Statistical modeling of intrinsic structures in impacts sounds

Sofia Cavaco[a)]
Computer Science Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213

Michael S. Lewicki[b)]
Computer Science Department and Center for the Neural Basis of Cognition, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, Pennsylvania 15213

(Received 4 May 2006; revised 20 March 2007; accepted 20 March 2007)

This paper presents a statistical data-driven method for learning intrinsic structures of impact sounds. The method applies principal and independent component analysis to learn low-dimensional representations that model the distribution of both the time-varying spectral and amplitude structure. As a result, the method is able to decompose sounds into a small number of underlying features that characterize acoustic properties such as ringing, resonance, sustain, decay, and onsets. The method is highly flexible and makes no *a priori* assumptions about the physics, acoustics, or dynamics of the objects. In addition, by modeling the underlying distribution, the method can capture the natural variability of ensembles of related impact sounds. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2729368]

PACS number(s): 43.60.Cg, 43.60.Wy [RAL] Pages: 3558–3568

I. INTRODUCTION

When an object is struck, the sound that it produces is determined by the physical properties of the object, such as its size, geometry, and material, and also by the characteristics of the event, like the force and location of impact. It is possible to derive physical models of impact sounds given the relationship between the physical and dynamic properties of the object, and the acoustics of the resulting sound. Models of sounds have proven useful in many fields, such as sound recognition, identification of events or properties (like material or length) of the objects involved, sound synthesis, virtual reality, and computer graphics. However, physical models are limited because of the *a priori* knowledge they require and because they do not successfully model all the complexities of real sounds.

One model of impact sounds is the resonance model proposed by Gaver (1994, 1988). This model consists of a sum of amplitude-decaying sine waves:

$$y(t) = \sum_{n=1}^{N} \alpha_n e^{-\delta_n t} \sin(\omega_n t), \tag{1}$$

where ω_n is the frequency of partial n, α_n is the initial amplitude of this partial, and $e^{-\delta_n t}$ is decay function of the same partial. The values of parameters ω, α, and δ can be set from mathematical expressions derived from physics for a limited set of very simple geometries for which the functions of frequency, amplitude, and decay are known. It is also possible to deal with more complex geometries by fitting the parameters to recorded sounds (Pai *et al.* 2001). A limitation of this simplified, knowledge-based model is that it fails to account for the rich structure and variability of real impact sounds. For instance, it fails to model the complex structure of the attack and the variability of sounds resulting from roughness in the surfaces. A solution to overcome this problem was proposed by van den Doel *et al.* (2001); however, some knowledge about the surfaces of the objects and their contact dynamics is still required. Other physical models have been proposed (e.g., Avanzini and Rocchesso, 2001a, b; Lambourg *et al.*, 2001), but as with the above-noted models, they require knowledge of the acoustics, as well as the physics, dynamics of contact, and the surface texture of the objects.

In order to obtain a detailed description of the modes of vibration and parameters of objects with complex geometries, some knowledge-based techniques use rigid body simulators developed for computer graphics (James *et al.*, 2006; O'Brien *et al.*, 2001, 2002). These approaches permit the synthesis of very realistic sounds; however, they are computationally intensive and they require a detailed description of the objects.

A more fundamental limitation of all these approaches, however, is that it is difficult to derive from natural impact sounds intrinsic acoustic properties beyond those that are explicitly modeled by the equations. For instance, how can a ringing property or a nonexponential decay be modeled by Eq. (1)?

This leads to another motivation for this work, which is the extraction of intrinsic features from sounds. Algorithms have been developed to extract basic features of impact sounds, such as the decay rates or the average spectra, but these approaches fail to capture the acoustic richness and variability that is characteristic of natural impact sounds.

In this paper, we propose a statistical data-driven method for learning the intrinsic features that govern the acoustic structure of impact sounds. The method aims to characterize the structures that are common to sounds of the same type (for instance, if the impacts on the same rod have

[a)]Electronic mail: scavaco@cs.cmu.edu
[b)]Electronic mail: lewicki@cnbc.cmu.edu

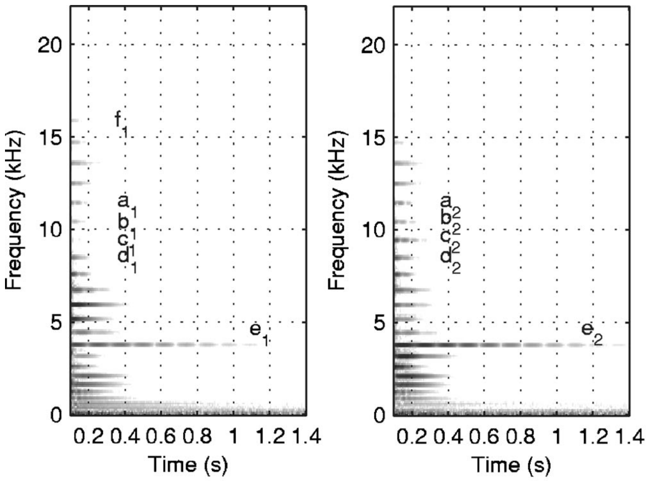

FIG. 1. Two spectrograms **S** (in decibels) of sounds (Al1 on the left and Al3 on the right) from impacts on an aluminum rod at approximately the same location and with approximately the same force (these spectrograms have been normalized). The relative power and temporal behavior of the partials varies from one instance to the other. For instance, in the left spectrogram, partial b_1 starts with a lower amplitude than partial a_1, while in the right spectrogram partial b_2 starts with a higher amplitude than a_2. The same happens with partials c and d: c_1 is weaker than d_1, while c_2 is stronger than d_2. Another example is the partial above 15 kHz. In the left spectrogram, this partial, f_1, is stronger than partial c_1, while in the right spectrogram c_2 is the strongest of the two. In fact, in the second spectrogram the partial above 15 kHz does not even appear.

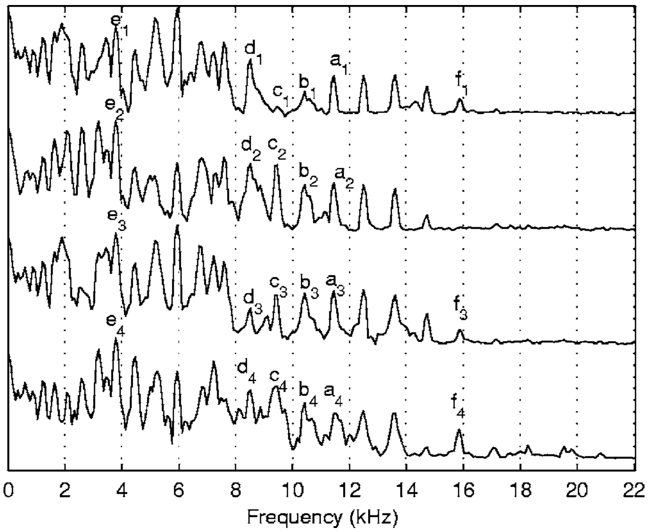

FIG. 2. Power spectra of four different impacts on aluminum (Al1, Al3, Al10, and Al19 from top to bottom) at approximately the same location and with approximately the same force. The relative power of the partials varies from one impact to another. (The partials are marked with the same labels as in Fig. 1.) Again, it can be seen that the relative powers of partials a, b, c, and d vary in the four power spectra. Also note that partial f appears in the first, third, and fourth lines (f_1, f_3, and f_4) but it is absent from the second line. Another interesting feature that can be observed is how the shape of the power spectrum changes from one sound to the next. For instance, note how partial a_1 is better defined than a_4.

a ringing property, the method should be able to learn a characterization of this intrinsic structure), as well as their variability (using the same example, the method should also capture the subtle variability of the ringing property in different impacts). At the same time, it aims for low dimensional representations of the sounds. This method requires no *a priori* knowledge and is used to create models of impact sounds that represent a rich variety of structure and variability in the sounds. The method is not restricted to learn an explicit set of properties of the sounds, and it has shown to be able to learn properties such as ringing, resonance, sustain, decay, and sharp onsets. To the best of our knowledge, this is the first statistical approach for modeling impact sounds.

II. MODELING INTRINSIC STRUCTURES

Our goal is to learn the intrinsic structure of sounds: We aim to decompose sounds in terms of the set of component signals that best describes them. For convenience, we assume the sounds are initially represented by a spectrogram, **S**. (Here we will refer to the rows of **S**, which are the power of frequencies over time, as *frequency bins* or *bins*, and we will refer to the columns of **S**, each of which is the power spectrum at a given time, as *frames*). Even though our method can be applied to a broader variety of sounds, here we will focus on impact sounds. To illustrate the data, Fig. 1 shows the spectrograms of two impact sounds on an aluminum rod (more details on how these sounds were produced and digitized are given in Sec. III).

