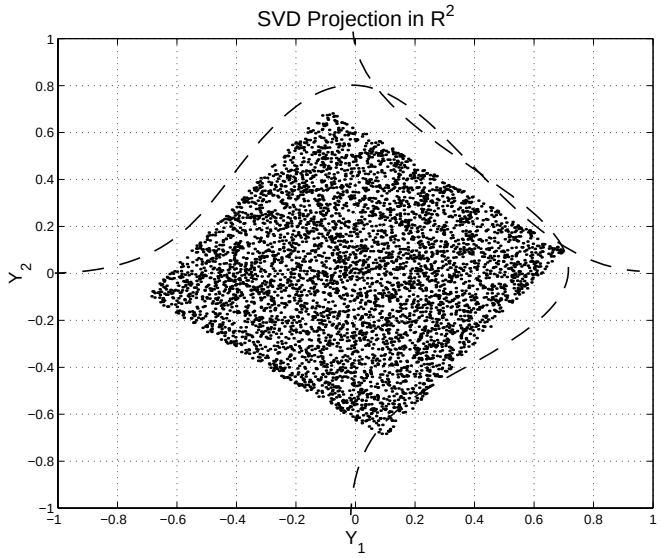

FIGURE 15. Scatter plot of SVD transformation of uniformly-distributed random variable Z. The SVD rotates the basis into the most Gaussian-like directions shown by the dashed lines. Clearly, this basis is not the best possible characterization of the input space.

3.3.16 Independent Component Analysis (ICA)

The concept of an ICA was first proposed in 1983 by Jutten and Herault who produced an iterative on-line algorithm, based on a neuro-mimetic architecture, for blind signal separation, see Jutten and Herault (1991). Their algorithmic solution to the problem of separating an unknown mixture of signals became the basis for a a number of different investigations into the application of statistical methods for identifying independent components within a data set. The blind source separation problem is related to the ICA problem by the need to identify statistically independent components within the data. For blind signal separation (BSS) the independent components correspond to *a-priori* unknown signals in a linear mixture,

Giannakis et al. (1989) used third-order cumulants to address the issue of identifiability of ICA. The resulting algorithm required an exhaustive search and is thus intractable for practical applications. Other mathematically-based techniques for identifying ICA were proposed by Lacoume and Ruiz (1989), who also used a cumulant-based method, and Gaeta and Lacoume (1990) proposed a maximum likely hood approach to the problem of blind identification of sources without prior knowledge.

An alternative method of investigating the existence of ICA was the method of Cardoso (1989) who considered the algebraic properties of fourth-order cumulants. Cardoso's algebraic methods involve diagonalization of cumulant tensors, the results of which are an ICA. Inouye and Matsui (1989) proposed a solution for the separation of two unknown sources and Comon (1989) proposed a solution for possibly more than two sources. These investigations form the mathematical foundations on which independent component analysis has continued to grow.

Recently many neural-network architectures have been proposed for solving the ICA problem. Using Comon's identification of information-theoretic quantities as a criteria for ICA Bell and Sejnowski (1996) proposed a neural network that used mutual information as a cost function. The resulting architectures were able to identify independently distributed components whose density functions were uni-modal. Bell's network maximizes the mutual information between the input and output of the neural network which has the effect of minimizing the redundancy. Amari *et al.* (1996) proposed a using a different gradient descent technique than Bell which they called the *natural gradient.* These, and many other, neural network-based architectures were proposed as partial solutions to the problem of blind signal separation.

Aside from the BSS problem of additive mixtures, several architectures have been proposed for addressing the problem of convolved mixtures of signals. Among these are architectures that employ feedback weights in their neural network architectures in order to account for convolution operations. A novel approach to the problem of convolutions of signals was proposed by Smaragdis (1997), in which the convolution problem is treated as a product of spectral components thus the architecture seeks to factor the spectral components into independent elements.

All of the techniques and architectures introduced above have been applied to the problem of separation of sources in one form or another. An alternate view of ICA is that it is closely related to PCA. This is the view that we take in this section. We develop the necessary mathematical background in order to derive an algorithm which is capable of producing a *basis* in which spectral components are lifted into independent distributions. Our methods are closely related to the algebraic methods of Comon and Cardoso and are seen as a higher-order statistical extension of the SVD.

3.3.17 The ICA Signal Model: Superposition of Outer-Product TFDs

For the purposes of feature extraction from a TFD using an ICA we must be explicit about our signal assumptions. Our first assumption is that the input TFD is composed of a number of *a-priori* unknown, statistically independent TFDs which are superposed to yield the observable input TFD. This assumption of superposition is represented as:

$$\mathbf{X} = \sum_{i=1}^{\rho} \chi_i + \sum_{j=1}^{\kappa} \Upsilon_j \qquad [97]$$

where χ_i are the latent independent TFDs of which there are ρ, and the Υ_j are an unknown set of noise TFDs of which there are κ. Observing that the superposition of TFDs is a linear operation in the time-frequency plane and under the assumption that the inverse TFD yields the corresponding latent superposition of signals then Equation 97 is interpreted as the frequency-domain representation of a blind signal separation problem. In this form the signal model defines the domain of signal compositions that we are operating under but it does nothing to define the form of the features that we might want to extract as characteristic components of the signals.

A second, stronger assumption is that each independent TFDs χ_i is uniquely composed from the outer product of an *invariant* basis function $\mathbf{y}_i$ and a corresponding *invariant* weighting function $\mathbf{v}_i$ such that:

$$\chi_i = \mathbf{y}_i \mathbf{v}_i^T. \qquad [98]$$

These functions are invariant because they are statistically stationary vectors which multiply, using the outer-product of two vectors, to form a TFD matrix. Under the assumption of the outer-product form of the TFD the vectors are defined to be stationary since there is no way to affect a time-varying transform.

This latter assumption seems on the surface quite limiting. After all many natural sounds are composed of non stationary spectral components which may shift in frequency during the course of the sound. However, recalling our framework from the previous chapter, the utility of a statistical basis decomposition comes not from the ability to fully characterize the transformational structure of a

sound, but it is in its ability to identify likely candidates to be treated as *invariants* for a sound structure. These invariants are to be subjected to further treatment in the next chapter in which we use them to identify the time-varying structure of a sound. We recall the observation of Warren and Shaw (1985) that structure must be defined as a form of persistence and a style of change. The statistical decomposition of TFDs provides much of this structure in the form of spectral invariants and temporal fluctuations, but time-varying frequency components are not represented by the techniques. We must define time-varying frequencies in terms of a form of persistence and it is this form that we seek to identify.

We conjecture that the time-varying components of a natural sound are constrained in their range between each time frame of a TFD, thus the change in an invariant is relatively small at each time frame. Recall from our discussion on auditory group theory that such small changes in an invariant component can be used to identify the form of a transformation. The basis techniques on which we rely for extraction of the invariant features are dependent upon the PDFs of the invariant components. Thus, under the assumption of small changes between frames, it is assumed that each PDF is stationary enough over a portion of the TFD that it is well represented by the ensemble statistics of the TFD.

However, the argument is a little more subtle than this. Since the statistical techniques presented herein are batch techniques, operating on the entire data matrix with no update rule, there is actually no dependence upon the order of the frames in a TFD. Thus we would get equivalent results if we presented the frames in a random order. So it is the time-average PDF of an independent spectral component that determines the form of an invariant. For example, if the component oscillates about a mean frequency such that the average density of the centre frequency is greater than the density of the peak frequency deviations then the distribution of the average PDF will be representative of the underlying invariant.

These arguments lead us to our third assumption for the ICA signal model: that the underlying invariant functions of the independent TFDs are distributed in time-frequency in such a way that their average PDF is, in fact, representative of an invariant component. In the case that they are not centered on a mean frequency value we observe that the statistics will yield a series of similar TFD basis components that differ by the nature of the underlying transformation. Since the basis decomposition techniques order the basis components by their singular values, i.e. their salience in the input TFD, we take the components that have larger singular values as being representative of the invariants that we seek. It is extremely unlikely that a single time-varying component will yield very high singular values for each of its mean spectra in the statistical decomposition. This leads us to assert that the decompositions are valid representatives of the underlying TFDs but care must be taken in interpreting the results.

By representing the basis components $\mathbf{y}_i$ and $\mathbf{v}_i$ as the columns of two matrices we arrive at an assumed signal model for the input TFD:

$$\mathbf{X} = \mathbf{Y}_\rho \mathbf{V}_\rho^{\mathrm{T}} + \Upsilon' \qquad [99]$$

where $\Upsilon = \sum_{j=1}^{\kappa} \Upsilon_j$ is the summed noise matrix. The components $\mathbf{Y}_\rho$ and $\mathbf{V}_\rho$ both have ρ columns. Thus for an $m \times n$ input TFD $\mathbf{X}$, $\mathbf{Y}_\rho$ is an $m \times \rho$ matrix and $\mathbf{V}_\rho$ is an $n \times \rho$ matrix. We call this model a superposition of outer-product TFDs and it defines the structure of the features that we seek in a statistical basis decomposition of a given input TFD.

3.3.18 ICA: A Higher-Order SVD

For our formulation of ICA we start with an SVD of a TFD data matrix:

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}}. \qquad [100]$$

The statistically independent basis that we seek is an orthogonal set of vectors that span the column space $\Re(\mathbf{X})$, thus the SVD factorization is a good starting point since $\mathbf{V}^{\mathrm{T}}$ already spans this space, but under a rotation of basis on the assumption of Gaussian input statistics.

