Chapter IV: Structured Sound Effects using Auditory Group Transforms

4.1 Introduction

In this chapter we develop signal processing techniques for implementing real-time, controllable sound-effects models. Since the domain of investigation of this thesis is the representation and synthesis of environmental sounds we have had to develop new signal representation methods in order to characterize the structure inherent in these sounds. The algorithms presented in this chapter are based on developing both the auditory group theory representation of sound structure and the use of statistical basis functions as structured audio source material. We start this chapter with a consideration of direct synthesis from the basis functions using inverse Fourier transform methods. Issues of reconstruction of time-frequency distributions from reduced basis representations will be covered

4.2 Resynthesis of Independent Auditory Invariants from Statistical Bases

In the previous chapter we investigated the problem of extracting statistically independent features from a time-frequency distribution under a signal model of superposition of outer-product independent TFDs. In this section we investigate the problem of reconstructing the independent TFDs from their statistical bases. These independent signals form the basic material for subsequent structured audio representation and synthesis.

4.2.1 Spectrum Reconstruction from Basis Components

Recall from the previous chapter the signal model of independent outer-product TFDs:

$$\chi = \mathbf{Y}_\rho \mathbf{V}_\rho^T = \sum_{i=1}^{\rho} \mathbf{Y}_i \mathbf{V}_i^T \qquad [121]$$

where χ is the spectrum reconstruction of the signal space of the analyzed sound. $\mathbf{Y}_i$ and $\mathbf{V}_i$ are the independent left and right basis vectors of each χ_i in χ of which there are ρ corresponding to

the estimated rank of the analyzed TFD. This equation specifies an algorithm for independent component resynthesis.

First let us investigate the reconstruction of a full-spectrum signal TFD from a basis set. Both the SVD and ICA basis components span the same subspace of a TFD. We can see this by investigating the form of the ICA analysis equation:

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{Q}^{T}\mathbf{P}^{T}\Lambda^{-1}\mathbf{D}\mathbf{D}^{T}\Lambda\mathbf{P}\mathbf{Q}\mathbf{V}^{T} = \mathbf{Y}\mathbf{Z}. \quad [122]$$

This equation describes the factorization of a TFD into a left basis set $\mathbf{Y}$ and a right basis set $\mathbf{Z}$ that span the full space of $\mathbf{X}$. But let us now consider the unitary transform $\mathbf{Q}$. The unitary transform produces orthogonal rotations of the basis vectors of an SVD: $\mathbf{U}$ and $\mathbf{V}$. This means that each plane of the SVD basis functions is mapped into the same plane, but it is rotated about the origin. Thus each plane of the SVD basis spans exactly the same Euclidean subspace as the resulting ICA basis under the unitary transform $\mathbf{Q}$. Furthermore, it is evident that the relation $\mathbf{Q}^{T}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^{T} = \mathbf{I}$ holds because $\mathbf{Q}$ is unitary. The upshot of this is that the reduced ensemble of ρ ICA basis vectors spans exactly the same subspace of $\mathbf{X}$ as the reduced ensemble of ρ SVD basis vectors, but the rotation of the basis within that subspace is different between the two, with ICA basis vectors pointing in directions that maximize the fourth-order cumulants of the underlying PDFs as described in the previous chapter.

Now let us consider the uniqueness constraints $\mathbf{P}, \Lambda$ and $\mathbf{D}$. $\mathbf{P}$ is a permutation matrix which is an orthonormal matrix and thus has the property $\mathbf{P}^{T}\mathbf{P} = \mathbf{I}$ which is the identity matrix. Λ is a diagonal scaling matrix and is used in conjunction with its trivial inverse (inversion of the diagonal entries) in Equation 122 which produces the relation $\Lambda^{-1}\Lambda = \mathbf{I}$. Finally, $\mathbf{D}$ is a matrix of real diagonal entries with unit modulus thus $\mathbf{D}\mathbf{D}^{T} = \mathbf{D}^{T}\mathbf{D} = \mathbf{I}$. Thus, for a reduced basis decomposition of ρ basis vectors, Equation 122 reduces to the following form:

$$\hat{\mathbf{X}}_{\rho} = \mathbf{U}_{\rho}\Sigma_{\rho}\mathbf{I}_{\mathbf{Q}}\mathbf{I}_{\mathbf{P}}\mathbf{I}_{\Lambda}\mathbf{I}_{\mathbf{D}}\mathbf{V}_{\rho}^{T} = \mathbf{U}_{\rho}\Sigma\mathbf{V}_{\rho}^{T} \quad [123]$$

where $\mathbf{I}_{Q}$ etc. indicates an identity transform due to a pair of complimentary matrices. This relation serves as a proof that the subspace spanned by the SVD basis is exactly the same as the subspace spanned by the ICA basis because they reconstruct exactly the same signal.

The utility of this proof is that, for a full spectrum reconstruction of the composite signal TFD χ, the resynthesized TFD is exactly the same for both the SVD and ICA cases. Thus there is no distinction between the two methods for the purposes of data compaction in the formulation of ICA that we developed in the previous chapter. However, since the *independent* basis components point in different directions through the same Euclidean subspace, there are quantitative differences between the two sets of individual bases. It is only the ensemble subspace that remains the same under the ICA transform. This is a desirable property of the ICA since it is an invertible represen-

tation of a TFD with respect to its SVD factorization. These points serve as a strong argument for using an algebraic form of an ICA rather than adopting *ad hoc* learning algorithm techniques.

If now consider the sum of independent TFDs formulation:

$$\chi = \mathbf{Y}_1\mathbf{V}_1^{T} + \mathbf{Y}_2\mathbf{V}_2^{T} + \ldots + \mathbf{Y}_\rho\mathbf{V}_\rho^{T} \quad [124]$$

we have shown that the effect of the ICA is to produce a different set of independent summation terms than an SVD, one that characterizes each independent component of χ in a more satisfactory manner than an SVD basis but which also preserves the vector space spanned by the SVD.

4.2.2 Example 1: Coin Drop Independent Component Reconstruction

As an illustration of these points we consider the spectrum reconstruction of the coin drop sound whose full spectrum is shown in Figure 26 in ChapterIII. To illustrate the first point, that the ICA and SVD span the same subspace of $\hat{\mathbf{X}}_\rho$, the 3-component reconstruction of the coin drop sound is shown in Figure 37 and Figure 38 for an SVD and an ICA basis respectively. These spectrograms are exactly the same thus empirically corroborating the proof of subspace equivalence.

We can see that the TFD reconstruction has captured most of the salient detail in the original TFD with only 3 components; the original non-factored TFD had 257 components ($\frac{N}{2}+1$ due to symmetry of the Fourier spectrum of a real sequence). The white patches in the TFD represent areas where the reconstruction produced negative values. Since this TFD is an STFTM each point in the TFD is a magnitude which, by its definition, cannot be negative for a complex spectrum. These negative values arise because a good deal of the original TFD space has been cropped by the projection onto a 3-dimensional subspace thus additive compensating spectral magnitudes have been eliminated. The negative values are clipped at a threshold of -100dB in order to create a sensible magnitude spectrum reconstruction. Comparison of the white patches with the original TFD reveals that these areas were of low magnitude in the original TFD. Only a very small portion of the reconstructed TFD needs to be clipped in the case of full independent-component signal spectrum reconstruction.

We now investigate the resynthesis of each independent component χ_i using both the SVD and ICA factorizations. Figure 40 and Figure 39 show a reconstruction of the first independent component of both the SVD and ICA factorizations respectively. This reconstruction can be viewed as the first summation term in Equation 124. What is immediately noticeable is that the SVD factorization produces an independent component TFD that seems to oscillate about two complimentary spectral bases producing a mesh-grid type pattern. This behavior stems from the tendency of an SVD factorization to contain negating components as well as additive components in each basis vector, which is due to the fact that the components are not really statistically independent. In contrast, a consideration of the ICA independent component shows clearly a behavior that matches the low and mid-frequency ringing components of the original coin drop TFD.