Natural sounds of the same type have a rich variability in their acoustic structure. For example, different impacts on the same rod can generate very different acoustic waveforms. In natural environments there is variability due to reverberation and background noise, but even when the sounds are recorded in anechoic conditions, there is variability that is due to factors such as the slight variations in the impact force and location (see Sec. III for details on the recording conditions). Figures 1–3 show that, even though different impacts on the same rod have very similar spectra, the relative power

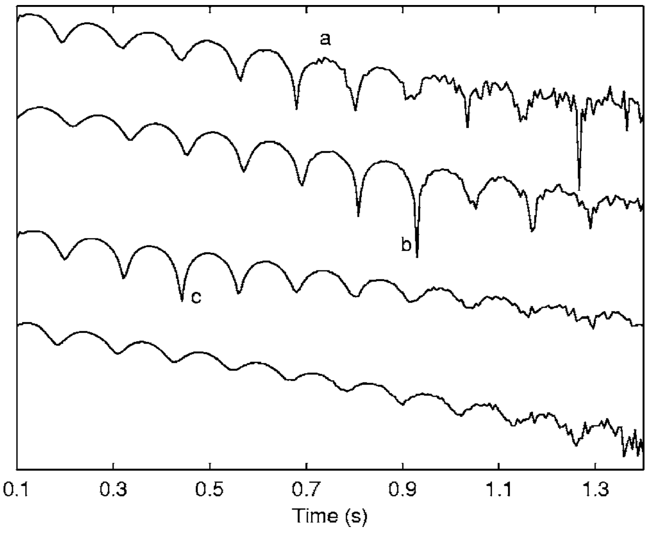

FIG. 3. The decay shape of the partial at 3.95 kHz in decibels (which is partial e in Figs. 1 and 2) for four different impacts on an aluminum rod (Al1, Al3, Al10, and Al19 from top to bottom) at approximately the same location and with approximately the same force. The temporal behavior of the partials varies from one impact to another; there is variability in the decay rate and beat pattern of this frequency bin. Note, for instance, the irregularities marked with an a in the first line, and notches b and c in the second and third lines. Also of interest is the consistency of the beating, which suggests that this rod has two close modes of vibration.

and temporal behavior of the partials varies from one instance to the other. These differences cannot be explained by a simple variation in amplitude of the whole spectrogram.

In spite of these variations, when these sounds are heard they are often perceptually very similar, that is, impacts from similar objects or materials have some common intrinsic structures that listeners can identify. Our goal is not to develop a perceptual model but rather to construct a model that learns the common intrinsic structures of similar sounds as well as their variability.

The model discussed here represents the sounds in terms of a set of component signals, in other words, it represents them in a new coordinate system. The form of the basis functions in the new coordinate system depends on the initial representation of the data, which here is the spectrogram. The frames are initially represented in an F-dimensional space with one dimension for each frequency bin f; let us call this space the *frequency space*. The bins are initially represented in a T-dimensional space, which we will call *time space*, with one dimension for each time frame t. A *spectral basis function* consists of a vector in the frequency space (that is, a spectra), while a *temporal basis function* consists of a vector in the time space (which can be thought of as a spectra's amplitude envelope). Using spectrograms as the initial representation allows us to model the sounds in spaces defined by spectral and temporal basis functions. (Section IV contains graphical examples of these basis functions.)

Given that the spectrogram $\mathbf{S}$, of size $(F \times T)$, is defined over a discrete set of frequencies, $f \in \{f_1, \ldots, f_F\}$, and a discrete set of time instants, $t \in \{t_1, \ldots, t_T\}$, we can define $\mathbf{S}$ as an ordered set of bins or as a sequence of frames. (Here we use only the power spectrum, and we ignore the phase component.) The model, which we call the bin model, or M_b, expresses the spectrogram $\mathbf{S}$ as an ordered set of bins. These are modeled as linear combinations of temporal basis functions $\boldsymbol{\phi}_i$:

$$\mathbf{b}_f = \sum_{i=1}^{I} \boldsymbol{\phi}_i c_{i,f}, \tag{2}$$

where $\mathbf{b}_f$ is the transpose of the fth bin of $\mathbf{S}$. $\boldsymbol{\phi}_i$ is scaled at this bin by coefficient $c_{i,f}$.[1] The value of I depends on the technique used to lear the basis functions $\boldsymbol{\phi}_i$. Here $I \leqslant T$ (see Sec. III and Appendix C in the supplementary material for further details). The basis functions $\boldsymbol{\phi}_i$ describe the temporal regularities in the bins in the data set, that is, in $\mathbf{S}$. These basis functions can describe a single sound, or the temporal regularities of a set of related sounds simply by including the appropriate spectrogram bins in the data set. (Section IV A shows how to learn $\boldsymbol{\phi}_i$.) The vectors of coefficients are commonly called *source signals*. Since the vector that consists of the coefficients that are associated with basis function $\boldsymbol{\phi}_i$, that is $\mathbf{c}_i = (c_{i,f_1}, \ldots, c_{i,f_F})^T$, ranges over the frequency space, here we call it a *spectral source signal*. (For graphical examples of spectral source signals see Sec. IV A.) Spectral source signal $\mathbf{c}_i$ scales basis function $\boldsymbol{\phi}_i$ across frequencies.

In order to represent the spectrograms of different sounds with a fixed basis $\boldsymbol{\Phi}$ (where $\boldsymbol{\Phi}$ represents the set of temporal basis functions $\boldsymbol{\phi}_i$), the model requires different spectral source signals to scale each basis function $\boldsymbol{\phi}_i$, i.e., there will be one set of spectral source signals for each sound. We distinguish these variables with an upper index k, that is, the set of spectral source signals associated with sound k, which is the set containing $\mathbf{c}_1^k, \ldots, \mathbf{c}_I^k$, is represented by $\mathbf{C}^k$.[2] We can thus rename some of the above-used variables to take into account the sound they refer to. Equation (2) can thus be rewritten as

$$\mathbf{b}_f^k = \sum_{i=1}^{I} \boldsymbol{\phi}_i c_{i,f}^k, \tag{3}$$

where $\mathbf{b}_f^k$ is the transpose of the fth bin of $\mathbf{S}^k$, i.e., the spectrogram of sound k, and the scalar $c_{i,f}^k$ is the fth element of $\mathbf{c}_i^k$.

If we consider all F bins in $\mathbf{S}^k$, Eq. (3) can be rewritten as

$$(\mathbf{S}^k)^T = \boldsymbol{\Phi}\ \mathbf{C}^k, \tag{4}$$

where the ith column of matrix $\boldsymbol{\Phi}$ contains $\boldsymbol{\phi}_i$, and the ith row of $\mathbf{C}^k$ contains $(\mathbf{c}_i^k)^T$. (See Appendix A in the supplementary material for figures of the matrices.)

Thus far, M_b describes the temporal structure, but not the spectral structure inherent in the spectral source signals $\mathbf{c}_i^k$. We can extend M_b to consider the regularities in the spectral source signals for an ensemble of related sounds. Instead of describing the temporal shape of a given bin, this part of the model describes the spectral source signals $\mathbf{c}_i^k$. These signals are modeled as a linear combination of spectral basis functions $\boldsymbol{\psi}_j^i$:

$$\mathbf{c}_i^k = \sum_{j=1}^{J} \boldsymbol{\psi}_j^i v_{i,j}^k, \tag{5}$$

where the scalar $v_{i,j}^k$ is a scaling coefficient. The spectral basis functions $\boldsymbol{\psi}_j^i$ describe the spectral regularities in the spectral signals. Again, the value of J depends on the technique used to learn the basis functions $\boldsymbol{\psi}_j^i$. Here, $J \leqslant F$ (see Appendix C in the supplementary material for further details). (Section IV B shows how to learn $\boldsymbol{\psi}_j^i$.)

We can now consider the previous equation at a given frequency bin f and express $c_{i,f}^k$ as follows:

$$c_{i,f}^k = \sum_{j=1}^{J} \psi_{j,f}^i v_{i,j}^k, \tag{6}$$

where $c_{i,f}^k$, and $\psi_{j,f}^i$ are the values of $\mathbf{c}_i^k$, and $\boldsymbol{\psi}_j^i$ at frequency bin f, respectively. (In other words, they are the fth values of vectors $\mathbf{c}_i^k$ and $\boldsymbol{\psi}_j^i$, respectively.)

Finally, combining Eqs. (3) and (6) it follows that the bins of $\mathbf{S}^k$ can be expressed as

$$\mathbf{b}_f^k = \sum_{i=1}^{I} \sum_{j=1}^{J} \boldsymbol{\phi}_i \psi_{j,f}^i v_{i,j}^k. \tag{7}$$

This shows that $\mathbf{S}^k$ can be modeled by temporal bases $\boldsymbol{\Phi}$, spectral bases $\boldsymbol{\Psi}$ (where $\boldsymbol{\Psi}$ contains all spectral basis functions $\boldsymbol{\psi}_j^i$), and a set of coefficients $\mathbf{V}^k$ (where $\mathbf{V}^k$ contains coefficients $v_{i,j}^k$), that is, $\mathbf{S}^k = M_b(\boldsymbol{\Phi}, \boldsymbol{\Psi}, \mathbf{V}^k)$. (For more details and figures of the matrices used in this model, see Appendix A in the supplementary material.)