We would like the ICA to be an orthogonal basis like that of the SVD but we impose different constraints corresponding to maximization of higher-order cumulants in the PDFs of the data space. We define the matrix $\mathbf{Z} = \mathbf{V}^{\mathrm{T}}$ which is the matrix of random vectors whose PDFs we seek to factor. Now, the random vector $\hat{\mathbf{z}} \in \mathbf{Z}$ has statistically independent components if and only if:

$$p_{\mathbf{Z}}(\hat{\mathbf{z}}) = \prod_{i=1}^{N} p_{\mathbf{Z}_i}(\hat{\mathbf{z}}_i). \qquad [101]$$

where N is the dimensionality of $\hat{\mathbf{z}}$. Thus we seek to factor the joint probability density function of $\hat{\mathbf{z}}$ into the product of its marginal PDFs in order to achieve statistically independent basis components.

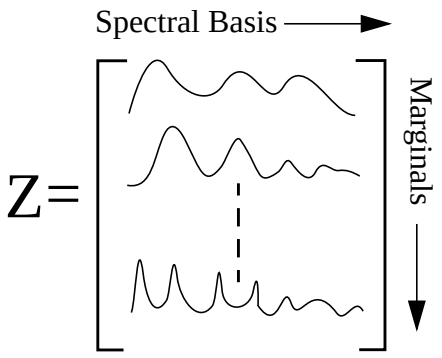

FIGURE 16. Illustration of the orientation of spectral basis components in the ICA decomposition of the variable Z. The marginals are the orthogonal complement of the basis functions.

As an illustration of the effect of the factorization let us consider $\mathbf{X}$ as a TFD in spectral orientation. The column space of $\mathbf{X}$ is occupied by the frequency-bins of the chosen time-frequency transform with each row of $\mathbf{X}$ an observation or time-slice of the TFD. The matrix $\mathbf{V}$ is a basis for the column space of $\mathbf{X}$ with each column of $\mathbf{V}$ corresponding to a spectral basis component. The matrix $\mathbf{Z}$, then, contains the spectral basis functions row-wise as shown in Figure 16. Now, the random vector $\hat{\mathbf{z}}$ has a joint PDF which operates down the columns of the matrix $\mathbf{Z}$. A successful factorization of the joint probability density function of $\mathbf{Z}$ will therefore result in statistically independent rows of $\mathbf{Z}$ which corresponds to a statistically-independent spectral basis for the spectrally-oriented TFD.

We could equally decide to represent the TFD in temporal orientation. The column space of $\mathbf{X}$ would thus have variates corresponding to the temporal amplitude functions of a TFD, each weighting an underlying spectral basis component which, in the case of temporal orientation, is represented by the left singular vectors which are the columns of $\mathbf{U}$. A successful factorization would result in a set of statistically-independent amplitude basis functions. The orientation that we choose may have a significant effect on the resulting characterization of a TFD. We explore issues of data matrix orientation later in this chapter.

For the defined matrix $\mathbf{Z}$, we can effect the factorization shown in Equation 101 by a rotation of the basis $\mathbf{V}$. This corresponds to rotating the rows of $\mathbf{Z}$ such that they point in characteristic directions that are as statistically independent as possible based on a criteria which we shall soon define. Thus the ICA can be thought of as an SVD with a linear transform performed by a new matrix $\mathbf{Q}$ such that:

$$\mathbf{Z}_{\mathrm{ICA}} = \mathbf{Q}\mathbf{Z}_{\mathrm{SVD}} = \mathbf{Q}\mathbf{V}^{\mathrm{T}}. \qquad [102]$$

3.3.19 Information-Theoretic Criteria For ICA

Having defined the form of the ICA we now seek to define a criteria for statistical independence that will yield the factorization of Equation 101. In order to do this we must define a distance metric δ between the joint-probability density function $p_{\mathbf{z}}(\hat{\mathbf{z}})$ and the product of its marginals:

$$\delta\left(p_{\mathbf{z}}(\hat{\mathbf{z}}), \prod_{i=1}^{N} p_{\mathbf{z}_i}(\hat{\mathbf{z}}_i) \right) \quad [103]$$

In statistics, the class of *f*-divergences provides a number of different measures on which to base such a metric. The Kullback-Leibler divergence is one such measure and is defined as:

$$\delta(p_x, p_z) = \int p_x(u) \left(\log \frac{p_x(u)}{p_z(u)} \right) du . \quad [104]$$

Substituting Equation 103 into Equation 104 yields:

$$I(p_{\mathbf{z}}) = \int p_{\mathbf{z}}(\hat{\mathbf{z}}) \left(\log \frac{p_{\mathbf{z}}(\hat{\mathbf{z}})}{\prod_{i=1}^{N} p_{\mathbf{z}_i}(\hat{\mathbf{z}}_i)} \right) d\hat{\mathbf{z}} \quad [105]$$

where $I(p_{\mathbf{z}})$ is the average mutual information of the components of $\mathbf{z}$. The Kullback-Leibler divergence satisfies the relation:

$$\delta(p_x, p_z) \geq 0 \quad [106]$$

with equality if and only if $p_x(\mathbf{u}) = p_Z(\mathbf{u})$, Comon (1994). Thus, from Equation 103, the average mutual information between the marginals $\hat{\mathbf{z}}_i$ becomes 0 if and only if they are independent, which implies that information of a marginal does not contribute to the information of any other marginal in the joint PDF of $\mathbf{z}$.

3.3.20 Estimation of the PDFs

Having defined a suitable criteria for ICA we must now tackle the problem of estimation of the PDF of $\mathbf{z}$ since the densities are not known. We do, however, have data from which to estimate the underlying PDFs.

The Edgeworth expansion of a density $\mathbf{z}$ about its best Gaussian approximate ϕ_z for zero-mean and unit variance is given by:

$$\frac{p_z(u)}{\phi_z(u)} = 1 + \frac{1}{3!}k_3h_3(u) + \frac{1}{4!}k_4h_4(u) + \frac{10}{6!}k_3^2h_6(u) + \frac{1}{5!}k_5h_5(u) + \frac{35}{7!}k_3k_4h_7(u) \quad \textbf{[107]}$$
$$+ \frac{280}{9!}k_3^3h_9(u) + \frac{1}{6!}k_6h_6(u) + \frac{56}{8!}k_3k_5h_8(u) + \frac{35}{8!}k_4^2h_8(u) + \frac{2100}{10!}k_3^2k_4h_{10}(u)$$
$$+ \frac{15400}{12!}k_3^4h_{12}(u) + o(m^{-2})$$

where k_i denotes the cumulant of order i of the scalar variable u and $h_i(u)$ is the Hermite polynomial of degree i defined by the recursion:

$$h_0(u) = 1, \qquad h_1(u) = u \qquad . \quad \textbf{[108]}$$
$$h_{k+1}(u) = uh_k(u) - \frac{\partial}{\partial u}h_k(u)$$

With a method for estimating the PDF from an ensemble of data we are able to proceed with parameterizing the linear transform $\mathbf{Q}$ so that the ICA basis vectors in $\mathbf{Z}$ satisfies our independence criteria as closely as possible.

3.3.21 Parameterization and Solution of the Unitary Transform Q

In order to obtain the rotation matrix $\mathbf{Q}$ a parameterization in terms of the Kullback-Leibler divergence on $\mathbf{z}$ is utilized.

With the solution of $\mathbf{Q}$ we arrive at a form for the ICA transform:

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{Q}^T\mathbf{Q}\mathbf{V}^T \quad \textbf{[109]}$$

since $\mathbf{Q}$ is unitary, the quantity $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$, thus rotations of the basis components do not affect the reconstruction of $\mathbf{X}$.

3.3.22 Uniqueness Constraints

The formulation of ICA in this manner does not specify the basis uniquely. In fact, it expresses an equivalence class of decompositions for which there are infinitely many possibilities. In order to define a unique ICA additional constraints must be imposed on the form of Equation 109.

Firstly, the decomposition is invariant to permutations of the columns. Thus the same criteria for ordering of basis components as the SVD is utilized; namely that the basis components are permuted in decreasing order of their variances. We denote by $\mathbf{P}$ the permutation matrix that performs this ordering. Permutation matrices are always invertible and they have the property $\mathbf{P}^T\mathbf{P} = \mathrm{I}$. The second criteria for uniqueness stems from the fact that statistics are invariant under scaling. That is, the PDF of a scaled random vector is the same as the unscaled vector's PDF. A scaling is chosen such that the columns of $\mathbf{V}$ have unit norm. We denote by Λ the invertible diagonal matrix of scaling coefficients. Finally an ICA is invariant to sign changes in the basis components. The unique-

ness constraint is chosen such that the sign of the largest modulus is positive. We denote by $\mathbf{D}$ the diagonal matrix comprising values from [1,-1] which performs this sign change. As with the other uniqueness constraint matrices, the sign-change matrix is trivially invertible.

These uniqueness constraints give the final form of the ICA:

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{Q}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\Lambda^{-1}\mathbf{D}\mathbf{D}^{\mathrm{T}}\Lambda\mathbf{P}\mathbf{Q}\mathbf{V}^{\mathrm{T}} \quad \textbf{[110]}$$

with

$$\mathbf{Z} = \mathbf{D}^{\mathrm{T}}\Lambda\mathbf{P}\mathbf{Q}\mathbf{V}^{\mathrm{T}} \quad \textbf{[111]}$$

and

$$\mathbf{Y} = \mathbf{U}\Sigma\mathbf{Q}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\Lambda^{-1}\mathbf{D}. \quad \textbf{[112]}$$

Since $\mathbf{X}$ and $\mathbf{Z}$ are both given we can compute the left basis by a projection of the data against the right basis vectors:

$$\mathbf{Y} = \mathbf{X}\mathbf{Z}^{\mathrm{T}} = \mathbf{V}\mathbf{Q}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\Lambda^{\mathrm{T}}\mathbf{D}. \quad \textbf{[113]}$$

The outputs covariance Φ_{Y} is diagonalized by the unitary transform $\mathbf{Q}$ but, unlike the SVD, this diagonalization is based on the contrast defined by fourth-order cumulants.