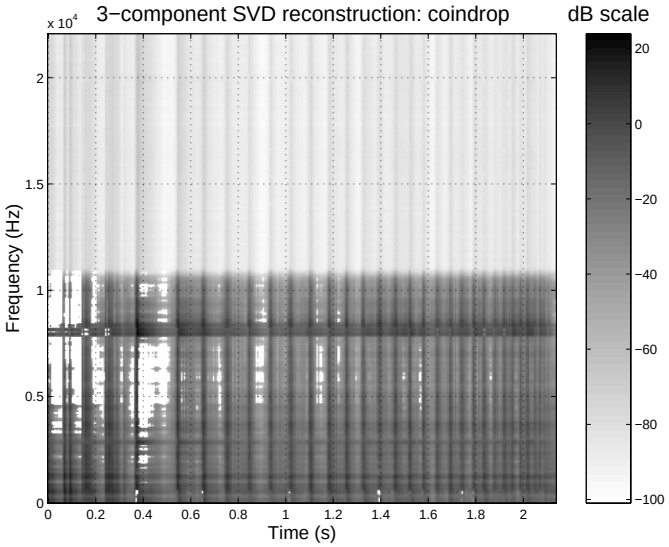

FIGURE 37. SVD 3 basis-component reconstruction of coin drop TFD.

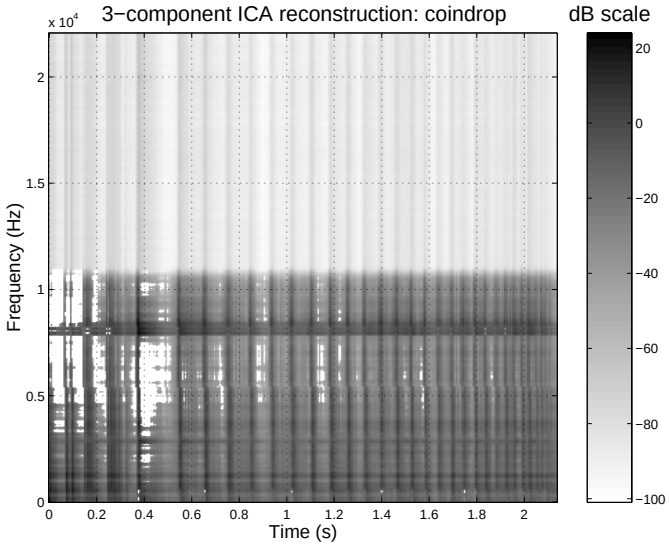

FIGURE 38. ICA 3 basis-component reconstruction of full coin-drop TFD.

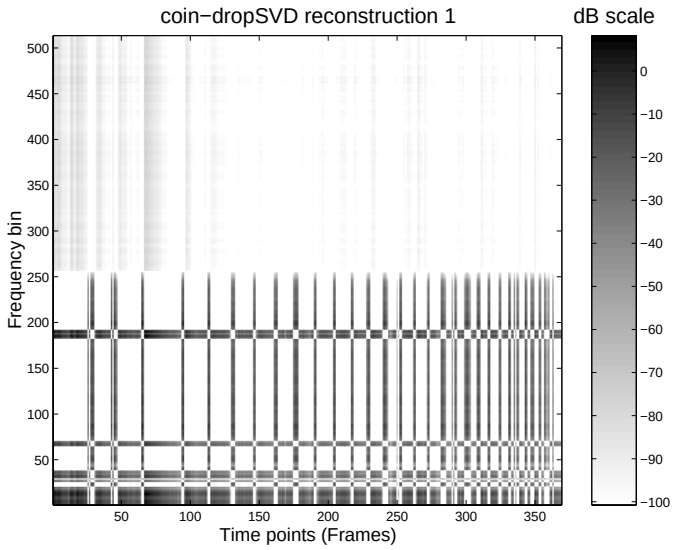

FIGURE 40. SVD reconstruction of first coin drop independent TFD.

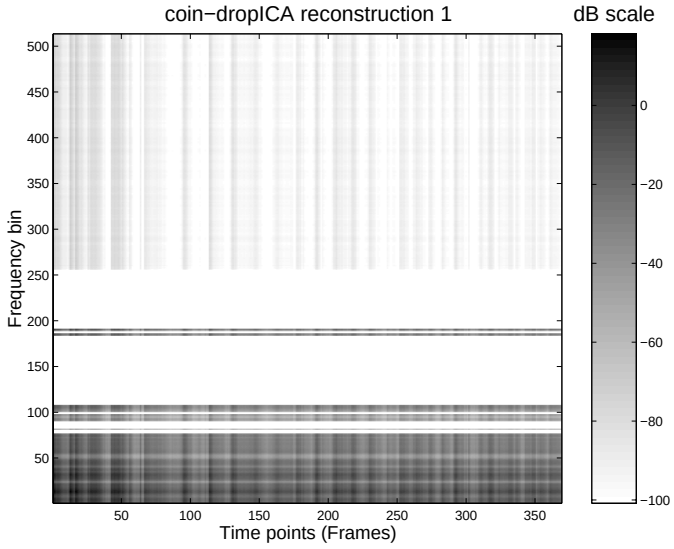

FIGURE 39. ICA reconstruction of first coin drop independent TFD.

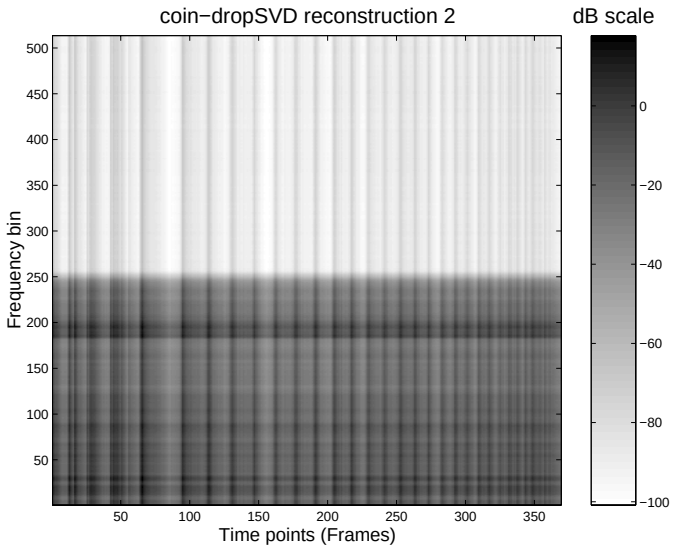

FIGURE 41. SVD reconstruction of second coin drop independent TFD.

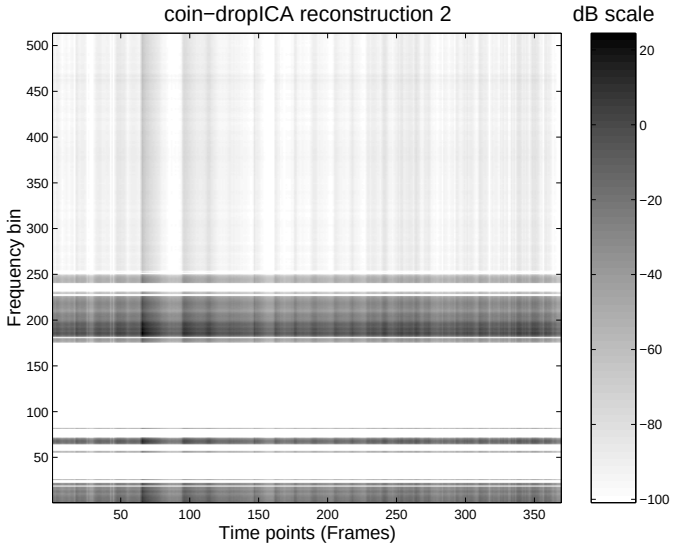

FIGURE 42. ICA reconstruction of second coin drop independent TFD.