The model is thus defined by two sets of basis functions, and the objective is to find the sets of basis functions with which the data can be better described: ideally only a few basis functions would be needed to accurately describe the data with less redundancy. In Sec. IV, we show that the basis functions can be learned effectively by redundancy reduction techniques.

As mentioned before, we can define $\mathbf{S}^k$ as an ordered set of bins or as a sequence of frames. Model M_b describes the data as an ordered set of bins, and it is possible to build an alternative model that describes the data as a sequence of frames. Yet depending on the techniques used to learn the basis functions, model M_b is more appropriate than the alternative model, in the sense that it may give a better description of the statistics of the data used in this study (see Sec. III for a description of the data and Appendix B in the supplementary material for further details). Therefore, here we focus only on model M_b, and we do not describe the alternative model.

III. METHODS AND TECHNIQUES

We used a set of impact sounds that were produced using four rods with the same length and diameter, but made of different materials. A wooden rod, with a much shorter length but the same diameter, was used as a mallet. Several impacts on each rod were recorded in an anechoic chamber. The location of the impacts and the impact force varied slightly from one instance to the next, since the rods were hit by hand. The sounds were digitized using a sampling frequency of 44 100 Hz.

The spectrograms of the sounds were computed using a 11.6 ms sliding Hanning window. Successive frames overlapped by 5.8 ms. Like with any other system that uses spectrograms, there is a trade off between spectral and temporal resolution. Even though the type of structures obtained for different resolutions is the same, the choice of spectral versus temporal resolution affects the representation: the shapes of the structures obtained differ slightly; for instance a structure that includes a sharp onset can look more or less sharp depending on the resolution. Here, we only report the results obtained using an intermediate resolution of 512–point fast Fourier transform.

We use principal component analysis (PCA) and independent component analysis (ICA) to learn the sets of basis functions from Sec. II. PCA and ICA are redundancy reduction techniques that look for the axes that best describe the distribution of the data. These techniques are used to represent high dimensional data in a (usually lower dimensional) space with less redundancy. The data are expressed as a linear transformation of the basis functions (i.e., the axes that define the new space). Given a set of M *source signals* of size N [represented by an $(M \times N)$ matrix $\mathbf{Y}$ with one signal per row] mixed into a set of M *signal mixtures* [represented by an $(M \times N)$ matrix $\mathbf{X}$ with one signal mixture per row] PCA and ICA learn a $(M \times M)$ matrix $\mathbf{W}$ that allows extracting the source signals from matrix $\mathbf{X}$:

$$\mathbf{Y} = \mathbf{W}\mathbf{X}. \tag{8}$$

If $\mathbf{A} = \mathbf{W}^{-1}$ this equation can be rewritten as

$$\mathbf{X} = \mathbf{A}\mathbf{Y}. \tag{9}$$

The two techniques differ importantly in the way they model the distribution of the data, and in their constraints. PCA is a second-order statistical method that assumes a Gaussian distribution and is restricted to orthogonal basis functions (that are the eigenvectors of the data covariance matrix). This technique decomposes a set of signal mixtures into a set of decorrelated signals and can be used to reduce the dimensionality of the data by considering I basis functions, where $I < M$ (in which case only I source signals are obtained). ICA is a generative model that decomposes a set of signal mixtures into a set of maximally independent source signals. This higher-order statistical method models multivariate data with non-Gaussian distributions and is not restricted to orthogonal basis functions. ICA contains PCA as a special case when the marginal distributions of signals are assumed to be Gaussian and the bases are restricted to be orthogonal. [For more details on ICA and PCA, see Hyvärinen *et al.* (2001) or Stone (2004).]

For instance, in the case of the first part of M_b and when we consider K impact sounds, matrix $\mathbf{X}$ consists of the horizontal concatenation of transposed spectrograms $((\mathbf{S}^1)^T, (\mathbf{S}^2)^T, \ldots, (\mathbf{S}^K)^T)$, $\mathbf{A}$ is the spectral basis $\mathbf{\Phi}$, and $\mathbf{Y}$ is the horizontal concatenation of the matrices of spectral source signals $(\mathbf{C}^1, \mathbf{C}^2, \ldots, \mathbf{C}^K)$. A signal mixture is the concatenation of one transposed frame from each of the K spectrograms, and there are T signal mixtures. Therefore, $I \leq T$ in Eqs. (2)–(7).

We used a built-in function from MATLAB to do PCA and the Fast ICA software package by Hyvärinen *et al.* (2001) to do ICA (for more details see Appendix C in the supplementary material). Because both PCA and ICA model the variation around the data mean, we used both the data matrix and its negative, i.e., we used the *extended matrix* $(-\mathbf{S}^T, \mathbf{S}^T)$, so that the mean would be zero. This was done so that the model described the signal rising and falling from zero, rather than the spectrogram mean.

IV. RESULTS

In this section, we show how to learn representations of the intrinsic structures of impact sounds. We show that the method developed in Sec. II can be used to characterize the structures of a single sound or the structures of sets of related sounds. In the latter case, the method learns representations of the structures that are common to the set of sounds and models their natural variability.

Section IV A explores the first part of the model, which is characterized by the temporal basis functions $\mathbf{\Phi}$, while Sec. IV B explores the second part of the model, which is characterized by the spectral basis functions $\mathbf{\Psi}$. Finally, Sec. IV C illustrates how the natural variability of related sounds is represented by the model.

A. Temporal basis functions $\mathbf{\Phi}$

There are two ways of applying ICA and PCA to spectrograms: these techniques can be used to do a *spectral analysis* of $\mathbf{S}^k$, in which the signal mixtures and source signals are considered to be spectra, or a *temporal analysis* of $\mathbf{S}^k$, in which the signal mixtures and source signals are considered to be temporal signals. *Spectral analysis* considers the frames (or power spectra) of $\mathbf{S}^k$ as a linear combination of independent or uncorrelated spectral source signals (for ICA and PCA, respectively). Here the goal is to decompose $\mathbf{S}^k$ into this set of spectral source signals. (For more details see Appendix B in the supplementary material.)

In order to learn the set of temporal basis functions $\mathbf{\Phi}$ and decompose the spectrograms into sets of spectral source signals, we apply spectral PCA and ICA to the spectrogram of a single impact or to the spectrograms of different impacts on the same rod. For instance, in order to learn the temporal basis functions $\mathbf{\Phi}$ and find the sets of spectral source signals $\mathbf{C}^1, \mathbf{C}^2, \ldots, \mathbf{C}^K$ for K sounds, model M_b does a spectral analysis on matrix $((\mathbf{S}^1)^T, (\mathbf{S}^2)^T, \ldots, (\mathbf{S}^K)^T)$, where $(\mathbf{S}^1)^T$ to $(\mathbf{S}^K)^T$ are time aligned, so that the matrix has one row (or transposed frame) that corresponds to the start of all K impacts. The temporal basis functions $\mathbf{\Phi}$ are time varying functions that represent temporal properties of different subspectra of the sounds. Each spectral source signal $(\mathbf{c}_i^k)$ is associated with a particular temporal basis function $(\boldsymbol{\phi}_i)$ that represents a component of the signal's temporal behavior.

1. One impact sound

We start with the spectrogram $\mathbf{S}$ of a single sound. Figures 4(a) and 4(b) show six out of the ten most dominant basis functions (i.e., $\boldsymbol{\phi}_1 - \boldsymbol{\phi}_{10}$) learned by ICA.[3] As can be seen, ICA is able to isolate temporal properties of the sound: see for instance $\boldsymbol{\phi}_b$ in Fig. 4(a), which represents a ringing property of the sound, $\boldsymbol{\phi}_d$ in the same figure, which represent a decay property of the sound, $\boldsymbol{\phi}_a$ in Fig. 4(a) and $\boldsymbol{\phi}_e$ in Fig. 4(b), which represent sustain properties, and the sharp basis functions like $\boldsymbol{\phi}_a$ and $\boldsymbol{\phi}_d$ in Fig. 4(b), and $\boldsymbol{\phi}_c$ in Fig. 4(a) which are related to impact (i.e., attack) properties of the sounds.

While ICA can model the data using nonorthogonal basis functions, PCA models the data with orthogonal bases. Consequently, the temporal basis functions learned by PCA can differ from those learned by ICA. Figure 5 illustrates the results obtained by PCA of the spectrogram of the sound of an impact on an aluminum rod. This figure shows that the dominant basis function, $\boldsymbol{\phi}_1$, has a much smoother shape than the other basis functions. This basis function shapes the overall decay of all partials. In fact, the results show that PCA extracts a dominant basis function $\boldsymbol{\phi}_1$ that represents most of the temporal structure of the sound (Fig. 6). On average, this basis function accounts for more than 68% of the temporal variation in $\mathbf{S}$. This property of the dominant basis function is due to the lack of variation in the spectral structure of the sound over time. (As an example of this regularity, Fig. 1 shows that there is not much variation in which partials are active over time.) $\boldsymbol{\phi}_1$ has the ability to account for the temporal behavior of this spectral structure. To illustrate this point, Fig. 7 shows the average power spectrum of a sound from an impact on an aluminum rod and spectral source signal $\mathbf{c}_1^{\mathrm{Al2}}$ obtained by PCA of the spectrogram of this sound. $\boldsymbol{\phi}_1$ describes the temporal behavior of spectra $\mathbf{c}_1^{\mathrm{Al2}}$, which, as can be seen in Fig. 7, is very similar to the sound's power spectrum, which represents the spectral structure of the sound over time.