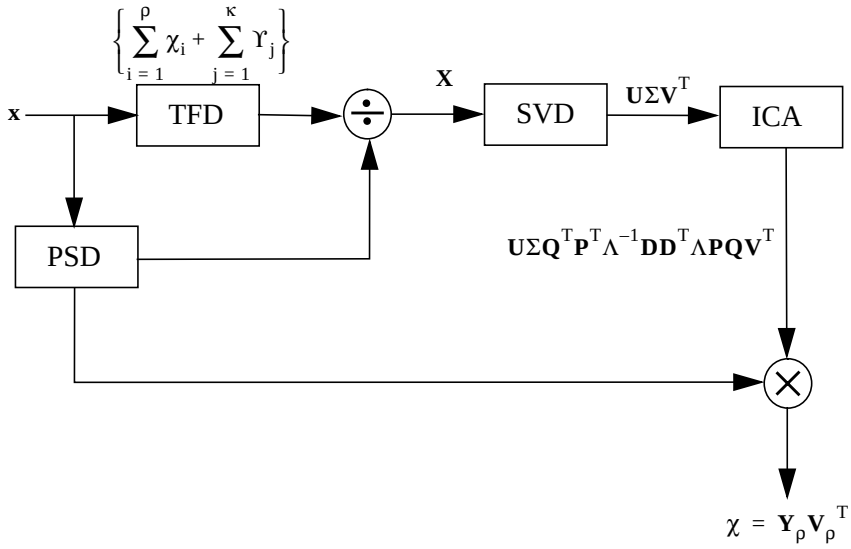

FIGURE 17. Signal flow diagram of independent component analysis of a time-frequency distribution. The input is a signal $\mathbf{x}$ whose latent variables χ_i we seek. An optional power-spectral density normalization is applied followed by the SVD and ICA transforms. The output is a set of ρ basis vectors which span the signal space of the TFD of $\mathbf{x}$.

3.4 Independent Component Analysis of Time-Frequency Distributions

So far we have discussed the mechanics of the ICA and SVD algorithms in relative isolation from their application to sound analysis. We are now in a position to discuss the general application of independent component analysis to time-frequency distributions. In this section we investigate methods for utilizing an ICA for identifying features in the TFD of a signal. The feature extraction problem involves the estimation of many unknown variables in the signal. As we shall see, using the signal assumptions defined above in conjunction with a careful application of an ICA we are able to obtain results that appear to meet our demands to a remarkably high degree.

3.4.1 Method

The method proceeds as follows. An input signal $\mathbf{x}$ is assumed to contain ρ independent components that are combined under the signal model of Equation 97. All of ρ, χ_i and Υ_j are assumed unknown *a-priori*, Figure 17.

A time-frequency transform produces a TFD which expresses the signal model in the time-frequency plane. For many natural sounds, the power of the signal in increasing energy bands may decrease rapidly due to the predominance of low-frequency energy. Thus, from the point of view of statistical measures on the data, the variates are scaled by the average spectrum of the TFD. In order to compensate for the power-loss effect at high frequencies, and to *sphere* the data to a reasonable scale in all variates, an optional power-spectral density estimation and normalization step is incorporated.

The power spectral density of a TFD is calculated using Welch's averaged periodogram method. The data sequence $x[n]$ is divided into segments of length L with a windowing function $w[n]$ applied to each segment. The periodogram segments form a separate TFD which can be estimated from the analysis TFD in the case of an STFT. In the most general case, however, this may not be possible so we represent PSD normalization as a separate path in the signal flow diagram of Figure 17.

The periodogram or the l th segment is defined as:

$$I_l(\omega) = \frac{1}{LU}\left|X_l(e^{jw})\right|^2 \qquad \textbf{[114]}$$

where L is the length of a segment and U is a constant that removes bias in the spectral estimate and $X_l(e^{jw})$ is a short-time Fourier transform frame as described previously. The average periodogram for a signal $x[n]$ is then the time-average of these periodogram frames. If there are K frames in the periodogram then the average periodogram is:

$$\bar{I}_l(\omega) = \frac{1}{K}\sum_{l=0}^{K-1} I_l(\omega). \qquad \textbf{[115]}$$

Thus the average periodogram provides an estimate of the power-spectral density (PSD) of $x[n]$. Assuming that the PSD is nowhere equal to zero we can perform the normalization of the TFD by division in the frequency domain as indicated in the figure. Once the data matrix $\mathbf{x}$ is obtained an SVD is performed which yields a factorization of the data matrix into left and right basis vectors and a matrix of corresponding singular values, see Equation 93.

From the singular values Σ, the rank ρ of the TFD can be estimated. In order to do this we pick a criteria $\Psi \in [0 \ldots 1]$ that specifies the amount of total variance in the TFD that we wish to account for in the resulting basis. In the case of data compaction applications Ψ is chosen relatively high, typically around 0.95, so that the reconstruction from the reduced basis results in as little loss as possible. However, for the purposes of feature extraction we can choose Ψ much lower since we seek to characterize the primary features in the data space rather than account for all the variance. Typical values for Ψ were determined empirically to be in the range $0.2 \le \Psi \le 0.5$. Given this variance criteria, estimation of the rank ρ of $\mathbf{x}$ is achieved by solving the following inequality for ρ:

$$\frac{1}{\text{trace}(\Sigma^2)} \sum_{i=1}^{\rho} \Sigma^2(i, i) \geq \Psi . \qquad [116]$$

This estimate of the rank of the data matrix provides a good approximation of the number of statistically independent components in the TFD. Thus the following ICA problem can be reduced from the problem of generating a full set of independent columns in the basis space to that of generating exactly ρ independent columns. Since the singular vectors of the SVD are sorted according to their singular values in decreasing order of importance, we choose the first ρ columns of $\mathbf{V}$ for the estimation and solution of the ICA.

Thus the first ρ right singular vectors of the SVD are used to obtain a basis with an ICA, the vectors are transposed and stored in a matrix $\mathbf{Z}$ which is the observation matrix for the ICA decomposition. An iterative procedure is employed which first estimates the cumulants for each pair of rows $(\hat{\mathbf{z}}_i, \hat{\mathbf{z}}_j)$ in $\mathbf{Z}$; of which there are $\rho\frac{(\rho-1)}{2}$ pairs. From these cumulants the angle α that minimizes the average mutual information $I(p_{\mathbf{z}})$, defined in Equation 105, is calculated such that the unitary transform $\mathbf{Q}^{(i, j)}$ performs a rotation about the angle α in the orthogonal plane of $(\hat{\mathbf{z}}_i, \hat{\mathbf{z}}_j)$. It can be shown that a set of planar rotations, derived from estimates of α, that maximize the pairwise independence (i.e. minimize the average mutual information) of the rows of $\mathbf{Z}$ are a sufficient criteria for independence. That is, pair-wise independence specifies global independence. For a proof of this conjecture see Comon (1994). After each iteration, $\mathbf{Z}$ is updated by applying the unitary transform $\mathbf{Z} = \mathbf{Q}^{(i, j)}\mathbf{Z}$. The iterations continue on the set of row pairs in $\mathbf{Z}$ until the estimated angles α become very small or until the number of iterations k has reached $1 + \sqrt{\rho}$.

After these operations have been performed, $\mathbf{Z}$ contains a set of ρ basis components in the rows which are as statistically independent as possible given the contrast criteria of maximization of fourth-order cumulants in $\mathbf{Z}$. As discussed previously, these components are not unique for the statistics are invariant to scaling, ordering and sign changes in the moduli of the vector norms. Applying uniqueness constraints we first compute the norm of the columns of $\mathbf{V} = \mathbf{Z}^{\mathrm{T}}$ which specify the entries of the diagonal scaling matrix Λ. In order to solve for the ordering the entries of Λ are sorted in decreasing order; this specifies the permutation matrix $\mathbf{P}$ whose rows generate the said ordering of entries in Λ. Finally a diagonal matrix of entries with unit modulus an possibly different signs is constructed such that the entry of the largest modulus in each column of $\mathbf{Z}$ is positive real; this specifies the matrix $\mathbf{D}$.

With the full specification of the ICA in hand we can compute a basis for the row-space of $\mathbf{X}$ using the relation:

$$\mathbf{Y}_\rho = \mathbf{X}\mathbf{Z}_\rho^{\mathrm{T}} = \mathbf{V}_\rho\mathbf{Q}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\Lambda^{\mathrm{T}}\mathbf{D} \qquad [117]$$

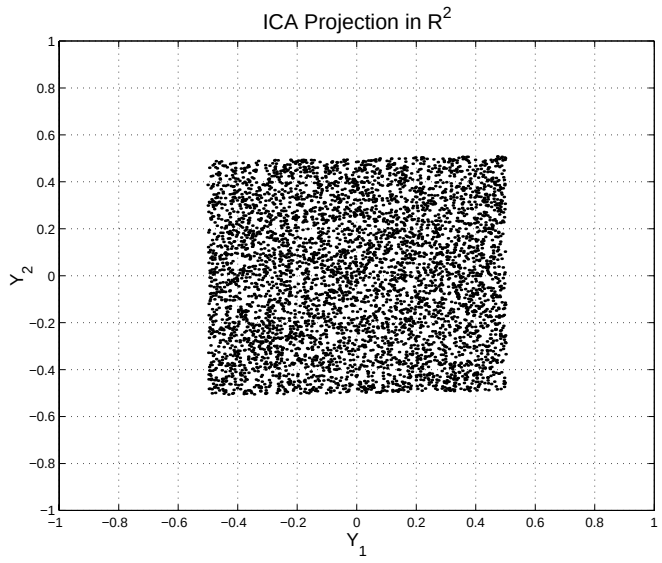

FIGURE 18. Scatter plot of the output of an ICA for the input of an arbitrary linear transformation of a bi-variate uniform distribution. The plot shows that the PDFs have been accurately characterized by the ICA since they have been separated correctly (compare with Figure 15 on page 97).