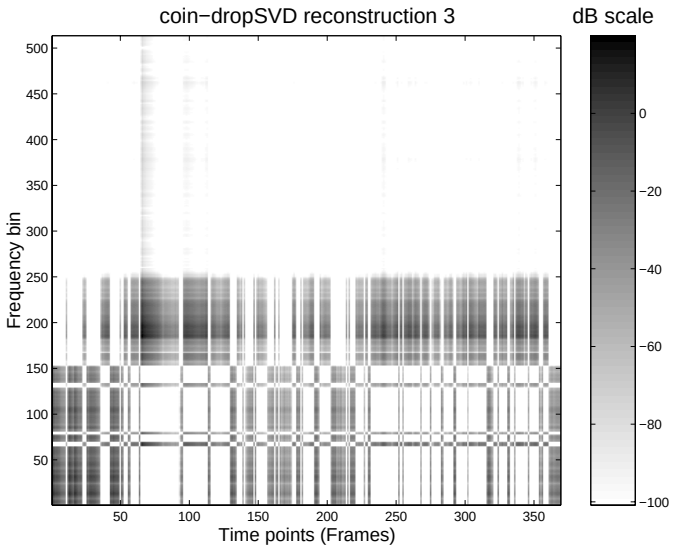

FIGURE 44. SVD reconstruction of third coin drop independent TFD.

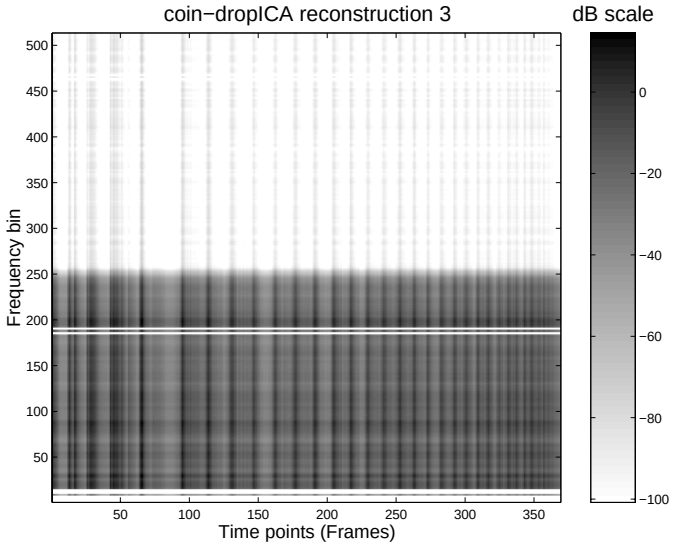

FIGURE 46. ICA reconstruction of third coin drop independent TFD.

Figure 41 and Figure 42 show TFD reconstructions of the second independent component for the SVD and ICA respectively. In this example the SVD component exhibits more stable behavior than in the previous example, but it has not successfully decorrelated the high-frequency ring component from the wide-band bounce-impacts. We can see this by the presence of both horizontal and vertical structures in the spectrum. The second ICA component has produced complimentary ringing components to those found in the first ICA component. These seem to correspond to mainly to high-frequency modes in the coin with some low and mid-frequency ringing also present.

Figure 44 and Figure 46 show TFD reconstructions of the third independent component. Again, the SVD has resorted to the mesh-grid pattern seen in the first SVD independent component. The third ICA component, however, reveals a clear vertical structure corresponding to the impacts of the coin on a surface with very little horizontal structure present in the TFD. This independence of this structure is due to the fact that the wide-band signal is an instantaneous excitation function for the coin, thus it is itself independent of the ringing from a physical point of view. It is the cause of the ringing, thus we see that the ICA has successfully decomposed the coin drop sound into independent components that make sense from a physical perspective as a source/filter model. This example has shown how we can reconstruct a set of two-dimensional TFDs from pairs of one-dimensional vectors. The vectors represent the most important structure of the TFD in terms of its orthogonal time-varying amplitude and time-static spectral components.

4.2.3 Example 2: Bonfire Sound

In this example we discontinue discussion of the SVD since we have sufficiently covered its limitations from the point of view of TFD feature extraction as well as TFD independent component resynthesis. The full spectrum TFD of the bonfire sound is shown in Figure 19 in ChapterIII. The 3-component reconstruction for the bonfire sound is here shown in Figure 47. The signal TFD reconstruction shows that the basis vectors have successfully captured both the wide-band erratic crackling as well as the continuous low-pass and wide-band noise densities of the sound.

Inspecting the independent TFD resynthesis we see from Figure 48 that the first independent component has captured some of the wide-band crackling components of the sound, with the dark-gray regions representing energy approximately 20-30dB above the light-gray regions. The second independent component, shown in Figure 50, shows a much more characteristic wide-band intermittent crackling component than the first. We note that the first component contains energy at lower frequencies, at around the 60th frequency bin, than the second which is perhaps the source of the independence. The third component, shown in Figure 49, clearly shows a horizontal continuous noise structure which contains none of the intermittent crackling thus demonstrating that the ICA has successfully separated the continuous noise component from the intermittent wide-band component.

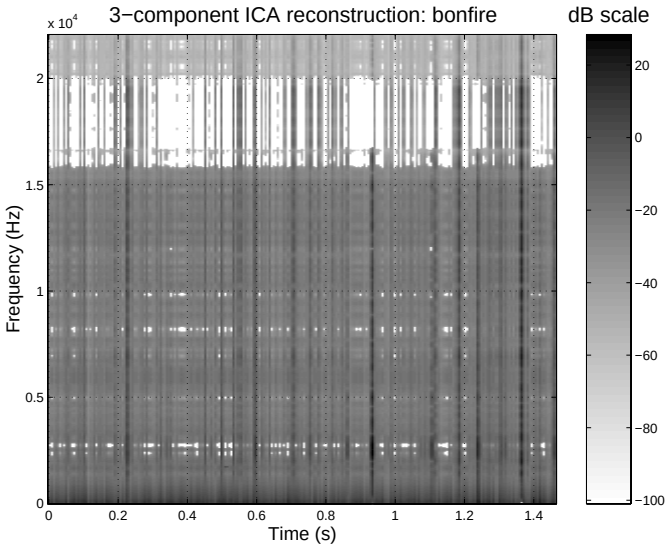

FIGURE 47. ICA 3-component reconstruction for the bonfire sound. (See Figure 19 on page 111 for full-spectrum TFD).

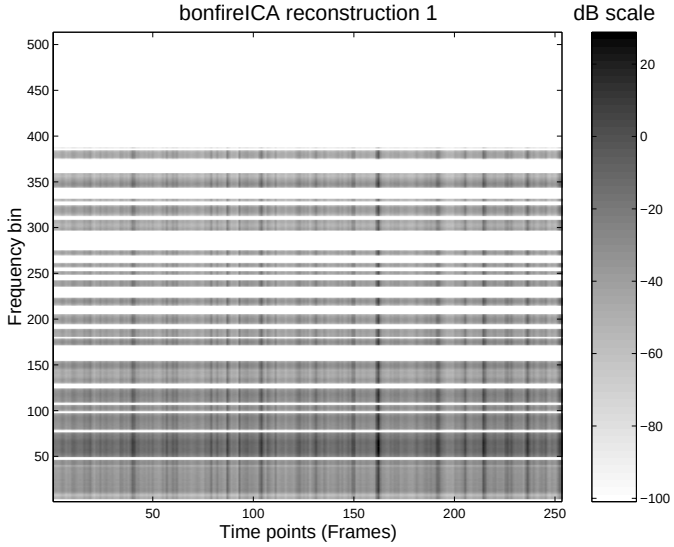

FIGURE 48. First Independent TFD reconstruction from ICA basis.