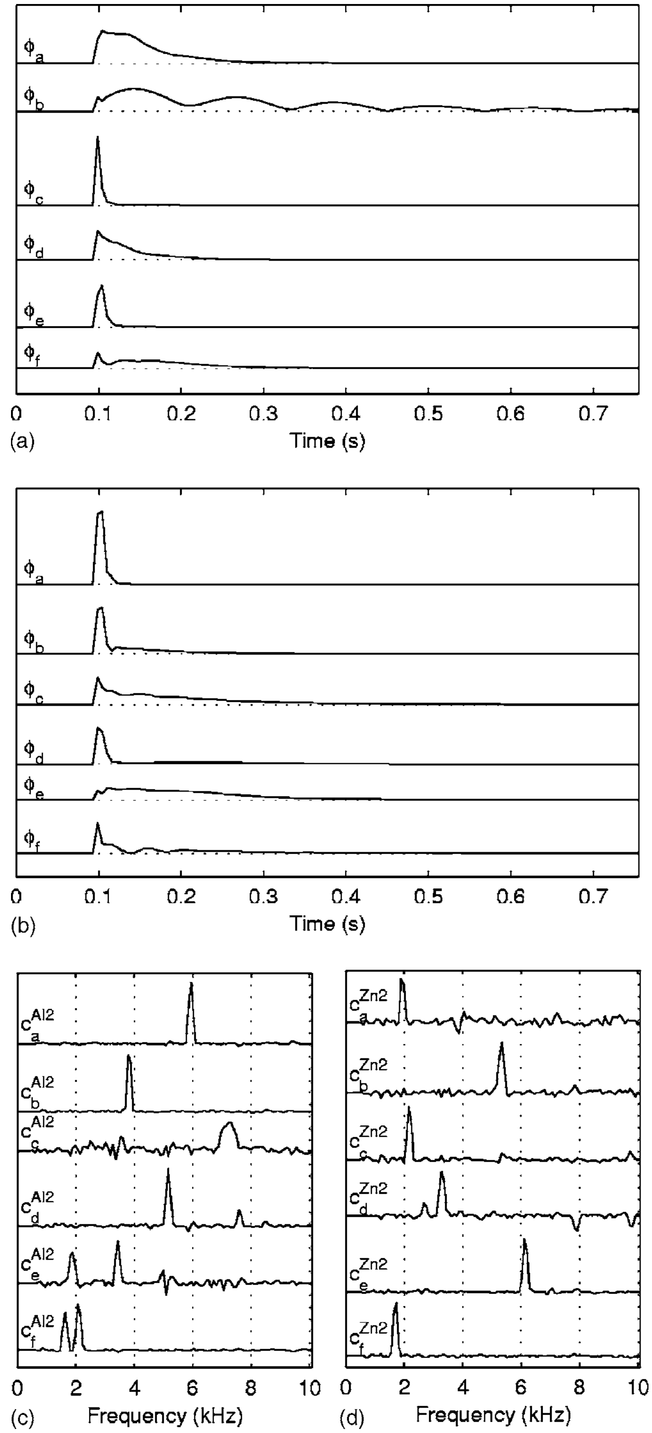

FIG. 4. Temporal basis functions $\mathbf{\Phi}$ learned by ICA of the spectrogram of: (a) a sound (Al2) from an impact on an aluminum rod; (b) a sound (Zn2) from an impact on a zinc plated steel rod. In each case, six out of the ten most dominant basis functions are shown in decreasing order of dominance from top to bottom. The corresponding spectral source signals for Al2 (c) and Zn2 (d) are shown also from top to bottom.

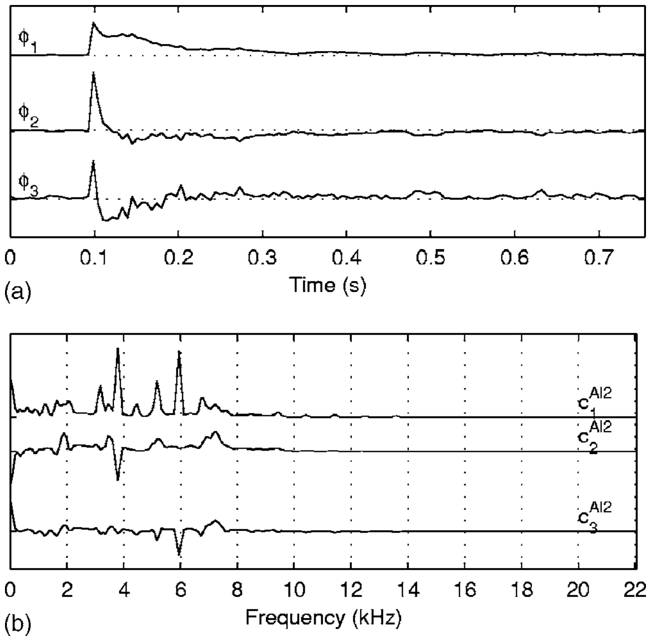

FIG. 5. Temporal basis functions $\boldsymbol{\Phi}$ and spectral source signals ($\mathbf{C}^{\mathrm{Al2}}$) obtained by PCA of the spectrogram of a single sound (Al2) from an impact on an aluminum rod. (a) The first three basis functions are shown from top to bottom. (b) The first three spectral source signals are shown from top to bottom.

Other less significant basis functions account for temporal behaviors that differ from the overall decay shape described by $\boldsymbol{\phi}_1$. For example, the temporal shapes of $\boldsymbol{\phi}_2$ and $\boldsymbol{\phi}_3$ account for variations in the temporal behavior of subspectra $\mathbf{c}_2^{\mathrm{Al2}}$ and $\mathbf{c}_3^{\mathrm{Al2}}$ (Fig. 5). (Note also that these subspectra contain common partials with the spectral structure of the sound, but, as can be easily seen in this figure, they account for much less of the spectral structure of $\mathbf{S}$ than $\mathbf{c}_1^{\mathrm{Al2}}$ does. The same is true for other sounds. The less variance a basis function accounts for, the fewer partials its spectral source signal shares with $\mathbf{S}$.) In contrast to what was seen with ICA, these basis functions are not as directly related to temporal properties of the sounds. (Note that since the same sound, Al2, was used in both Figs. 4(a) and 5, these are directly comparable.)

As seen earlier, ICA obtains a greater variety of basis function shapes: some are similar to the most significant PCA basis functions, but ICA is also able to learn basis functions that capture structures besides decay. In fact, there seems to be a more direct relation between the shape of the basis functions learned by ICA and temporal properties like ringing, resonance, decay, impact (or attack), etc. As a consequence, ICA needs more basis functions to explain the variance of $\mathbf{S}$ (Fig. 6). On average, the most significant basis function ($\boldsymbol{\phi}_1$) accounts only for about 27% of the temporal variation in $\mathbf{S}$ compared to 68% for PCA.

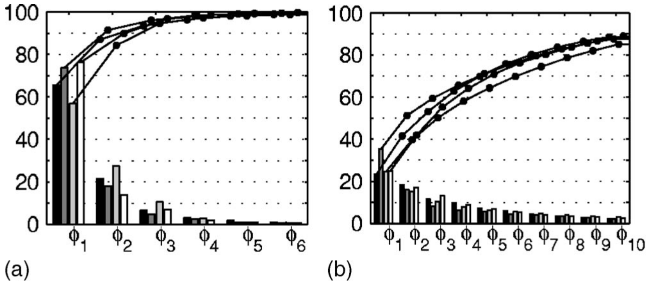

FIG. 6. Percentage of variance explained by the basis functions in $\boldsymbol{\Phi}$. The spectrograms from ten impact sounds from each rod (aluminum in black, zinc plated steel in dark grey, steel in light grey, and wood in white) were used. $\boldsymbol{\Phi}$ was learned by spectral analysis on one spectrogram at a time. The ten results obtained for each rod were averaged. Only the values for the first six or ten temporal basis functions are shown. The dots on the curves show the cumulative sums of the percentages. In (a) $\boldsymbol{\Phi}$ was learned by PCA. In (b) $\boldsymbol{\Phi}$ was learned by ICA.