this operation is equivalent to performing the unitary transform and uniqueness operations on the first ρ left singular vectors of the preceding SVD scaled by their singular values:

$$\mathbf{Y}_\rho = \mathbf{U}_\rho \Sigma_\rho \mathbf{Q}^T \mathbf{P}^T \Lambda^{-1} \mathbf{D}. \quad \textbf{[118]}$$

With these bases in place we are able to specify the form of the latent independent TFDs which form the independent features of the original TFD:

$$\chi = \mathbf{Y}_\rho \mathbf{V}_\rho{}^T \quad \textbf{[119]}$$

thus each column $\mathbf{Y}_j$ and $\mathbf{V}_j$ specifies a basis vector pair for an independent TFD χ_i, and the independent χ_i 's sum to form the signal TFD of $\mathbf{X}$, which is an approximation $\hat{\mathbf{X}}$. The residual spectrum $\mathbf{X} - \hat{\mathbf{X}}$ specifies the near-uncorrelated noise components of $\mathbf{X}$ which is also obtainable by an ICA transform of the SVD basis components that were not used in the identification of χ:

$$\Upsilon = \mathbf{X} - \hat{\mathbf{X}} = \mathbf{Y}_{M-\rho} \mathbf{V}_{N-\rho}^T. \quad \textbf{[120]}$$

As a final comment on the ICA before we demonstrate its application to sound analysis we again consider the bi-variate uniform distribution of Figure 14. Recall that the SVD basis did not ade-

quately characterize the PDFs due to its Gaussian basis criteria, see Figure 15; an ICA transform of the SVD basis produces the basis rotation shown in Figure 18, which demonstrates that the ICA is capable of characterizing non-Gaussian distributions. In fact, the bi-variate uniform distribution is one of the most difficult joint-PDFs for an ICA to characterize and it bodes well for the algebraic approach that we were able to factor this example correctly, see (Bell and Sejnowski 1995; Amari *et al.* 1996).

3.5 Examples of Independent Component Analysis of TFDs

We now give some examples of the application of ICA to analysis of noisy and textured natural sounds. These sounds have traditionally been very difficult to characterize with sinusoidal-based analysis techniques such as the dual-spectrum representations considered earlier in this chapter. ICA characterizations are not limited to noisy spectra, however. A harmonic sound will also have a discernible PDF which can be separated from other components. In fact, PCA techniques have been successfully applied to harmonic spectra in previous research as outlined previously; see, for example Bayerbach and Nawab (1991). These studies have demonstrated the applicability of PCA techniques to sinusoidal tracking applications. In the following examples, therefore, we focus on the much harder problem of analysis and characterization of sounds with very noisy TFDs.

3.5.1 Example 1: Bonfire sound

1. Method

The first example is that of a bonfire; the spectrogram is shown in Figure 19. The discernible features in this sound are a low-pass continuous Gaussian noise component and a number of intermittent wide-band crackles. We would like the ICA to treat these as separable features of the TFD. An ICA analysis was applied to the bonfire sound with no PSD normalization since there was no frequency band in which energy was disproportionately high compared with the other bands.

2. Results

The singular values of the SVD of the bonfire sound are shown in Figure 20. There is a steep roll-off in the first three singular values followed by a steady exponential decay for the remainder of the components. The first three singular values account for 40% of the total variance in the bonfire signal. This is a very high quantity for just three components, so we would like to investigate the parts of the sound that they represent.

Figure 21 and Figure 22 show the SVD and ICA basis vectors for the bonfire sound respectively. The left singular vectors of the SVD decomposition correspond to amplitude functions through time of the TFD. These characterize the row-space of the TFD in spectral orientation. Each of the three components shown seem to exhibit both the intermittent crackling properties as well as the Gaussian noise sequence properties described above. However, they are not well separated into statistically-independent components. An inspection of the right SVD singular vectors similarly shows that the wide-band and low-pass components are mixed between the basis vectors. Thus we conclude that the SVD has not characterized the bonfire sound satisfactorily.

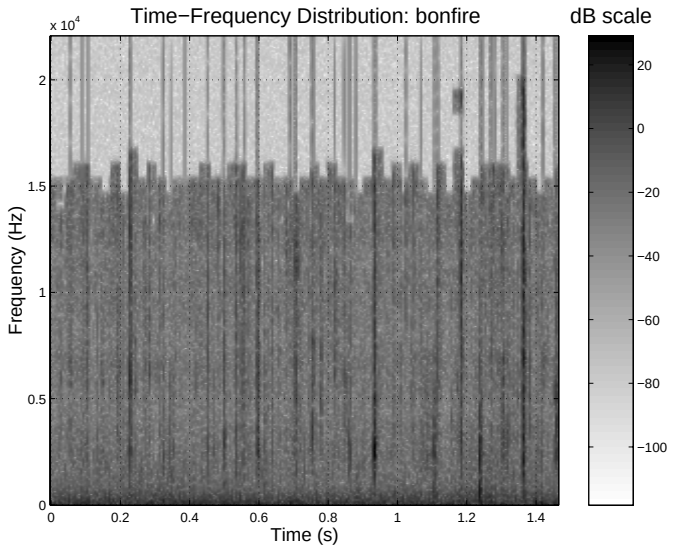

FIGURE 19. STFT spectrogram of bonfire sound. The sound contains intermittent wide-band click elements as well as low-pass and wide-band continuous noise elements.

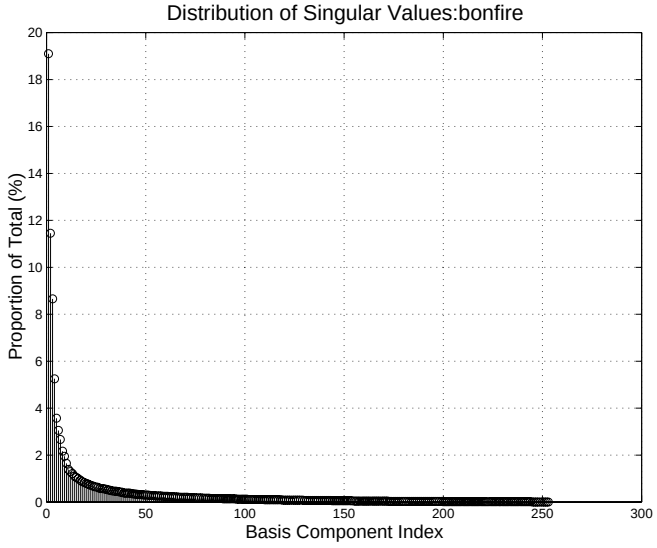

FIGURE 20. Singular values of bonfire sound. The first three singular values account for 40% of the total variance in the data. This implies that they are good candidates for features.

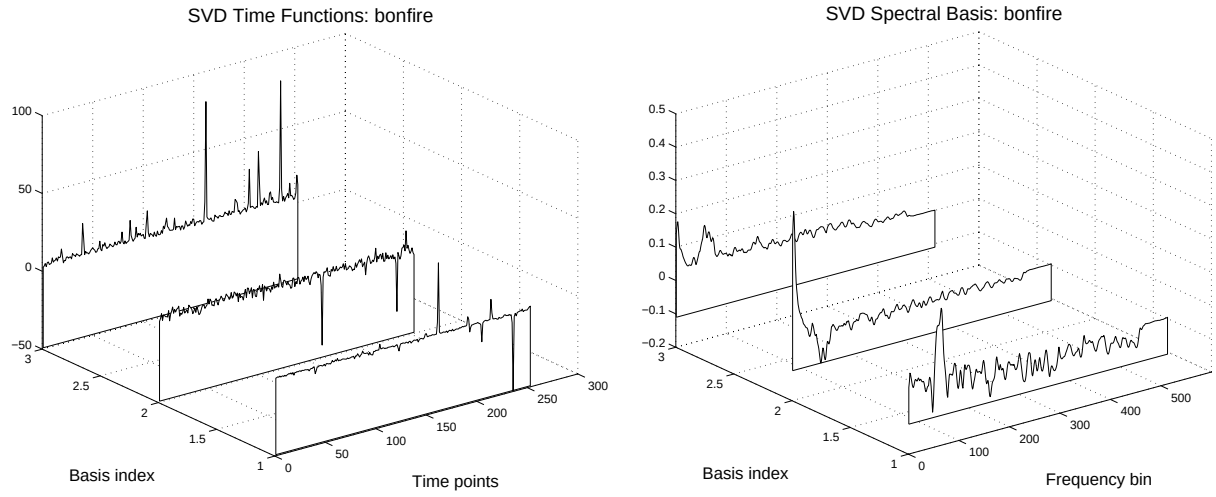

FIGURE 21. SVD basis vectors of a bonfire sound. The left singular vectors seem to mix both the continuous noise elements as well as the erratic impulses. The right singular vectors exhibit a similar mixing of spectral features with notches in some spectra occurring at the peaks of others.