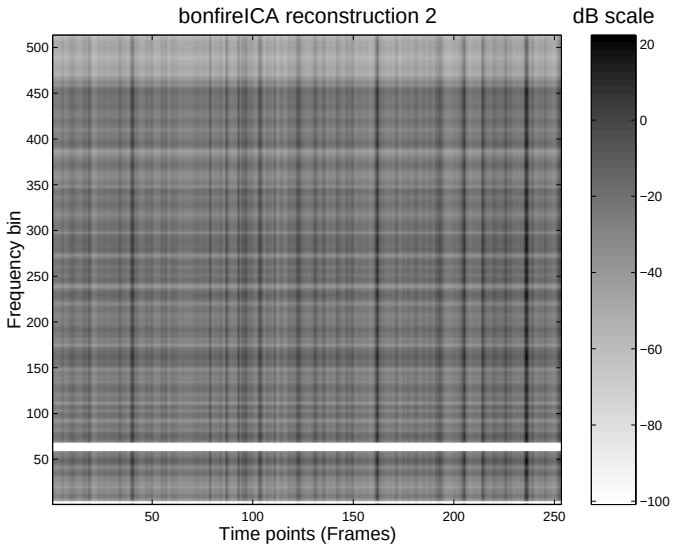

FIGURE 50. Second Independent TFD reconstruction from ICA basis.

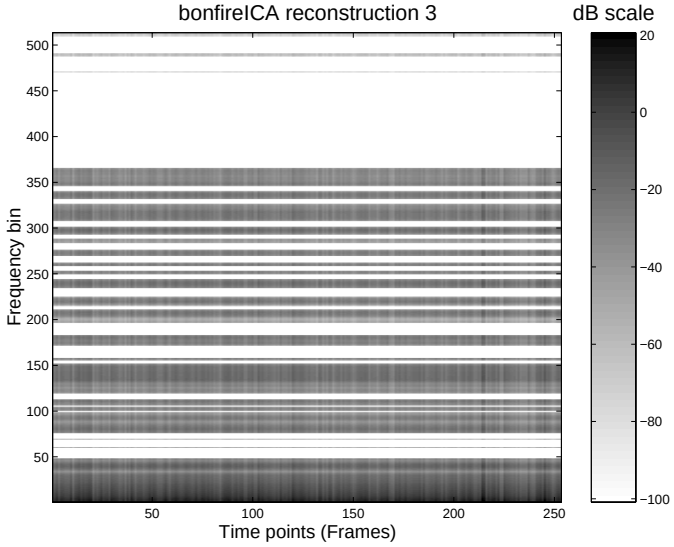

FIGURE 49. Third Independent TFD reconstruction from ICA basis.

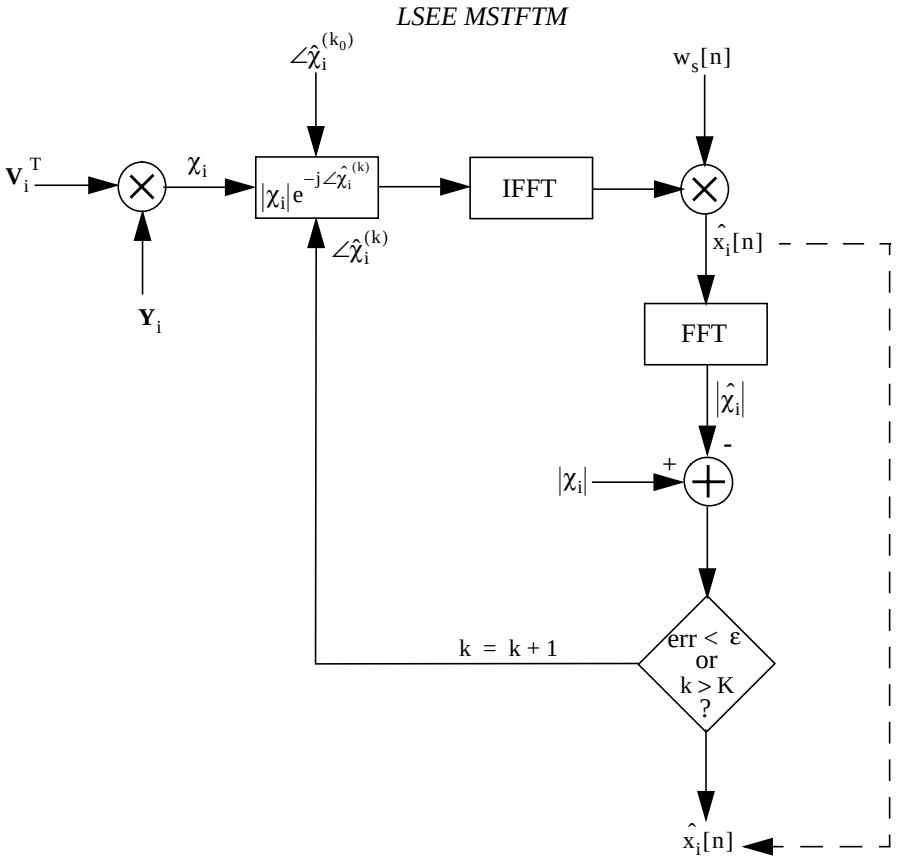

FIGURE 51. Independent component signal reconstruction algorithm. Least-Squares Error Estimation Modified Short-Time Fourier Transform Magnitude, based on Griffin and Lim (1984).

4.2.4 Signal Resynthesis from Independent Component Spectrum Reconstruction

Having obtained a set of independent magnitude spectrum reconstructions the problem at hand is how to estimate a signal for the independent component. Recall that we are assuming a magnitude-only spectrum representation for the TFD. We further assume, in this section, that the TFD can be transformed into an STFTM representation under some mapping. This is the approach used, for example, in Slaney et al. (1996) in which a correlogram representation is transformed into an STFTM representation for the purposes of correlogram inversion. Such a framework is quite general and enables us to proceed with little loss of generality in the methods.

Figure 51 shows a general algorithm for estimating a signal from a modified short-time Fourier transform magnitude representation. The STFTM TFD is constructed using the outer-product of the independent component vectors as described previously. Following the algorithm of Griffin and Lim (1984), which is also the algorithm used by Slaney et al. (1996), we seek to estimate a phase spectrum for the TFD such that the inverse transform yields a signal whose forward transform produces a magnitude TFD that minimizes the error with respect to the specified independent component in the least-squares sense.

To understand the problem consider that an arbitrary initial phase spectrum $\angle\hat{\chi}_i^{(k_0)}$ combined with the specified TFD magnitude $|\chi_i|$ in general is not a valid STFT since there may be no sequence whose STFT is given by $|\chi_i|e^{-j\angle\hat{\chi}_i^{(k_0)}}$, see Griffin and Lim (1984). The LSEE MSTFTM algorithm shown in Figure 51 attempts to estimate a sequence whose STFTM $|\hat{\chi}_i|$ is closest to the specified $|\chi_i|$ in the least squared error sense. By expressing the distance between the estimated and specified spectrum as a distance measure:

$$\delta\left\{\hat{x}_i[n], |\chi_i|e^{-j\angle\hat{\chi}_i}\right\} = \sum_{m=-\infty}^{\infty}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\chi_i(\omega)-\hat{\chi}_i(\omega)\right|^2 d\omega \qquad \textbf{[125]}$$

and applying Parseval's relation:

$$\delta\left\{\hat{x}_i[n], |\chi_i|e^{-j\angle\hat{\chi}_i}\right\} = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\left|x_i[mH-n]-\hat{x}_i[mH-n]\right|^2 \qquad \textbf{[126]}$$

a quadratic solution for $\hat{x}_i[n]$ is found the form of which is an overlap-add synthesis procedure with an error-minimizing windowing function $w_s[n]$:

$$x_i[n] = \sum_{m=-\infty}^{\infty} w_s[mH-n]\hat{x}_i[mH-n] \qquad \textbf{[127]}$$

$$w_s[n] = \left(\frac{2\sum_{m=0}^{\left(\frac{L}{H}\right)-1}\frac{H}{L}}{\sqrt{4a^2+2b^2}}\left[a+b\cos\left(\frac{2\pi n}{L}+\frac{\pi}{L}\right)\right]\right) \qquad \textbf{[128]}$$

where $a = 0.54$ and $b = -0.46$ are the Hamming window coefficients and H and L are the STFT hop size and window lengths respectively. Equation 128 holds only if $L = 4H$.