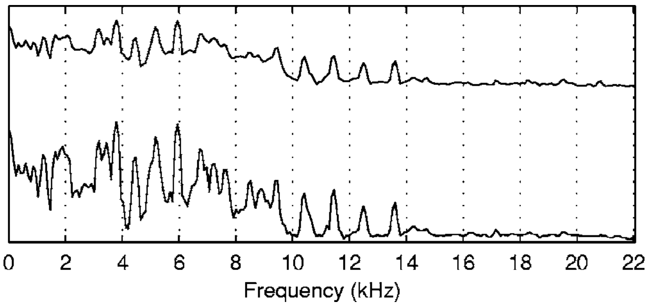

FIG. 7. Power spectrum of a sound (Al2) from an impact on an aluminum rod. The bottom line shows the power spectrum. The line on the top shows spectral source signal $\mathbf{c}_1^{\mathrm{Al2}}$ found by PCA of the spectrogram of this sound. Note how both lines show high energy on the same partials. (Here, the source signal $\mathbf{c}_1^{\mathrm{Al2}}$ looks different than in Fig. 5 because it is plotted in a logarithmic scale.)

Up to this point, we have considered the basis functions; now we will consider the spectral source signals. Because here we consider the spectrogram $\mathbf{S}$ of a single sound, there is only one spectral source signal $\mathbf{c}_i^k$ associated with each basis function $\boldsymbol{\phi}_i$. This source signal consists of the partials that have the time varying shape described by $\boldsymbol{\phi}_i$. In other words, the source signals consist of partials that have similar time varying shape. Unlike the source signal of the dominant basis function obtained by PCA, with ICA there is no source signal that accounts for most of the spectral structure in $\mathbf{S}$. ICA separates partials with different time varying shapes into different spectral source signals, which is better suited to represent the variability in the sounds. This point is illustrated by Figs. 4(c) and 4(d), which show the spectral source signals for six out of the ten most dominant basis functions obtained by spectral ICA. As can be seen, when ICA is used, the partials in one spectral source signal are typically not present in the remaining source signals. From another perspective, ICA learns basis functions that more directly relate to the underlying acoustic properties. This desirable effect allows ICA to extract more interesting temporal structures of the sounds than those seen with PCA.

2. Ensemble of impact sounds

We will now consider the more general case of an ensemble of impacts on the same rod. In this case, the data matrix is defined over a set of K sounds aligned at time zero. The result of applying spectral ICA or PCA to this data is a set of temporal basis functions $\boldsymbol{\Phi}$ and K sets of spectral source signals $\mathbf{C}^k$. The temporal basis functions $\boldsymbol{\Phi}$ model the common temporal properties of the sounds, and each set of spectral source signals $\mathbf{C}^k$ represents the spectra of sound k

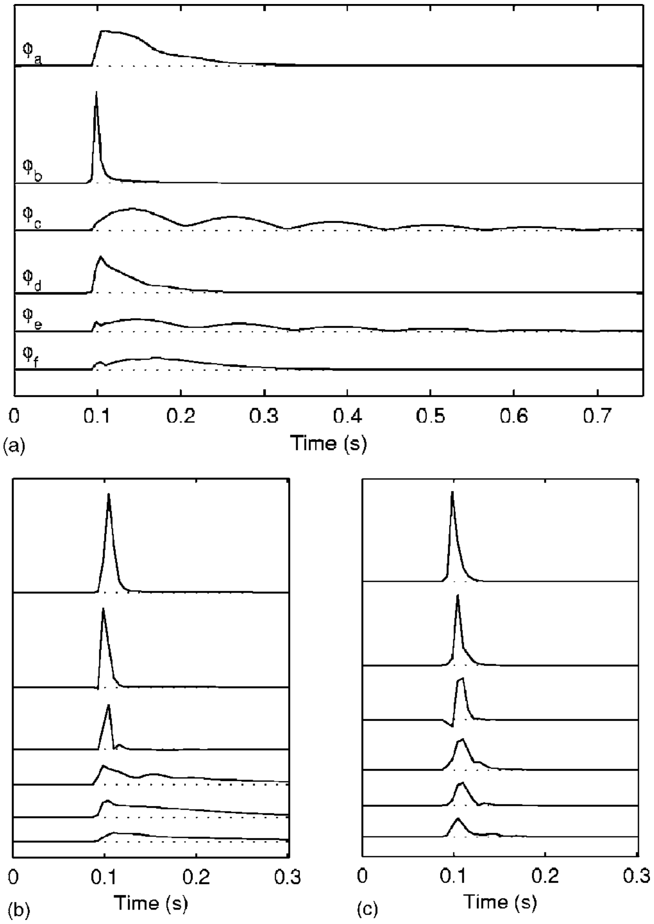

FIG. 8. Temporal basis functions $\mathbf{\Phi}$ learned by ICA of the set of: (a) ten sounds from impacts on an aluminum rod; (b) ten sounds from impacts on a zinc plated steel rod; and (c) ten sounds from impacts on a wooden rod. In each figure, six out of the ten most dominant basis functions are shown in decreasing order of dominance from top to bottom.

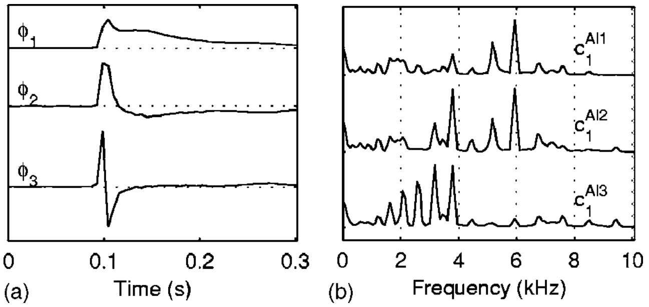

FIG. 9. Temporal basis functions $\mathbf{\Phi}$ and spectral source signals $\mathbf{C}^k$ obtained by PCA of the set of ten impacts on an aluminum rod. (a) The first three basis functions are shown from top to bottom. (b) The first spectral source signal for sounds Al1, Al2, and Al3.

that have the temporal properties described by $\mathbf{\Phi}$. The spectral source signals (say $\mathbf{c}_i^{k_1}$ and $\mathbf{c}_i^{k_2}$) associated with the same basis function $\boldsymbol{\phi}_i$ are the subspectra (of sounds k_1 and k_2, respectively) that share the temporal property described by $\boldsymbol{\phi}_i$.

The results for multiple impacts resemble those for a single impact due to the similarity in the underlying acoustic structure across impacts. This is clear with the basis functions learned by ICA, for instance, compare $\boldsymbol{\phi}_a$ in Figs. 4(a) and 8(a), and is particularly obvious with the most dominant basis function learned by PCA, for instance, compare the first line from Figs. 5(a) and 9(a). Even though the temporal basis functions in these figures are not exactly the same, they have very similar shapes.

Because more impacts on the same rod imply more variability, some acoustic structures that were represented by a single basis function in Section IV A 1, are now represented by multiple basis functions. For example, the ringing structure represented by $\boldsymbol{\phi}_b$ in Fig. 4(a) is now represented by both $\boldsymbol{\phi}_c$ and $\boldsymbol{\phi}_e$ in Fig. 8(a). In order to illustrate how the temporal variability is represented, we will examine these two basis functions more carefully. By inspecting the spectral source signals associated with $\boldsymbol{\phi}_c$ and $\boldsymbol{\phi}_e$ (see second and third plots in the middle column of Fig. 10) we can conclude that these two basis functions represent the temporal behavior of the partial at 3.95 kHz. In some impacts this partial has a temporal shape that is more closely described by $\boldsymbol{\phi}_c$ (observe that for Al1 there is a peak in $\mathbf{c}_c^{\mathrm{Al1}}$ but not in $\mathbf{c}_e^{\mathrm{Al1}}$), while in other impacts the partial's temporal shape is more closely described by $\boldsymbol{\phi}_e$ (for Al3 there is a peak in $\mathbf{c}_e^{\mathrm{Al3}}$ but not in $\mathbf{c}_c^{\mathrm{Al3}}$). Still in other impacts a mixture of both $\boldsymbol{\phi}_c$ and $\boldsymbol{\phi}_e$ is required to describe the partial's temporal shape (Al2 has a peak in both $\mathbf{c}_e^{\mathrm{Al2}}$ and $\mathbf{c}_c^{\mathrm{Al2}}$).

Even though on average the basis functions account for a smaller percentage of variance than in Sec. IV A 1 and more basis functions are needed to explain the same percentage of variance, the difference is not significant. For instance, the ten most dominant basis functions learned by spectral ICA of a single sound account for at most 88% of the variance, while when a set of ten sounds is used, the same number of basis functions explains at most 84% of the variance of the data.[4] Spectral PCA shows similar results: six basis functions suffice to explain around 99% of the variance on a single sound, while for a set of ten sounds, six basis functions can explain around 96% of the variance (Figs. 6 and 11).

The results shown here were obtained from sounds recorded in an anechoic chamber; however, we also tested the model with sounds recorded in a normal room (with background noise and reverberation). In this case, $\mathbf{S}$ represented not only the structure of the sound, but also the structure of the background noise. Consequently, apart from the temporal basis functions that accounted for the temporal structure of the sound, spectral PCA and ICA also learned some basis functions that described the temporal structure of the background noise (data not shown).