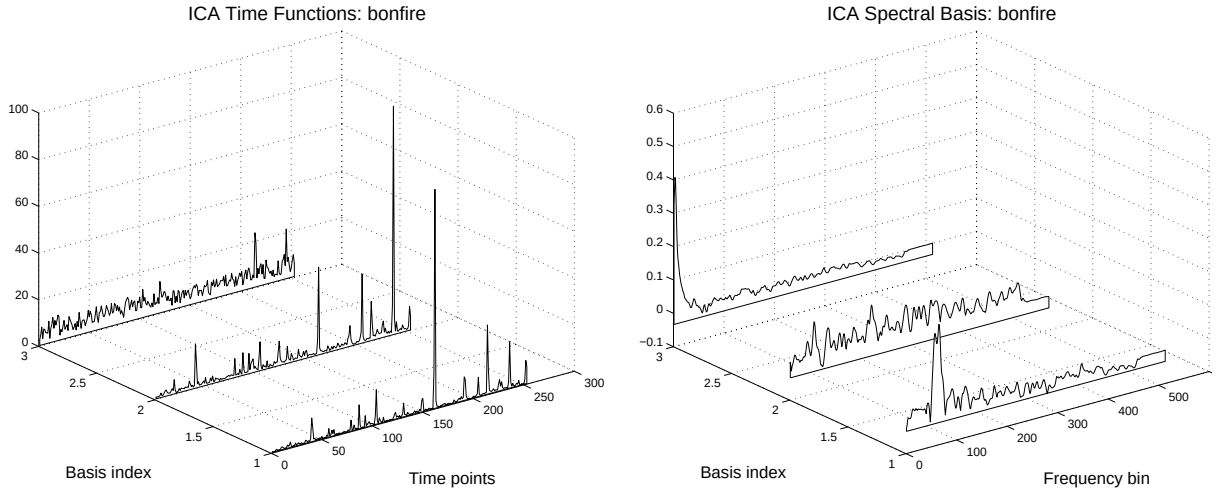

FIGURE 22. ICA basis vectors of the bonfire sound. The left singular vectors exhibit the desired characteristics of erratic impulses and continuous noise densities as independent components. The right singular vectors also exhibit independence with low-pass and wide-band components clearly distinguished.

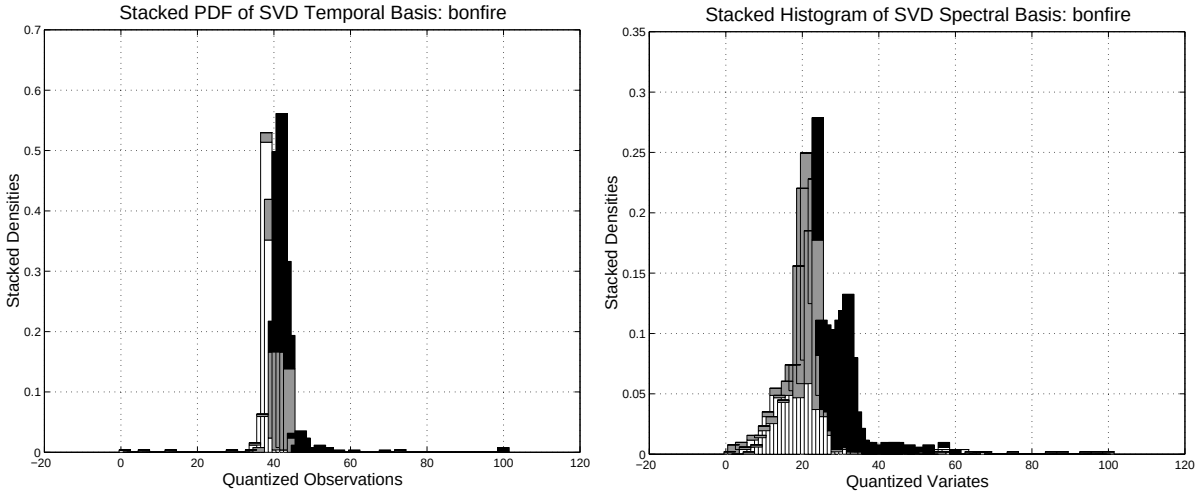

FIGURE 23. Stacked histograms of bonfire sound SVD basis vectors. The left singular vectors exhibit a clustering around the quantized 40. They are roughly evenly distributed each side implying an even distribution. The right singular vectors appear somewhat Gaussian and are centered at quantized values around 20 and 30.

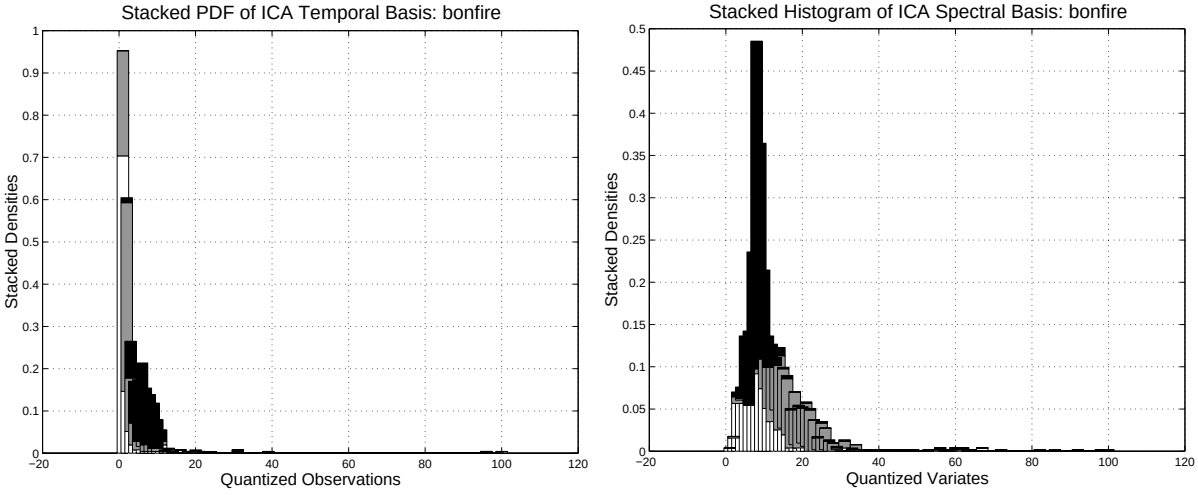

FIGURE 24. Stacked histograms of bonfire ICA basis vectors. The left singular vectors are skewed to the right-hand side resembling an exponential distribution. The right singular vectors are also skewed to the right. Several of the components are extremely peaky, suggesting a high kurtosis.

An inspection of the ICA basis functions, however, reveals a more promising characterization. The three left singular vectors show a clear distinction between continuous noise and intermittent crackling elements; the first two basis functions corresponding to the amplitude functions of the intermittent crackling and the third component corresponding to the amplitude function of the continuous noise component. Similarly, inspection of the right singular ICA vectors shows a clear separation between wide-band and low-pass components. The first two right basis functions correspond with the first two left functions and represent wide-band spectral components; namely, the spectral envelopes of the crackling components. The third right ICA basis vector shows a low-pass component, it corresponds with the continuous noise amplitude function of the third left basis vector and is the spectral envelope of the continuous noise component.

Figure 23 and Figure 24 show stacked histograms of the values of each set of basis vectors. The values were quantized into 150 bins and the histograms of each vector are stacked, in turn, on top of the histogram of its predecessor in order that they can be more easily inspected. The main difference between the SVD and the ICA histograms is that the SVD components are spread roughly evenly around a mean value, thus approximating the components with Gaussian-like PDFs. The ICA histograms, on the other hand, are skewed to one side, thus exhibiting underlying PDFs that are approximately exponential. The kurtosis (measure of fourth-order cumulants) of the PDFs of the ICA distribution is much higher than that of the SVD distribution, suggesting that the contrast function based on fourth-order cumulants is a successful method for contrasting a set of basis vectors against a joint-Gaussian PDF which has no cumulants above second order. Thus, if there exists higher-order cumulants in the estimated PDF of a given basis, the SVD will only find the best Gaussian approximation which, in the case of high kurtosis or skew in the actual PDF, will not be adequate for characterizing the basis components of a TFD.

Table 6 shows the values of the fourth-order cumulants for each of the right basis vectors for both SVD and ICA decompositions. The value of fourth-order cumulants is called *kurtosis* and it is a measure of the peakiness of a PDF. A kurtosis of 0 is a Gaussian distribution, with positive kurtosis being more peaky than a Gaussian and negative kurtosis being flatter than a Gaussian. For example, exponential distributions have a height positive kurtosis and uniform distributions have a low negative kurtosis. The sign of a distribution's kurtosis is called its modality. The table, then, shows that the kurtosis of the first and second components is greater than Gaussian for both SVD and ICA decompositions. However, the ICA has maximized the kurtosis to a much higher degree than the SVD suggesting that the Gaussian criteria does not point the basis vectors in the directions of

TABLE 6. Fourth-Order Contrast Values for SVD and ICA Right Basis Vectors of Bonfire

Basis Component	SVD Kurtosis	ICA Kurtosis
1	32.593	51.729533
2	4.736	18.102202
3	1.748	-0.640642
Contrast from Gaussian	1087.648	3004.044

greatest cumulants. The third component is somewhat Gaussian in both cases, suggesting that the third spectral component is in fact Gaussian. The contrast value at the bottom of the tables is the sum of squares of the kurtosis values and is used as a measure of deviation from normality (Gaussian-ness) for a PDF. Clearly the ICA has produced a greater deviation from normality, which suggests that higher-order PDFs exist in the signal. From the above results it is clear that ICA has done a better job of characterizing the features of the input TFD than SVD, thus we conclude that the ICA is a useful new technique for characterizing sound features.

3.5.2 Example 2: Coin dropping and bouncing sound

1. Method

The second example is that of a coin dropping and bouncing. The STFTM spectrogram of this sound is shown in Figure 26. Clearly discernible are the individual bounces which get closer through the course of the TFD, they exhibit a wide-band spectral excitation up to the Nyquist frequency, which in this case is 11.025 kHz because the original sound was sampled at only 22.050 kHz as compared with 44.1kHz for the other examples. This re-sampling does not affect our analysis since it just means that the spectrum is band-limited and it will serve as a good test of our methods for the restricted spectrum case. Since the coin is made out of a metallic material there is a high-frequency ringing component. This high-frequency component is observed to be continuous throughout the sound. We see this component centered at about 8kHz. Also discernible are a number of low and mid-frequency ringing components which are triggered by bounces, but do not continue for long. These components are metallic ring components with a rapid decay corresponding to internal damping in the coin and damping due to coupling with the surface of impact. We analyzed this sound with a 44.1kHz STFT in the same manner as the previous sounds, with no PSD normalization.

2. Results

The singular values of the coin drop sound are shown in Figure 25. The first three components account for approximately 33% of the total variance in the original sound. Again, as with the bonfire sound, the singular values decay rapidly at first followed by a steady exponential decay in the higher components.