The procedure is iterative, estimating a signal then transforming it back to the Fourier domain for the next iteration and converging to an optimal solution (in the quadratic sense) over successive iterations. A solution is considered adequate if the total magnitude error across the entire TFD satisfies $\left\{ \hat{x}_i[n], |\chi_i| e^{-j\angle\hat{\chi}_i} \right\} < \varepsilon$, with ε chosen as an arbitrarily small number.

The LSEE MSTFTM algorithm can be helped by the incorporation of different initial conditions for the phase estimates. For harmonic, or quasi-periodic sounds an initial linear-phase condition for the independent component TFD provided good results. For noise-based spectra initial phase conditions of a uniformly distributed random sequence produced good results. Using an appropriate choice for initial phase conditions in the MSTFTM algorithm we found that the solution converges satisfactorily after approximately 25 iterations.

4.3 Auditory Group Re-synthesis

4.3.1 Signal Modification using the LSEE MSTFTM

The independent component re-synthesis algorithm described above recovers a signal from a magnitude STFT so it can be used for signal modifications of an independent component feature. We take the final spectrum estimate of the MSTFTM algorithm as the spectrum for modified re-synthesis. The form of signal modifications for the MSTFTM follow closely the form of the phase vocoder which was previously discussed in Chapter II. The main auditory groups corresponding to phase vocoder transforms are: T_π and T_Ω which are the time-only and frequency-only transforms corresponding to resynthesis hop-size alterations and frequency-scale transforms with compensating time-only transforms as discussed in Chapter II.

Using these transforms, it is possible to implement a real-time synthesis algorithm based on the inverse FFT. The algorithm performs much in the same way as the phase vocoder implementing independent time-stretch and frequency-scale operations on the estimated LSEE MSTFTM TFD discussed in the previous section, see Figure 47. Because the transformations are associative, the order of the component transformation is not important.

Whilst it is possible to use the auditory group transformed IFFT as a real-time algorithm for independent control over sound features the implementation is somewhat expensive, even given the relative efficiency of the FFT. So rather than focusing upon an FFT-based implementation we turn our attention to more efficient techniques.

4.3.2 Efficient Structures for Feature-Based Synthesis

One way of improving on the ISTFT re-synthesis method is to construct an equivalent, more efficient, model in the time domain. The basis for an efficient filter-based implementation for independent component resynthesis is that the TFD of each independent component comprises essentially a single filter that is modulated in amplitude for each time frame. This amplitude modulation, together with the phase estimates provided by the MSTFTM, constitute the total information necessary to reconstruct the signal. Therefore it is possible for us to design filters for each independent component right spectral basis vector and drive it with filters designed from the left amplitude basis vectors of an independent component TFD.

There are two general approaches to the problem of filter design for independent component modeling: finite impulse response (FIR) and infinite impulse response (IIR) modeling. We start with a consideration of FIR modeling because of its close relationship to the inverse Fourier transform method of re-synthesis discussed above.

4.3.3 FIR Modeling

Figure 54, Figure 52, and Figure 55 show FIR filters for the three features of the bonfire ICA analysis. The FIR filters are obtained by a zero-phase inverse Fourier transform:

$$s_i[n] = \frac{1}{N}\sum_{k=0}^{N-1} V_i[k]e^{j\frac{2\pi k}{N}n} \qquad [129]$$

where $V_i[k]$ is a single DFT vector obtained from the Fourier magnitude values in $\mathbf{v}_i$, the i-th column of the right basis vectors. The Z-transform of $s_i[n]$ is $S_i(Z)$ and we shall refer to it shortly.

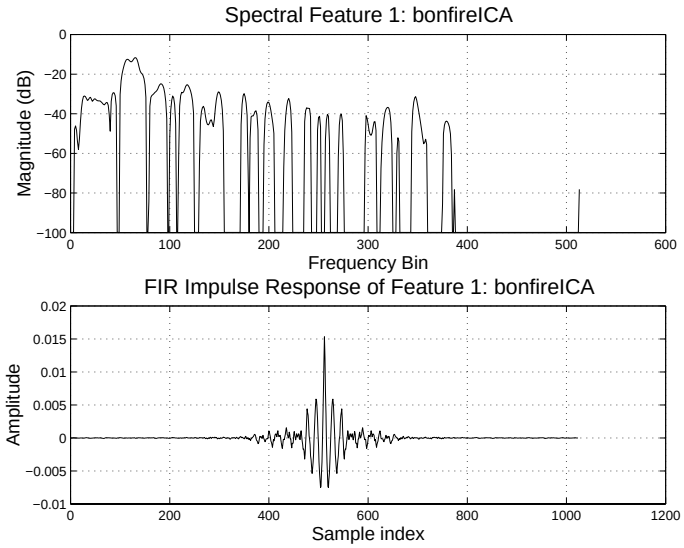

FIGURE 54. Bonfire sound: linear-phase FIR filter for spectral independent basis component 1.

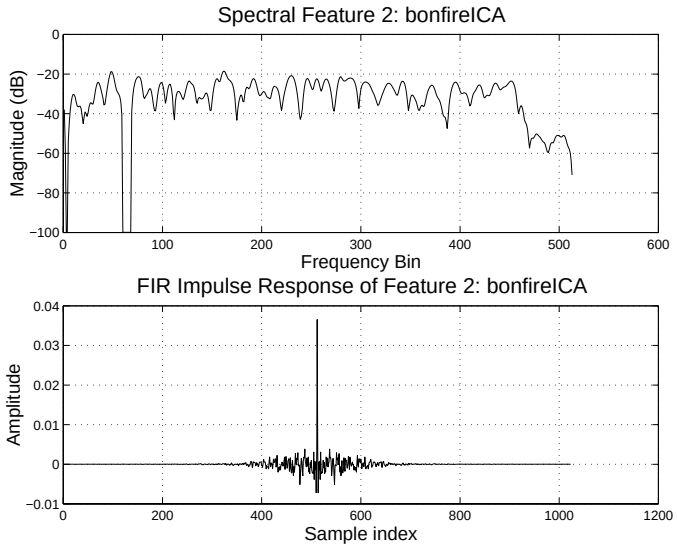

FIGURE 52. Bonfire sound: linear-phase FIR filter for spectral independent basis component 2.

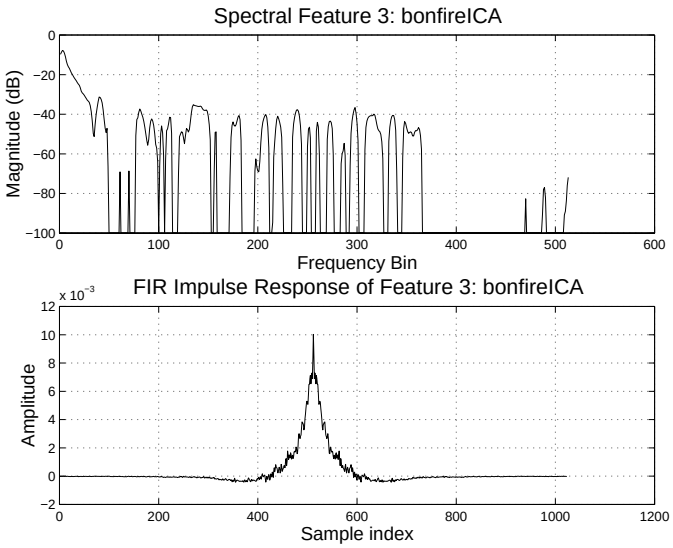

FIGURE 55. Bonfire sound: linear-phase FIR filter for spectral independent basis component 2.

These linear-phase FIR filters are used in conjunction with a matching set of IIR filters that are estimated from the left independent component basis vectors of the TFD. These IIR filters are low-order approximations of the time-amplitude basis functions and operate at the *frame rate* of the source STFT. In the examples that follow all of the time-amplitude IIR models of left basis vectors are 8th-order auto-regressive models obtained by an auto-covariance linear-predictive coding analysis (LPC). This type of analysis was discussed previously in Chapter II so we shall not explain it here.