The results are dependent on the sounds analyzed. Figure 8 shows that impacts on different rods are characterized by different basis functions. For instance, some of the basis functions that characterize impacts on aluminum have a longer duration than the basis functions that characterize impacts on wood. If sounds with different characteristics are used, the basis functions will reflect those characteristics.

B. Spectral basis functions $\mathbf{\Psi}$

The sets of spectral source signals $\mathbf{C}^1, \ldots, \mathbf{C}^K$ represent the subspectra associated with the temporal basis functions in $\mathbf{\Phi}$. Even though each $\mathbf{C}^k$ is specific to an individual sound

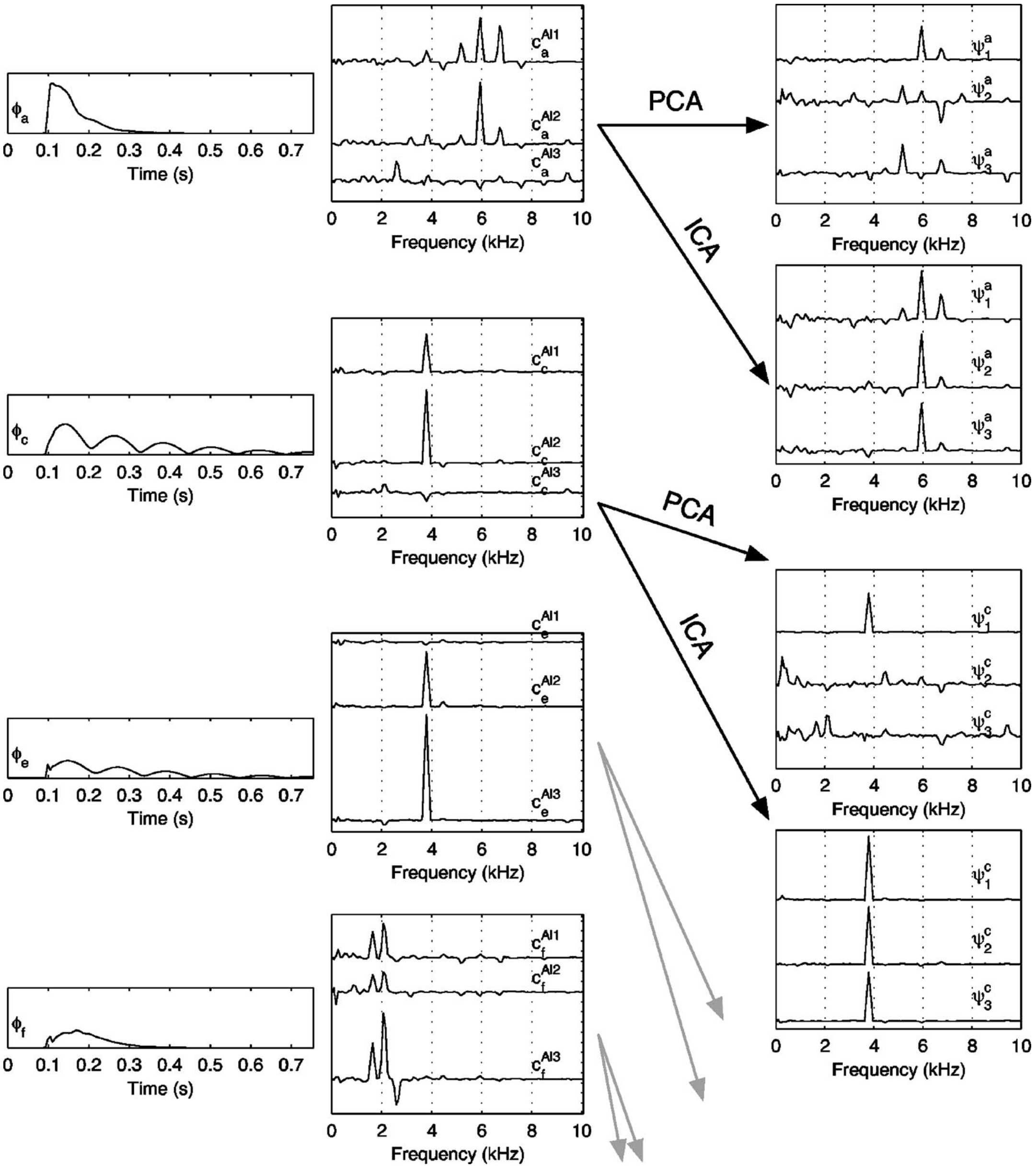

FIG. 10. Left column: Temporal basis functions $\boldsymbol{\phi}_a$, $\boldsymbol{\phi}_c$, $\boldsymbol{\phi}_e$, and $\boldsymbol{\phi}_f$ from Fig. 8(a). These are learned by ICA of the set of ten sounds from impacts on an aluminum rod. Middle column: The corresponding spectral source signals for sounds Al1, Al2, and Al3. Right column: Spectral basis functions $\boldsymbol{\Psi}$ obtained by analysis of the spectral source signals. The first and third figures in this column show the first three spectral basis functions from $\boldsymbol{\Psi}^a$ and $\boldsymbol{\Psi}^c$ learned by PCA. The second and fourth figures in this column show the first three spectral basis functions from $\boldsymbol{\Psi}^a$ and $\boldsymbol{\Psi}^c$ learned by ICA.

k, the sets of source signals do share common structures. This can be seen in Fig. 10. The middle column shows the spectral source signals obtained by spectral ICA of the set of ten sounds from an aluminum rod. Although the source signals show considerable variability, there is still much common structure. The same observations can be made on the results from spectral PCA [see Figs. 9(b) and 12(b)].

As explained in Sec. II, we can extend the approach to model the regularities in the spectral source signals. In the extended model, these regularities are represented by the set of spectral basis functions $\boldsymbol{\Psi}$, which is learned by applying PCA or ICA to matrices of spectral source signals, such that $\boldsymbol{\Psi}^i$ consists of the spectral basis functions that represent the regularities of the source signals associated with the tempo-

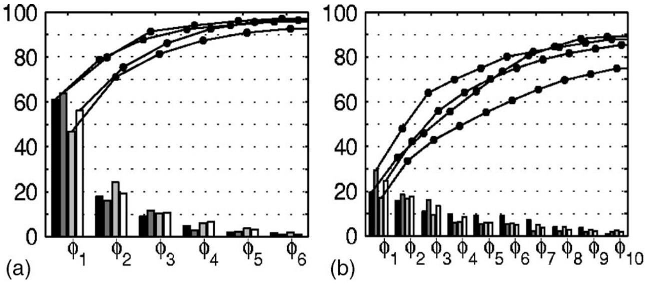

FIG. 11. Percentage of variance explained by the basis functions in $\boldsymbol{\Phi}$ learned by spectral analysis on the set of ten impacts on an aluminum rod (black), the set of ten impacts on a zinc plated steel rod (dark grey), the set of ten impacts on a steel rod (light grey), and the set of ten impacts on a wooden rod (white). Only the values for the first six or ten temporal basis functions are shown. The dots on the curves show the cumulative sums of the percentages. In (a) $\boldsymbol{\Phi}$ was learned by PCA. In (b) $\boldsymbol{\Phi}$ was learned by ICA.

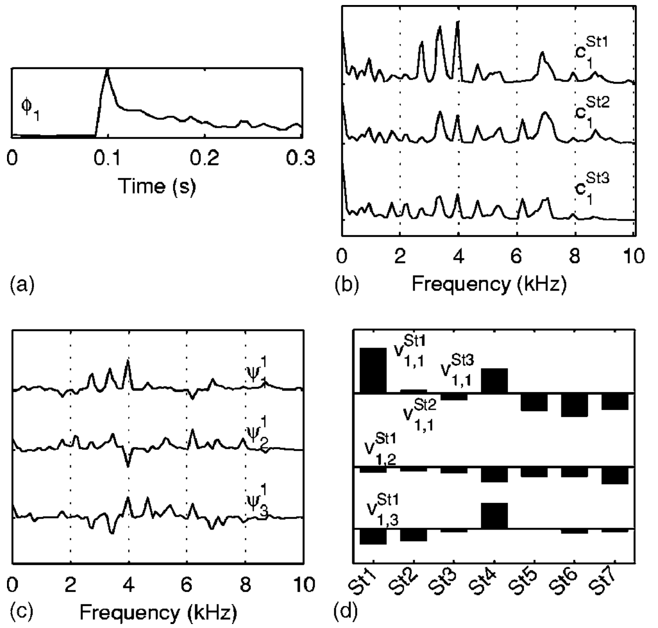

FIG. 12. Top row: Temporal basis functions $\mathbf{\Phi}$ and spectral source signals $\mathbf{C}^k$ obtained by spectral PCA of the set of ten impacts on a steel rod. (a) First (most dominant) basis function. (b) First spectral source signal for sounds St1, St2, and St3. Bottom row: Spectral basis functions $\boldsymbol{\Psi}$ and coefficients $\mathbf{V}^k$ (for $k \in \{\text{St1}, \text{St2}, \ldots, \text{St7}\}$) obtained by PCA of the source signals. (c) First three spectral basis functions from $\boldsymbol{\Psi}^1$. (d) Coefficients for spectral basis functions $\boldsymbol{\psi}_1^1$ to $\boldsymbol{\psi}_3^1$. The jth line, kth column shows $v_{1,j}^k$, that is, the coefficient for sound k and basis function $\boldsymbol{\psi}_j^1$.