Figure 27 and Figure 28 show the SVD and ICA basis vectors for the coin drop sound respectively. The left SVD basis vectors show some structural characterization. The individual bounces are relatively well delimited, but there is ambiguity across the functions. The right singular vectors have failed to separate the high-frequency, wide-band and low-mid components discussed above. Both the high-frequency components and the low-frequency components are mixed across the basis set. In contrast, we see that the left ICA basis vectors delineate three temporal behaviors: a decaying iterated amplitude sequence, a steady amplitude sequence with a single large spike and lastly we see the iterated impact amplitude functions for the bounce sequence. Similarly, the right basis vectors of the ICA show a clear separation of the TFD spectrum into three different spectral behaviors. The first is a low-pass component, corresponding with the rapidly decaying impact sequence of the first temporal vector which perhaps represents the initial high-velocity impacts of the coin with the impact surface, the second shows a predominance of mid and high-frequency spectral components corresponding with the ringing mentioned above and which is corroborated by the

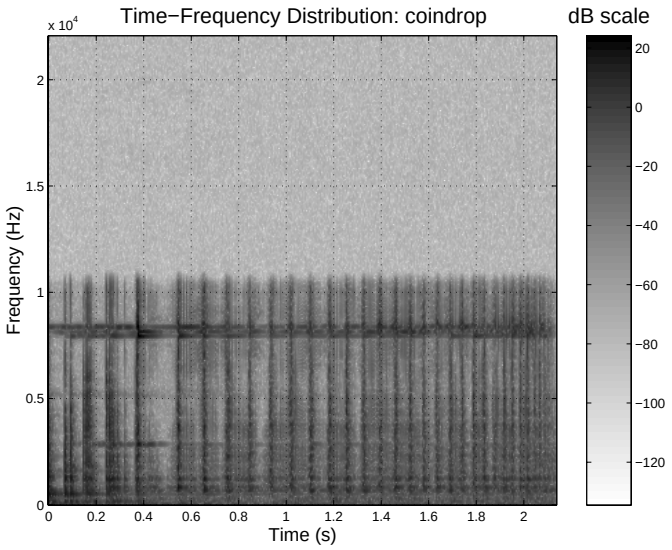

FIGURE 26. STFT spectrogram of coin drop sound. The features of interest in this TFD are the distinct high-frequency ringing component, the exponentially-decaying wide-band impact sequence (vertical striations) and low and mid-range ringing components.

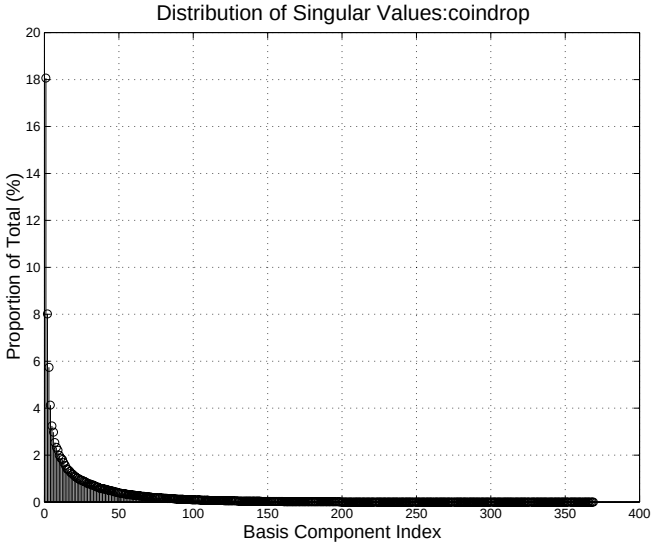

FIGURE 25. Singular values of coin drop sound. The first three components account for 33% of the origina variance in the TFD. The remaining components drop off gradually and exponentially.

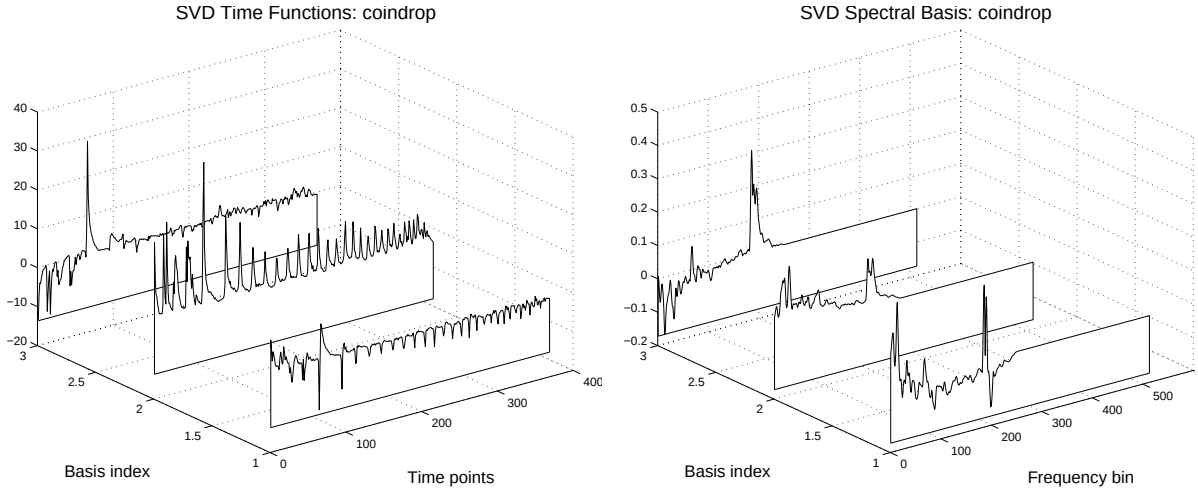

FIGURE 27. SVD basis vectors for the coin drop sound. The left singular vectors capture the structure of the coin bounces but there is some ambiguity. The right singular vectors fail to separate the high-pass, wide-band and low-pass components of the sound.

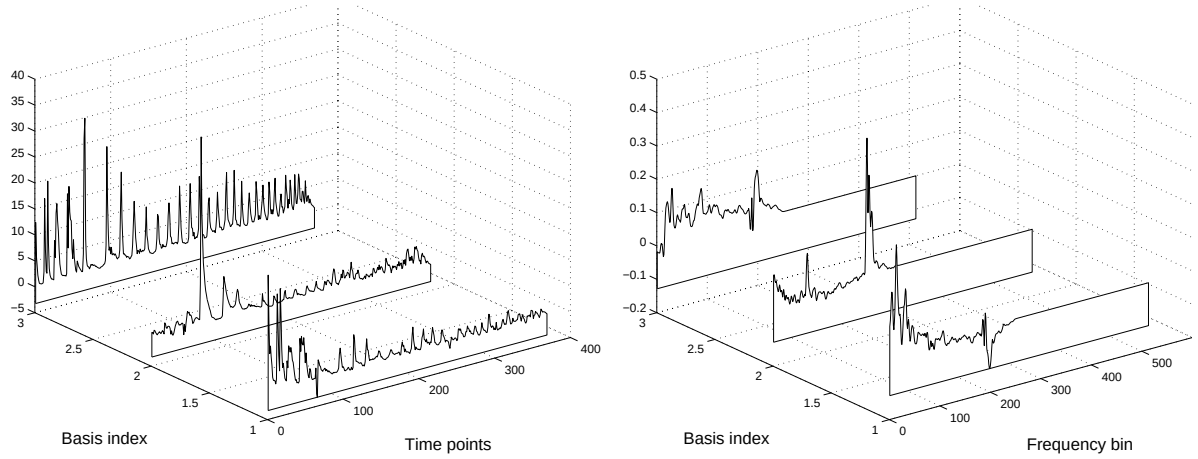

FIGURE 28. ICA basis vectors for the coin drop sound. The left singular vectors appear to reveal much of the temporal structure of the sound with the iterated behavior and the decaying envelope clearly delimited. The right singular vectors show a clear distinction between the low-frequency, high-frequency and wide-band components that we sought to characterize.

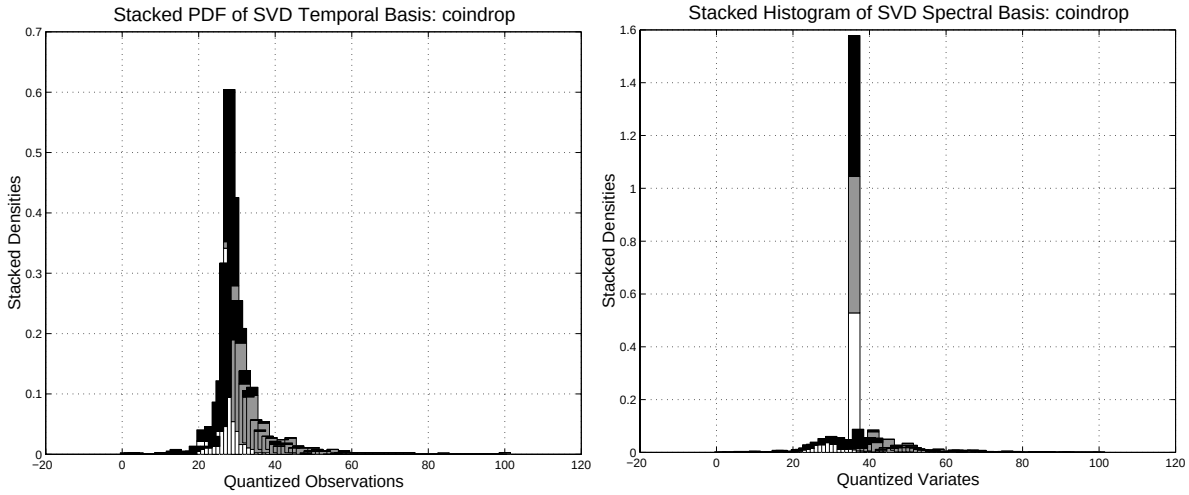

FIGURE 30. Stacked histograms of coin drop SVD basis functions.