The LPC analysis yields a set of filter coefficients for each independent component time function as well as an excitation signal. We can represent the form of this system as:

$$E_i(Z) = \frac{B_i(Z)}{A_i(Z)}, \qquad \textbf{[130]}$$

where $B_i(Z)$ are the zeros of the system and represent an excitation source signal for the amplitude function, and $A_i(Z)$ is an 8-th order prediction filter for the time-response of the amplitude function. These IIR models generate an impulse train spaced at the STFT frame rate. Time-stretch control is produced by altering the spacing of the impulse train which corresponds to a shift in the hop-size for the underlying STFT representation, this corresponds to the auditory group transform T_π which produces time-only alterations of a signal. Frequency-shift control is produced in the

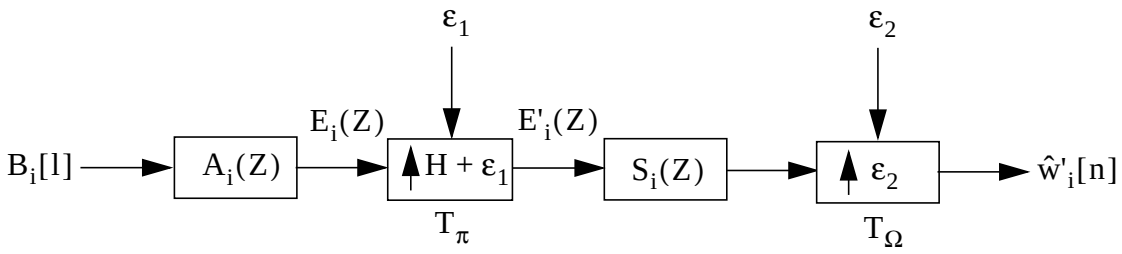

FIGURE 56. Implementation of structured independent component re-synthesis using a linear-phase FIR model $S_i(Z)$. The system function of the amplitude signal $E_i(Z)$ is specified at the Fourier transform frame rate. It is interpolated to the required hop size by an auditory group transform T_π which alters the global time structure of the sound. On the right, the transform T_π produces shifts in frequency by a factor ε_2.

same manner as the phase vocoder using the T_Ω frequency-only transform. The system flow diagram for FIR modeling of independent components is given in Figure 56. The input, on the left, is a source-excitation signal $B_i[l]$ expressed at the STFT frame rate, it's time response is generated by $A_i(Z)$ which is the prediction filter for amplitude functions. The time-scale-only auditory group transform produces shifts in the spacing of the amplitude functions which comprise a variably-spaced amplitude-varying impulse train $E'_i(Z)$. These frame-rate impulses are convolved with the FIR impulse response of the spectral basis component $S_i(Z)$ which is transformed by the frequency-shift-only auditory group transform. The result of these filtering operations is the synthesis of an independent component by separate control over the time-amplitude basis functions and the frequency basis functions of the underlying TFD.

By way of example, consider the IIR time-function models and excitation sequences in Figure 57-Figure 60. These signals are the systems-level implementation of the left ICA basis functions, and they are used in conjunction with the right-basis function FIR models described above. The first two figures show the IIR impulse response of the amplitude prediction filters. The figures show that the third independent component has a longer time response than the first. The third corresponds to a continuous noise component in the sound and the first corresponds to crackling components in the bonfire sound.

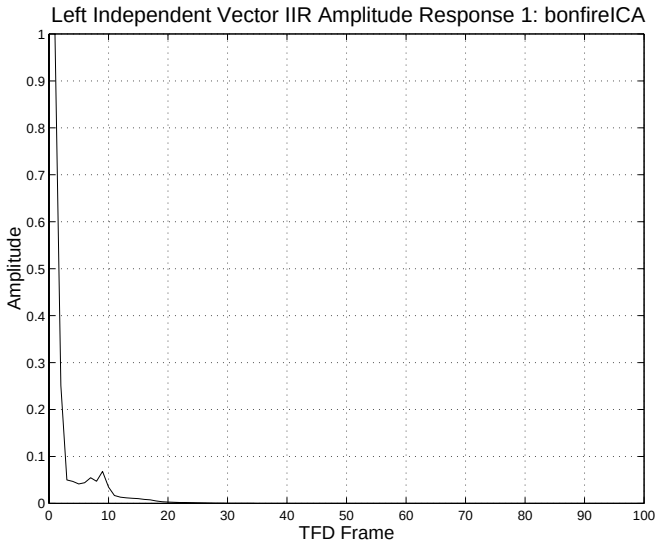

FIGURE 57. Impulse response of the first left independent vector IIR model of the bonfire sound.

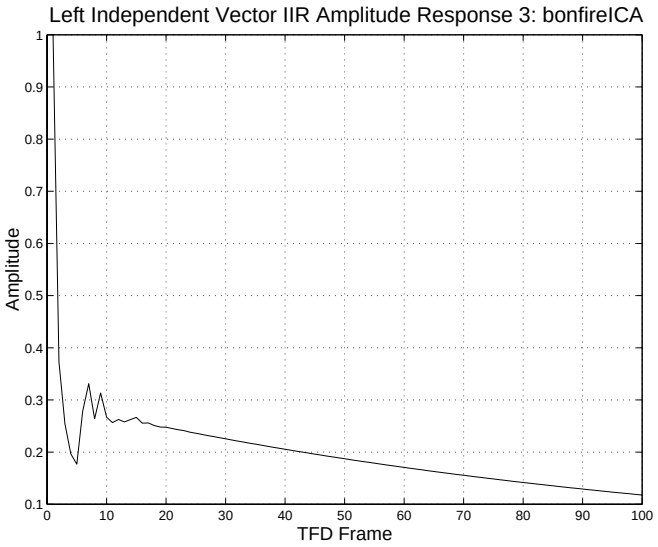

FIGURE 58. Impulse response of the third left independent vector IIR model of the bonfire sound.

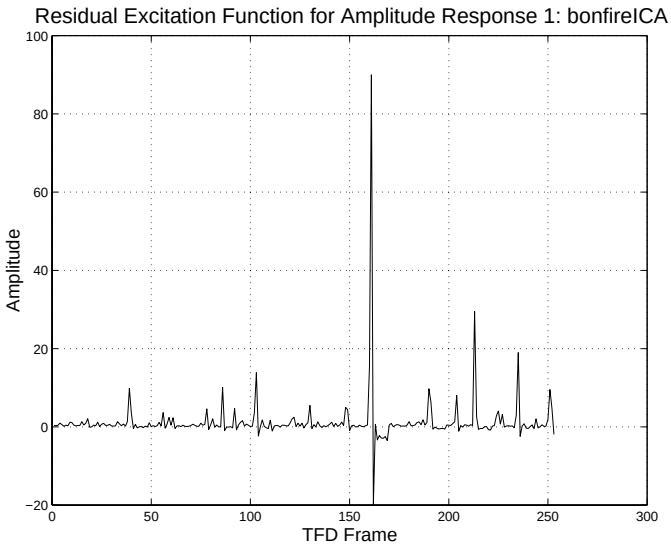

FIGURE 59. Excitation signal for first independent component amplitude function of the bonfire sound.

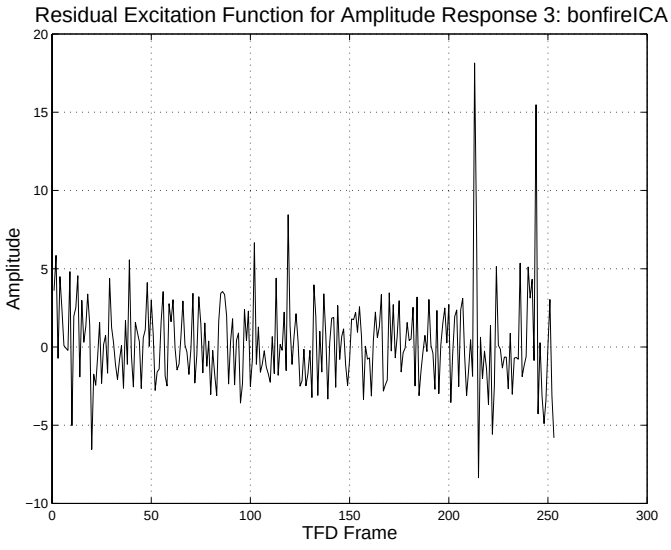

FIGURE 60. Excitation signal for third independent component amplitude function of the bonfire sound.