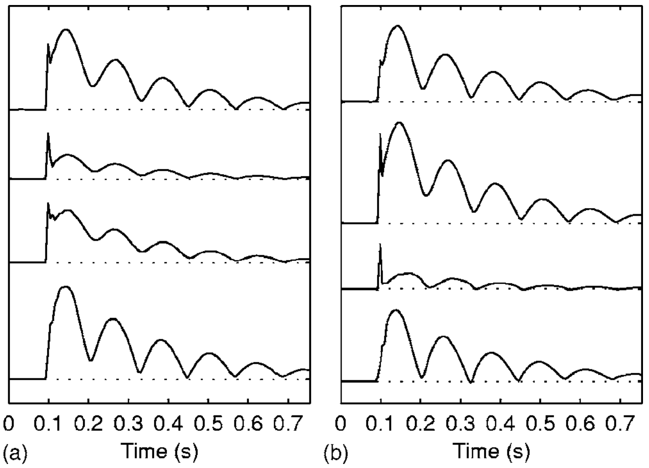

FIG. 13. The decay shape of the partial at 3.95 kHz (which is partial e in Figs. 1 and 2) from different spectrograms of impacts on an aluminum rod. (a) The original partials show considerable variability. The partials (from top to bottom) were extracted from the spectrograms of sounds Al2, Al4, Al9, and Al10. (The partial from Al10 looks different in Fig. 3 because there it was plotted in a logarithmic scale.) (b) The synthesized partials have a similar range of variability. The partials were extracted from four synthesized spectrograms. See the text for details.

ral basis function $\boldsymbol{\phi}_i$, that is, the regularities of source signals $\mathbf{c}_i^1, \ldots, \mathbf{c}_i^K$. ($\boldsymbol{\Psi}^i$ contains basis functions $\boldsymbol{\psi}_1^i, \ldots, \boldsymbol{\psi}_J^i$, and $\boldsymbol{\Psi}$ contains sets $\boldsymbol{\Psi}^1, \ldots, \boldsymbol{\Psi}^I$.)

Figure 12 shows the results obtained by PCA of the spectral source signals from PCA of the set of ten impacts on a steel rod.[5] Since PCA models the data with orthogonal bases, all basis functions within each set $\boldsymbol{\Psi}^i$ are orthogonal. Comparing $\boldsymbol{\psi}_1^1$ with $\mathbf{c}_1^{\text{St1}}$, it can be seen that the energy found in spectrum $\mathbf{c}_1^{\text{St1}}$ is being represented by this spectral basis function. For instance, note the three peaks between 2 and 4 kHz in both lines. Even though $\mathbf{c}_1^{\text{St2}}$ and $\mathbf{c}_1^{\text{St3}}$ have peaks in the same region, they show less energy in these partials. This variability is accounted for in part by other spectral basis functions in $\boldsymbol{\Psi}^1$ and in part by $\mathbf{V}^k$. Note how $v_{1,1}^{\text{St1}}$ has a much higher value than $v_{1,1}^{\text{St2}}$ and $v_{1,1}^{\text{St3}}$.

PCA can also be applied to the spectral source signals that have been obtained by spectral ICA. The first and third graphs in the right column of Fig. 10 show the results obtained by PCA of the spectral source signals from ICA of the set of ten impacts on the same aluminum rod. The set of spectral basis functions $\boldsymbol{\Psi}^a$ represents the regularities of the spectral source signals associated with $\boldsymbol{\phi}_a$. For instance, note how $\boldsymbol{\psi}_1^a$ represents the peaks close to 6 and 7 kHz, which can be seen in $\mathbf{c}_a^{\text{Al1}}$ and $\mathbf{c}_a^{\text{Al2}}$. Since these peaks are much lower (or negative) in $\mathbf{c}_a^{\text{Al3}}$, $v_{a,1}^{\text{Al3}}$ has a much lower value than $v_{a,1}^{\text{Al1}}$ and $v_{a,1}^{\text{Al2}}$ (these coefficients are not shown here).

The number of basis functions considered is arbitrary and depends on the application. It depends on how much of the structure of the sounds one needs to model. To completely represent the structure of the sounds, we need to be able to model all variability in all spectral source signals $\mathbf{c}_i^k$, and, therefore, we must consider all basis functions in $\boldsymbol{\Psi}$. However, the results show that when PCA is used there is a dominant component in each set $\boldsymbol{\Psi}^i$ that represents most of the structure in the spectral source signals $\mathbf{c}_i^k$. Thus, often a very good approximation of source signals $\mathbf{c}_i^k$ can be obtained by using a small subset of $\boldsymbol{\Psi}^i$.

Finally, we show some results from ICA of the spectral source signals that have been obtained by spectral ICA. Since the spectral basis functions learned by ICA are not restricted to be orthogonal, and not many sounds (and consequently not many spectral source signals)were used in this study, the basis functions $\boldsymbol{\Psi}$ learned by ICA are more tuned to specific spectral source signals, that is, they resemble more closely the shape of specific spectral source signals. As a consequence, the representations obtained by ICA are less compact than the representations obtained by PCA. The second and fourth graphs in the right column of Fig. 10 show the results obtained by ICA of the spectral source signals from spectral ICA of the set of ten impacts on an aluminum rod. It is interesting to note the similarities between individual spectral basis functions and spectral source signals. For instance, compare $\boldsymbol{\psi}_1^a$ to $\mathbf{c}_a^{\text{Al1}}$, and $\boldsymbol{\psi}_2^a$ to $\mathbf{c}_a^{\text{Al2}}$. See also how similar $\boldsymbol{\psi}_1^c$, $\boldsymbol{\psi}_2^c$, and $\boldsymbol{\psi}_3^c$ are.

C. Variability

Natural sounds have significant variability as was illustrated in Figs. 1–3. Because model M_b is adapted to represent the distribution of the ensemble of impact sounds, it also captures this variability. The variability is represented by different basis functions (like $\boldsymbol{\phi}_c$ and $\boldsymbol{\phi}_e$ in Fig. 10) and by the distribution of the coefficients $v_{i,j}^k$. To illustrate this, Fig. 13 shows that by giving different values to the coefficients $v_{i,j}^k$, one can use different combinations of the temporal and spectral structures represented by the basis functions in $\mathbf{\Phi}$ and $\boldsymbol{\Psi}$ to simulate the variability present in the sounds. By randomly sampling the coefficients $v_{i,j}^k$, we can generate different instances from the model distribution. Figure 13(a) shows the variation in the partial at 3.95 kHz, and Fig. 13(b) shows four different synthesis instances of the same partial (3.95 kHz), each extracted from a different synthesized spectrogram. To synthesize the spectrograms, we used the temporal basis functions ($\mathbf{\Phi}$) learned by spectral ICA of the spectrograms of ten aluminum rod impacts, the spectral basis functions ($\mathbf{\Psi}$) learned by PCA of the corresponding spectral source signals, and the coefficients obtained for one of the sounds ($\mathbf{V}^{\mathrm{Al4}}$). To simulate the variability caused by $\boldsymbol{\phi}_c$ and $\boldsymbol{\phi}_e$ (from Fig. 10) we varied the weightings of these two basis functions. That was done by varying the values of $\mathbf{v}_c^{\mathrm{Al4}}$ and $\mathbf{v}_e^{\mathrm{Al4}}$ for each synthesized spectrogram. The values were randomly sampled from the coefficient's distribution. (Note that in this way we are also varying the weightings of $\mathbf{\Psi}^c$ and $\mathbf{\Psi}^e$.) Figure 13 confirms that model M_b is suited to represent the natural variability of the sounds. The variations obtained by the model are similar to those in the ensemble of impact sounds [compare the variations in Fig. 13(a) to those in Fig. 13(b)].

V. DISCUSSION AND CONCLUSIONS

Our main goal here was to develop a data-driven method for learning a representation of the intrinsic structures of impact sounds. We showed that, by using PCA and ICA, it is possible to build a model that uses temporal and spectral basis functions that represent the intrinsic temporal and spectral structures of the sounds. The method can be used to characterize the structures of a single sound or the structures common to a set of impact sounds, in which case it also captures the natural variability in the structures. Obviously, if the method receives different inputs, it produces different outputs, but if the sounds are of the same type, the structures that the method learns are comparable. For instance, the results of analyzing one sound versus several sounds of the same type are very similar. The model does not require any *a priori* knowledge of the physics, acoustics, or dynamics of the objects and events and is able to represent the underlying acoustical structures in the sounds, which could offer advantages over previous knowledge-based models.