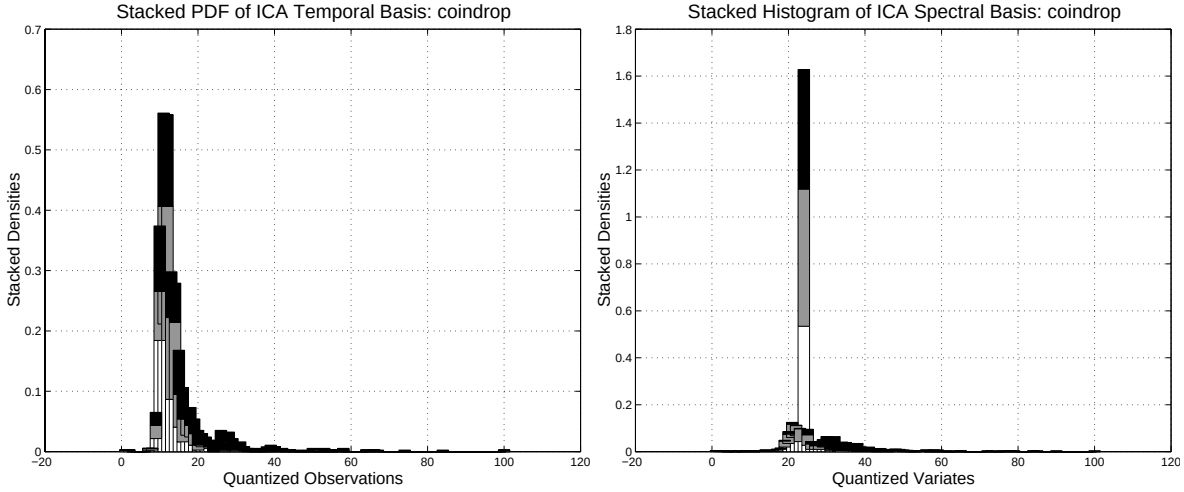

FIGURE 29. Stacked histograms of coin drop ICA basis functions.

continuous nature of the second ICA left amplitude function but for a single spike, and finally the third component exhibits the wide-band spectral content of the impacts, which again is supported by inspection of the left ICA amplitude functions. As with the previous example, these results suggest that the ICA has performed in a superior manner with respect to the characterization of features in the input TFD.

Figure 30 and Figure 29 show stacked histogram approximations of the PDFs of the coin drop sound for the SVD and ICA left and right vectors respectively. As with the last example, we can see that the SVD basis components are centered around a mean value and that the ICA values appear more skewed, suggesting the presence of higher-order cumulants in the underlying PDFs.

Table 7 shows the kurtosis values for the SVD and ICA right basis vectors respectively. As with the previous example, the ICA consistently maximizes the higher-kurtosis values whilst marginally changing lower values. This suggests that there is a single Gaussian component in the spectrum and that the other two components are non-Gaussian with relatively high kurtosis values. The contrast measure of the ICA compared with the SVD indicates that the ICA has a greater departure from normalcy and thus has performed better at characterizing the higher-order statistics of the input TFD.

TABLE 7. Fourth-Order Contrast Values for SVD and ICA Right Basis Vectors of Coin Drop

Basis Component	SVD Kurtosis	ICA Kurtosis
1	3.404	1.930
2	14.913	37.532
3	7.646	15.931
Contrast from Gaussian	292.476	1666.224

3.5.3 Example 3. Glass Smash Revisited

1. Method

Earlier in this chapter we gave an example of the application of SVD to the characterization of a glass smash sound. In this example we revisit the glass smash sound with an ICA analysis. The spectrogram of the glass smash sound is shown in Figure 8. In this example we not that the glass smash sound has a lot of energy in the low-frequency bands for the first few milliseconds of the sound due to the force of the impact. Since this energy is scaled disproportionately with respect to the energy of the subsequent glass particles shards we used a power-spectrum normalization of the input TFD as shown in the signal flow diagram of Figure 17.

2. Results

Figure 31 and Figure 32 show the singular values of the glass smash sound for non-normalized and PSD-normalized TFDs respectively. The subtle difference between the two is that the first component of the non-normalized TFD is much larger than the first in the PSD-normalized TFD. As we

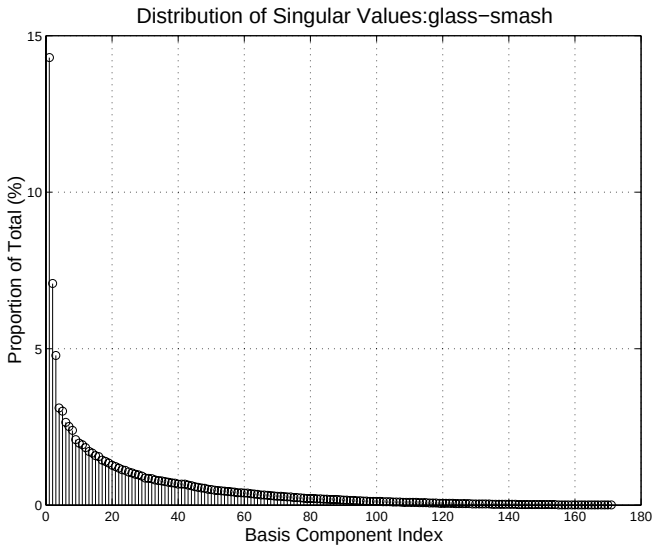

FIGURE 32. Singular values of glass smash PSD-normalized TFD decomposition.

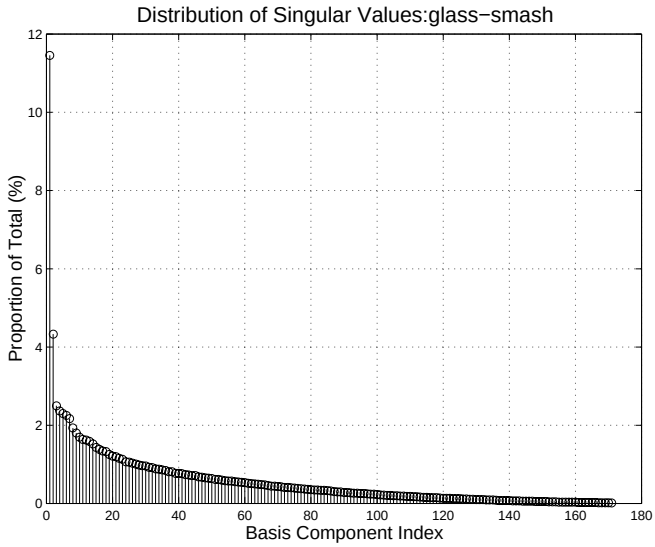

FIGURE 31. Singular values of glass smash non-normalized TFD decomposition.

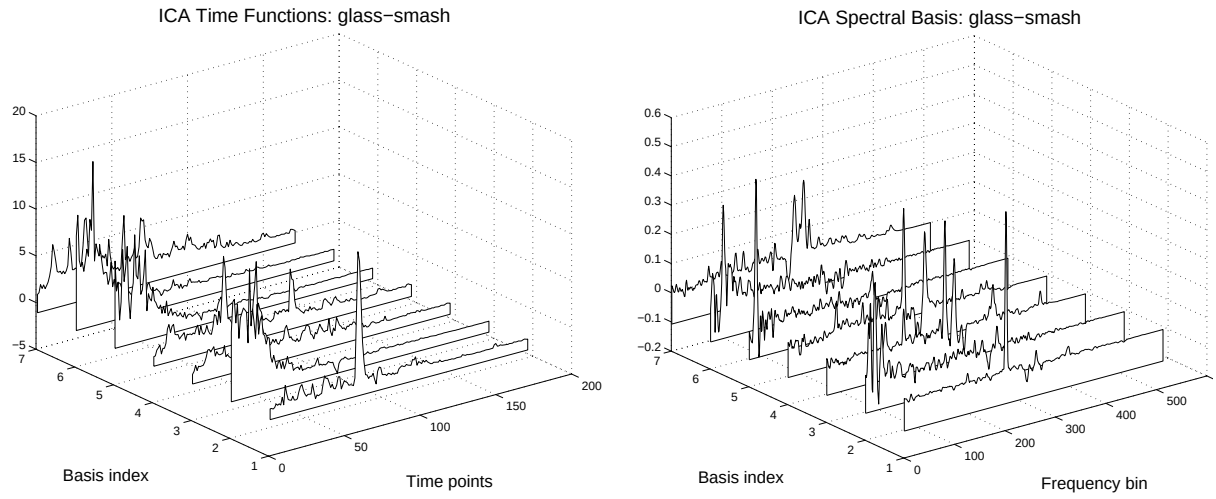

FIGURE 33. ICA basis functions for the non-normalized glass smash sound.