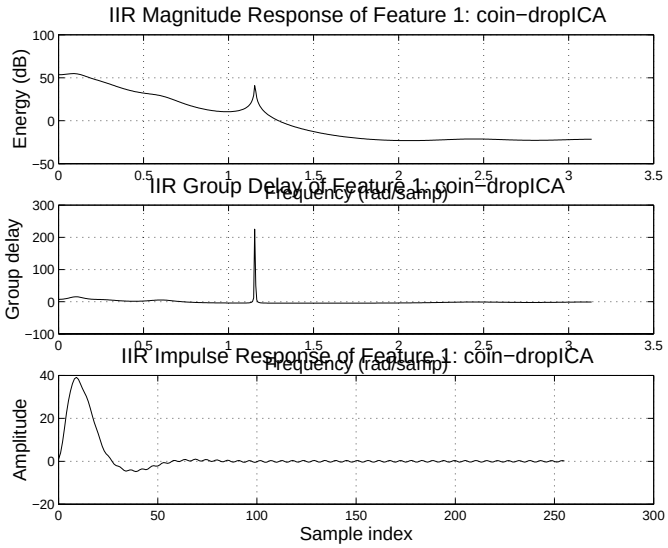

FIGURE 61. IIR Frequency Response, Group Delay and Impulse Response for the first independent component of the coin drop sound.

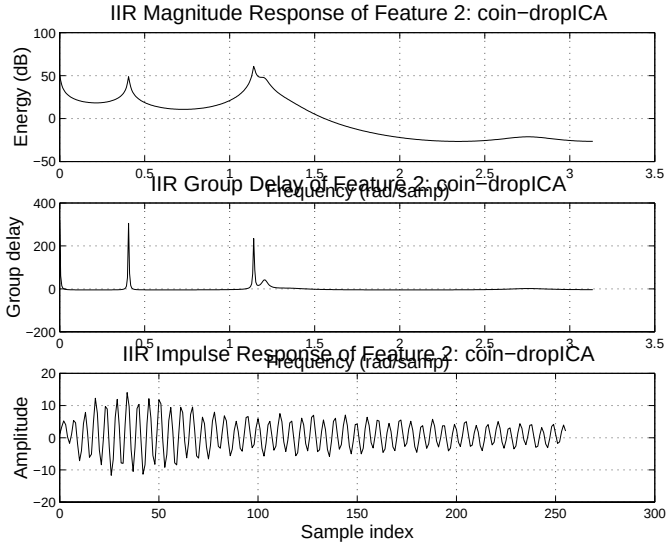

FIGURE 62. IIR Frequency Response, Group Delay and Impulse Response for the second independent component of the coin drop sound.

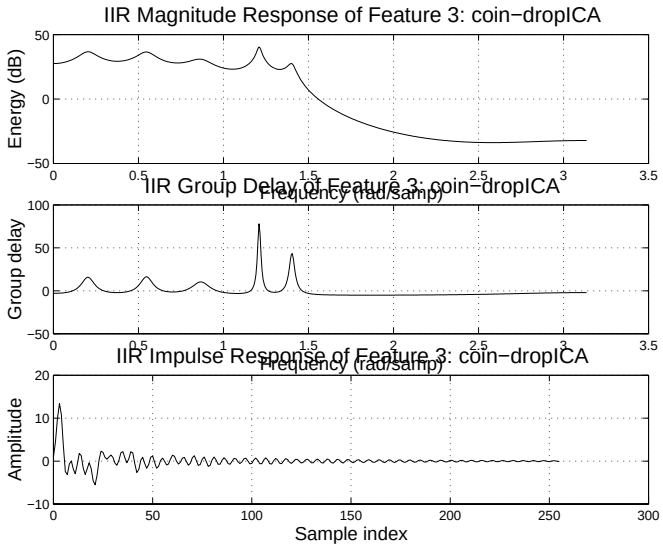

FIGURE 64. IIR Frequency Response, Group Delay and Impulse Response for the second independent component of the coin drop sound.

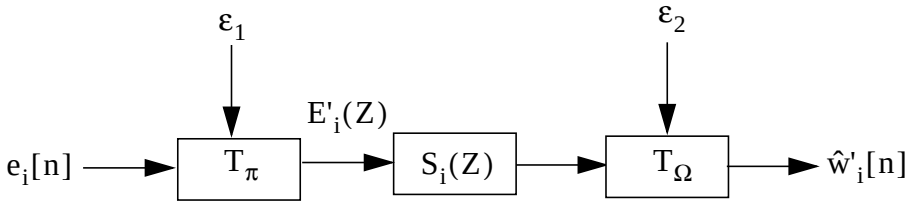

FIGURE 65. Implementation of structured independent component re-synthesis using an IIR system structure.

The second two figures show the excitation signals corresponding to the amplitude functions. We see a clear difference between the two illustrated components. The first shows an erratic impulsive behavior which is characteristic of crackling, and the second shows a continuous noise component that is characteristic of the bonfire sound.

4.3.4 IIR Modeling

The main problem with the FIR modeling technique is that it relies on a frame-rate much as the inverse short-time Fourier transform. In order to move away from the dependence on a frame rate we now consider the resynthesis of auditory invariants using IIR models. The method for IIR modeling of independent components starts with the MSTFTM signal for an independent component

$\hat{x}'_i[n]$. This signal is subjected to an LPC analysis using a higher-order than the LPC analysis used for FIR time-function modeling. The examples that we discuss in this section were obtained using a 12-th order LPC analysis. The LPC analysis yields a system function for each independent component as follows:

$$\chi_i(Z) = \frac{E_i(Z)}{S_i(Z)} \qquad [131]$$

where $\chi_i(Z)$ is the independent component sequence represented by its Z-transform, $E_i(Z)$ is an excitation sequence and $S_i(Z)$ is a spectral structure. This decomposition, for each of the ρ independent components gives us a convenient manner for extracting the components of a structured audio transform. Recall that the most general form of the structured audio transform was:

$$T_{structured}\{\mathbf{W}\} = \sum_{i=1}^{\rho} T_{U^{(i)}_{\varepsilon 1_i}}\{\mathbf{E}_i\} T_{V^{(i)}_{\varepsilon 2_i}}\{\mathbf{S}_i\}. \qquad [132]$$

the explicit separation of the signals $E_i(Z)$ and $S_i(Z)$ in Equation 131 gives us the final form of our analysis. Not only are we able to extract a number of independent feature signals from a TFD, but we now also have a deconvolution of the excitation and spectral-structure components of each independent component. This allows us to implement transforms of the type specified by Equation 132, which are well-formed structured audio transforms. This is the re-synthesis framework that we adopt for structured-audio re-purposing and control of sound effects. The system flow diagram for structured audio resynthesis using IIR models is shown in Figure 65.

By way of example for the IIR independent component resynthesis method consider Figure 62 - Figure 64. These figures show the frequency response and impulse response for each of the three independent components of the coin drop sound. The first component is generally low-pass with a narrow-band spike component, from the impulse response we determine that this component is heavily damped and very lowpass. The second component has a longer time response which corresponds to the ringing of the coin, this is also manifest as high-Q regions in the frequency response. The third component is wide band and corresponds to the impact component of the coin bounces. These figures demonstrate that the IIR models capture the features of each independent component quite well. Thus the efficient form of independent component resynthesis does a good job of characterizing the structure of the original TFD.

4.3.5 Characterization of Excitation functions

From the examples given above, we propose that the excitation functions can be generalized into four broad classes of behavior: impacts, iterations, continuous noise and scatterings. Each of these types of excitation function can be approximated with a parameterized unit generator function. Table 9 lists four common excitation functions that we use for modeling a wide range sound feature behaviours.

TABLE 9. Excitation Function Generators and their Modeling Uses

Function	Signal Variable	Modeling Applications
Iteration	I[n]	Periodic, bouncing
Gaussian	G[n]	Scraping, blowing
Poission	P[n]	Scattering, impact, spilling
Chaotic	C[n]	Turbulence, jitter

For example, the unit generator function of iterations is an impulse train with exponentially spaced impulses. By setting the time constant of the exponential to decay we create an excitation signal suitable for modeling bouncing events. A constant spacing between impulses is useful for generating excitation signals for periodic events such as hammering and footsteps.

4.4 Auditory Group Synthesis Models

Having established methods for approximating the spectral features of an ICA analysis with IIR filters and characterizing the excitation structures using the generator functions shown in Table 9, we now are in a position to describe the architecture of general-purpose sound synthesis techniques that use features extracted from recordings in order to generate novel audio content.

Figure 66 shows the general form of an auditory group model. The $T_{e_i}\{E_i(Z)\}$ elements represent excitation signals and their auditory group transforms, and the $T_{s_i}\{S_i(Z)\}$ elements are the spectral structures and their auditory group transforms. The signal processing network defined in this way essentially implements Equation 132.

Assuming that the spectral structures are approximated using IIR filter models as described above we can implement the auditory group transforms of spectral structures using pole manipulation algorithms. In Chapter 2 we described the invariance properties of size changes and changes in the Young's modulus of materials by inspecting the physical equations governing the acoustics. Both of these transformations of a spectral feature can be implemented efficiently by appropriate manipulation of the roots of the denominator polynomial of each $S_i(Z)$.

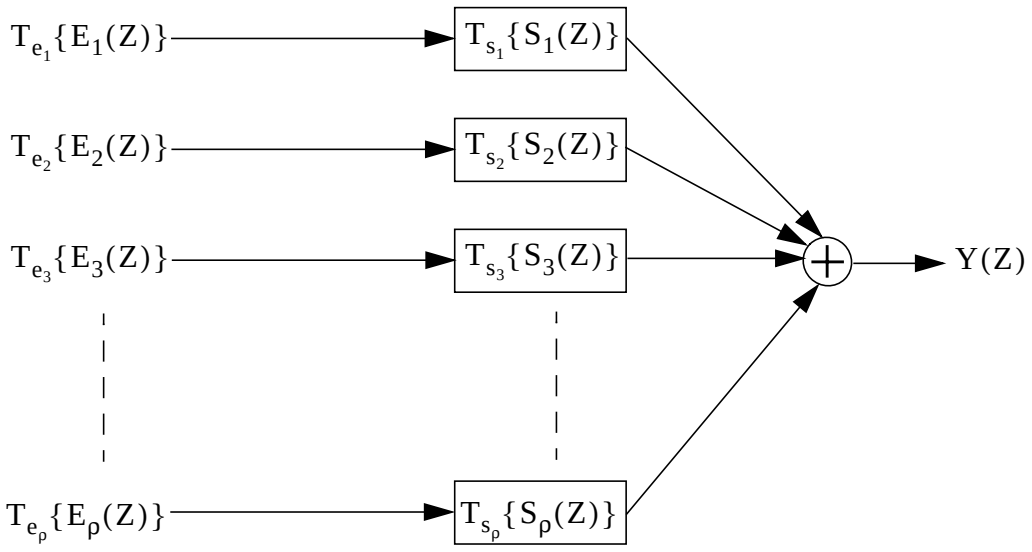

FIGURE 66. Schematic diagram of DSP implementation of multi-feature synthesis model with auditory group transforms T_e and T_s for each excitation signal and spectral structure respectively.

4.5 Future Directions

The system described thus far is capable of synthesizing a wide variety of sound types from single source/filter models to superpositions of source/filter models with time-varying behaviours. There are however some inherent limitations in the system as implemented thus far.

4.5.1 Orthogonality of ICA Transform

The first limitation of the current system is that the ICA transform is generated by an orthogonal rotation of the basis set from an SVD. Common (1994) points out that an ICA, if it exists for a given signal, is not constrained to be an orthogonal transform. Thus it should be possible to develop an algorithm that generates a non-orthogonal basis for an ICA. It is our expectation that improved separation performance for independent signal components will result from relaxing the constraint of basis orthogonality.

4.5.2 Weyl Correspondence and Transformational Invariant Tracking

Whilst we have shown that auditory group transforms can be used to specify physically-meaningful changes in an invariant, we have not fully demonstrated the possibility of tracking group transforms across a sound. For example, the sound of chirping birds would require a time-varying analysis that could separate the invariants from the frequency sweep transform generated by the chirp.

A number of recent developments in group-theoretic signal processing may be of importance to the correspondence and tracking problem. In particular, the time-frequency strip filters of Weisburn and Shenoy (1996) show promise as a method for tracking chirps in the time-frequency plane. Time-frequency strip filters are implemented as a special case of a *Weyl* filter, which relates the Weyl Correspondence to time-frequency analysis and thus provides group-theoretic origins for tracking transformational invariants.

4.5.3 On-Line Basis Estimation

A further requirement for successful tracking of invariants under transformation is that the basis estimation should utilize an on-line algorithm. The ICA algorithm that we developed in Chapter III serves our purposes well as long as the statistics of the input matrix are approximately constant across the matrix. For sounds with rapidly varying components, such as birds chirping, we require that the basis components be re-estimated for each frame of input, based on previous input. There are many such algorithms described in the ICA literature, for example Amari *et al.* (1996). A combination of on-line basis estimation and invariant tracking using time-frequency strip filters will allow greater accuracy in the analysis and characterization of complex everyday sound events.

4.6 Summary

In this chapter we first demonstrated that independent component basis vectors are better features of a sound than the corresponding singular value decomposition feature set, even though the two sets of basis vectors span exactly the same subspace of a time-frequency distribution. The independent components are used to reconstruct independent TFDs for each component of a sound. We gave methods for estimating a signal from the independent component TFDs based on an iterative procedure for minimizing phase errors.

These independent component signals can be used to further simplify the sound characterization by estimation of a set of filters using either FIR or IIR modeling techniques. The IIR modeling techniques were shown to be simpler in form than the FIR techniques but they are a little more computationally expensive. This extra expense, however, is eliminated when we consider the problem of phase modeling for FIR-based re-synthesis thus suggesting that IIR synthesis is a better model for implementing efficient resynthesis of independent components. The IIR filter model explicitly break each independent component signal into an excitation function and a spectral structure thus the combination of ICA analysis and IIR modeling results in a multi-source blind deconvolution of the latent statistical components in a TFD. This signal model is more ambitious

than most that are commonly used for audio signal processing algorithms and proves extremely powerful for audio re-purposing and control. The structure of the IIR resynthesis scheme was shown to be analogous to that of a well-formed auditory group transform thus satisfying all the conditions of a structured audio transform. To date there have been no audio signal processing methods capable of multi-source blind deconvolution and this technique may prove useful for application areas other than those presented.

In order to control sounds for structured re-purposing we presented a small collection of excitation modeling functions whose signals are representative of a wide range of natural sound behaviors. It was shown that a combination of these functions can be used to generate many different sound instances from a single input matrix representation. This excitation matrix is subject to control by auditory group transforms for generating novel features in sound structures. We also discussed transformation techniques for spectral-structure components that are used for modeling physical object properties such as size and materials.

The goal of this thesis was, at the outset, to find a method for representing and controlling natural sounds by their structured content. In this chapter we have demonstrated techniques for synthesizing and controlling the independent features of a sound in the desired manner. Our conclusion then, is that the said methodologies for representing and synthesizing natural sounds comprise a good working collection of tools with which to carry out the desired transforms.