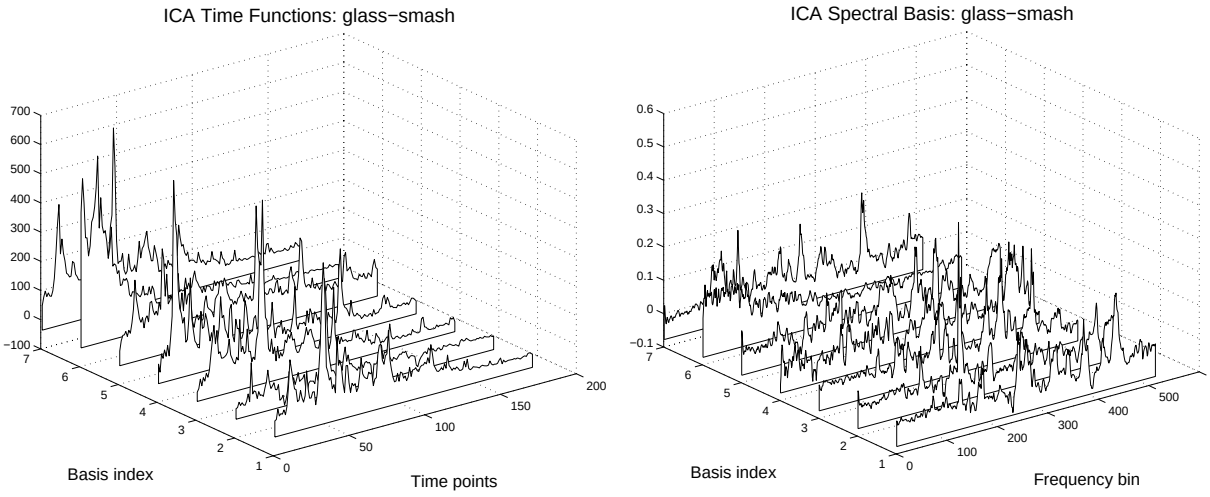

FIGURE 34. ICA basis functions for the PSD-normalized glass-smash sound.

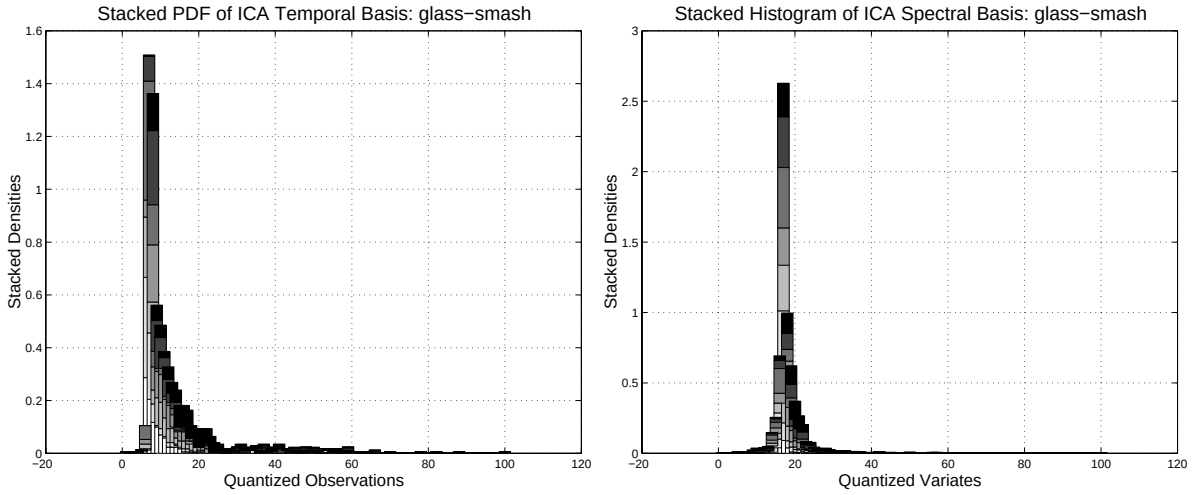

FIGURE 35. Stacked histograms of ICA basis functions for the non-normalized glass smash sound.

shall see, this first component corresponds with the low-frequency impact noise and the PSD normalization has had the effect of reducing its dominance for the decomposition.

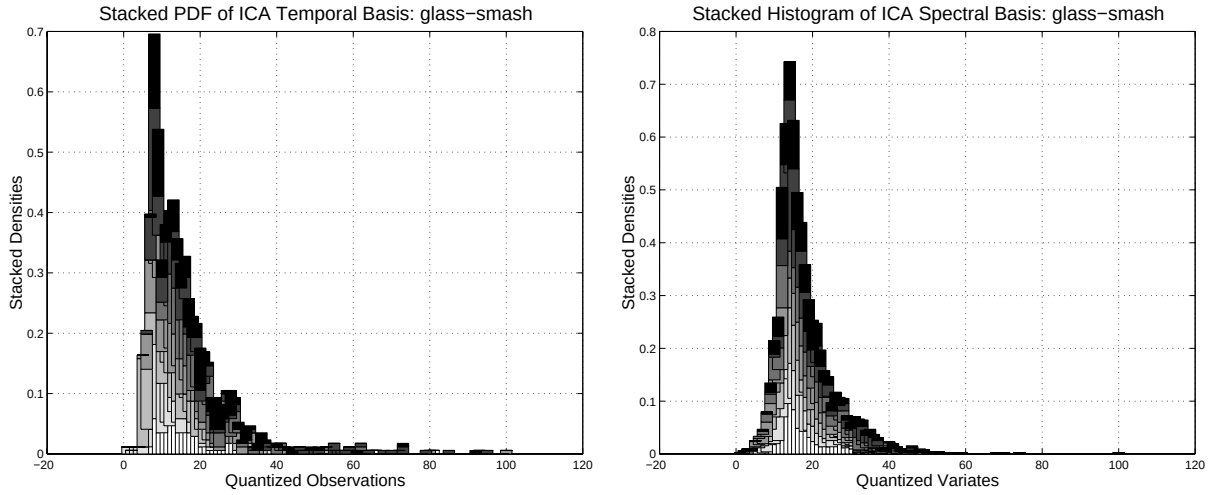

FIGURE 36. Stacked histograms of ICA basis functions for the PSD-normalized glass smash sound.

shall see, this component corresponds with the low-frequency impact noise. The effect is to enhance the presence of otherwise diminished components.

Figure 33 and Figure 34 show the basis vectors both the non-normalized and PSD-normalized input TFD respectively. The non-normalized functions show no less than three low-frequency impact components, suggesting an over-representation of this portion of the sound. The PSD-normalized vectors, however, show a more even spread of features, with only one feature corresponding to the low-frequency impact component. Amongst the other distinguishable components in the PSD-normalized basis are the presence of high-frequency narrow-band components which are not represented by the non-normalized spectrum. These components correspond with individual glass particles that are scattered throughout the spectrum. Thus they all exhibit high-Q narrow-band components, but with different fundamental frequencies. These observations are corroborated by inspection of the left-amplitude functions which show decaying-iterated time functions for these spectral components. From these results we conclude that the PSD-normalized spectrum has helped to significantly reduce the bias toward the low-frequency components in the input TFD. We suggest that PSD normalization may be a necessary tool for feature extraction from impact and explosion sounds due to the disproportionate representation of low frequency components in their TFDs.

Table 8 shows the values of the kurtosis for SVD and ICA decompositions of the PSD-normalized glass smash TFD. The ICA consistently maximizes the kurtosis over the SVD decomposition thus suggesting the presence of higher-order cumulants in the underlying PDFs. The contrast measure from normalcy suggests that the ICA has characterized the higher-order spectral structure of the glass-smash sound to a much greater degree than the SVD. The resulting basis vectors are therefore statistically independent to a much greater degree than the SVD basis vectors suggesting that ICA is again a better choice of decomposition for the input TFD.

TABLE 8. Fourth-Order Contrast Values for SVD and ICA Right Basis Vectors of Glass Smash

Basis Component	SVD Kurtosis	ICA Kurtosis
1	-1.827016	2.825701
2	0.024450	1.715675
3	1.463005	6.736605
4	1.218856	4.580318
5	2.932699	3.636371
6	1.359961	2.056869
7	7.047597	35.665819
Contrast from Gaussian	67.083415	1366.793833

3.6 Summary

In this chapter we have introduced the general notion of statistical basis decomposition of an arbitrary time-frequency distribution as a means for extracting features from sounds whose spectral and temporal properties are *a-priori* unknown. We developed a framework for investigating these techniques by considering principal component analysis. We have shown that PCA is not generally suitable for the decomposition of sound TFDs due to its reliance on a covariance representation which has the effect of decreasing the dynamic range of the input TFD with respect to the numerical accuracy of a given implementation. PCA was also shown not to characterize the different vector sub-spaces of a TFD; namely, the row space, the column space and the null space. Since, by our definition of a signal model, an input TFD is not necessarily assumed to be of full rank we sought a solution that would perform the said characterization.

The singular value decomposition was introduced as a method which directly decomposes a non-square matrix. This enables decomposition of a TFD without the need for covariance representation. This has the advantage of increasing the dynamic range of the decomposition over the standard PCA techniques. In addition to the advantage of rectangular decomposition, the SVD also provides a characterization of the vector spaces outlined above. This enables us to identify features for the row-space and column space of a TFD as well as enabling us to ignore the null-space components. Furthermore, this decomposition enables us to estimate the rank of the input TFD which provides a reasonable estimate of the number of statistically independent components in the input signal. However, we demonstrated that an SVD is limited by its assumption of Gaussian input statistics and showed that, for many sounds, such an assumption is not valid.

In order to address these problems we discussed a higher-order statistics extension of the SVD called independent component analysis. An ICA was shown to decompose the TFD of several complicated natural sounds in a manner that showed better feature characteristics than the corresponding SVD decomposition. The source of the better performance was the ability of the ICA to maximize cumulants at higher order than was possible using an SVD. Finally we showed that, in some cases, it is necessary to pre-process a TFD in order to remove bias toward high-energy low-frequency components. This was the case with a violent impact sound, glass smashing, and we demonstrated the superior performance of a power-spectral-density normalized TFD input over the standard non-normalized input. The ICA was also shown to outperform the SVD in this case.

In the larger context of our thesis, the techniques presented herein serve to identify features in a TFD. As discussed in this chapter, these features are considered to correspond directly to the structural invariants in an auditory group theory representation of a sound. Thus these methods are seen as a part of an auditory group analysis scheme. Whilst these components appear to be good candidates for sound-structure characterization, they do not offer a good characterization of time-varying components in a TFD. In the next chapter we consider how the statistical bases extracted from a TFD can be used to characterize time-varying structures within a sound in order to construct controllable models for sound-effect synthesis.