The temporal structures of the sounds are represented by the temporal basis functions $\mathbf{\Phi}$, which are learned by spectral analysis of the spectrograms. The spectral structures of the sounds are represented by the spectral basis functions $\mathbf{\Psi}$, which are obtained in a second step by PCA or ICA of the spectral source signals associated with the temporal basis functions $\mathbf{\Phi}$.

Spectral ICA is able to decompose spectrograms into a small number of underlying features (represented in the temporal basis functions $\mathbf{\Phi}$) that characterize acoustic properties such as ringing, resonance, sustain, decay, and onsets. Since the method is not restricted to learn explicit features (or structures) of the sounds, the representations obtained include new information that was not represented by previous (physical) models. For instance, features that are more abstract than simple decay rate or average spectra, like features that characterize ringing, or decay shapes that are not exponential, can now be modeled and easily extracted from the sounds. Spectral PCA gives compact representations of the temporal structures in the spectrograms. For instance, six basis functions can explain 96% or more of the variance of the data (see Sec. IV A 2). Such low dimensional characterizations of the data can present advantages over previous physical models. For example, since impact sounds can have hundreds of partials (van den Doel *et al.*, 2002), modeling them with Eq. (1) would mean using a very big N. When the objective is to model only the perceptually relevant portions of the sound, many less partials can be used (that is, N can be substantially smaller), yet determining which partials should be used is also a difficult question (van den Doel *et al.*, 2002).

Brown and Smaragdis (2004) have used ICA to separate different notes from two-note musical trills. In another study the same authors have used non-negative matrix factorization (NMF), which is another redundancy reduction technique, to analyze polyphonic musical passages (Smaragdis and Brown, 2003). Although these approaches are related to those presented here, their goal was to separate notes from musical segments with more than two notes. Even though the analyses used in both these studies resemble our analysis method, there are some fundamental differences. The main difference is that we are partitioning individual sounds according to the temporal behavior of the partials, whereas in their studies the sounds are being segmented according to events; we are interested in representing the structure of sounds of the same type efficiently, whereas they are interested in segmenting sound events. Also, while their analyses are appropriate for highly harmonic sounds, transient sounds with high structure variability are better described by our method, given that here individual sounds are represented by more than one temporal and spectral basis functions.

Most work with redundancy reduction techniques (like ICA, PCA, NMF, singular value decomposition, and sparse coding) and spectrograms or other time-frequency structures (like constant Q-transforms and wavelets), focus on the source separation problem, and, as with the above-mentioned two studies, it segments sounds according to events (e.g., Barros *et al.*, 2002; Casey and Westner, 2000; Smaragdis, 2004; Virtanen, 2004). Some MPEG-7 audio features are obtained using similar techniques, and there has been work on sound classification, recognition, and event detection using these features (e.g., Kim *et al.*, 2004; Xiong *et al.*, 2003). All these studies use techniques similar to those used in the method presented here, but their goals are very different and, to the best of our knowledge, the method presented is the first to partition individual sounds according to the temporal behavior of the partials. Even though this paper does not discuss sound classification and recognition, the basis functions learned by the method presented here, may be particularly useful to such applications.

Although here we have only considered impact sounds, namely impacts on rods, we predict that this model can be used to represent other types of transient acoustic events. The work presented considers only the spectral content of the signals. Nonetheless, there is also complex structure in the phase of the signals, which is important for synthesizing sound waveforms from the model.

ACKNOWLEDGMENTS

This work was supported in part by fellowships from Fundação Calouste Gulbenkian (Portugal) and Fundação para a Ciência e a Tecnologia (Portugal) to S.C. and a grant from the National Science Foundation to M.S.L.

APPENDIX

See EPAPS Document No. E-JASMAN-121–046706 for Appendices A, B, and C. This document can be reached via a direct link in the online article's HTML reference section or via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html).

[1] Here matrices are represented in bold upper case, vectors, which are column vectors unless the transpose is used, are represented in bold lower case, and scalars are represented in lower case. The horizontal concatenation of matrices **A** and **B** is $(\mathbf{A},\mathbf{B})$.

[2] We use upper indexes to distinguish different variables of the same type, so for instance $\mathbf{X}^1$ and $\mathbf{X}^2$ are two different matrices of the same type. Lower indices are used to index values within a matrix or vector.

[3] In order to make the graphs more readable, some of the basis functions $\boldsymbol{\phi}_i$ and corresponding spectral source signals $\mathbf{c}_i^k$ have been multiplied by -1.

[4] Since the basis functions given by PCA are orthogonal, the sum of the variances that they explain gives the total variance explained. However, the same is not true for the basis functions given by ICA, which are not restricted to being orthogonal. In this case, the sum of the variances may correspond to a quantity that is bigger than the actual variance explained by the basis functions.

[5] In order to make the graphs more readable, some of the basis functions $\boldsymbol{\psi}_j^j$ and corresponding coefficients $v_{i,j}^k$ have been multiplied by -1.

Avanzini, F., and Rocchesso, D. (**2001a**). "Controlling material properties in physical models of sounding objects," Proceedings of the International Computer Music Conference 2001, La Habana, Cuba, pp. 91–94.

Avanzini, F., and Rocchesso, D. (**2001b**). "Modeling collision sounds: Nonlinear contact force," in Proceedings of the COST G-6 Conference on Digital Audio Effects, 2001, Limerick, Ireland.

Barros, A. K., Rutkowski, T., Itakura, F., and Ohnishi, N. (**2002**). "Estimation of speech embedded in a reverberant and noisy environment by independent component analysis and wavelets," IEEE Trans. Neural Netw. **13**, 888–893.

Brown, J., and Smaragdis, P. (**2004**). "Independent component analysis for automatic note extraction from musical trills," J. Acoust. Soc. Am. **115**, 2295–2306.

Casey, M. A., and Westner, A. (**2000**). "Separation of mixed audio sources by independent subspace analysis," Proceedings of the International Computer Music Conference, Berlin, Germany.

Gaver, W. W. (**1994**). *Using and Creating Auditory Icons*, Auditory Display: Sonification, Audification and Auditory Interfaces, editied by G. Kramer (Addison-Wesley, Reading, MA), pp. 417–446.

Gaver, W. W. (**1988**). "Everyday listening and auditory icons," Ph.D. thesis, University of California at San Diego, San Diego, CA.

Hyvärinen, A., Karhunen, J., and Oja, E. (**2001**). *Independent Component Analysis* (Wiley, New York).

James, D. L., Barbič, J., and Pai, D. K. (**2006**). "Precomputed acoustic transfer: Output-sensitive, accurate sound generation for geometrically complex vibration sources," ACM Trans. Graphics **25**, 987–995.

Kim, H. G., Berdahl, E., and Sikora, T. (**2004**). "Study of MPEG-7 sound classification and retrieval," Fifth International ITG Conference on Source and Channel Coding, Erlangen, Germany.

Lambourg, C., Chaigne, A., and Matignon, D. (**2001**). "Time-domain simulation of damped impacted plates. II. Numerical model and results," J. Acoust. Soc. Am. **109**, 1433–1447.

O'Brien, J. F., Cook, P. R., and Ess, G. (**2001**). "Synthesizing sounds from physically based motion," Proceedings of ACM SIGGRAPH (Los Angeles, California) pp. 529–536.

O'Brien, J. F., Shen, S., and Gatchalian, C. M. (**2002**). "Synthesizing sounds from rigid-body simulations," in Proceedings of ACM SIGGRAPH Symposium on Computer Animation (San Antonio, Texas) pp. 175–181.

Pai, D. K., van den Doel, K., James, D. L., Lang, J., Lloyd, J. E., Richmond, J. L., and Yau, S. H. (**2001**). "Scanning physical interaction behavior of 3d objects," Proceedings of ACM SIGGRAPH (Los Angeles, California) pp. 87–96.

Smaragdis, P. (**2004**). "Non-negative matrix factor deconvolution; Extraction of multiple sound sources from monophonic inputs," Proceedings of the Fifth International Conference on Independent Component Analysis and Blind Signal Separation, Lecture Notes in Computer Science, edited by C. G. Puntonet and A. Prieto (Granada, Spain) pp. 494–499.

Smaragdis, P., and Brown, J. (**2003**). "Non-negative matrix factorization for polyphonic music transcription," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 177–180.

Stone, J. V. (**2004**). *Independent Component Analysis, A Tutorial Introduction* (MIT, Cambridge, MA).

van den Doel, K., Kry, P. G., and Pai, D. K. (**2001**). "Foleyautomatic: Physically-based sound effects for interactive simulation and animation," Proceedings of ACM SIGGRAPH, Los Angeles, California.

van den Doel, K., Pai, D. K., Adam, T., Kortchmar, L., and Pichora-Fuller, K. (**2002**). "Measurements of perceptual quality of contact sound models," Proceedings of the International Conference on Auditory Display (Kyoto, Japan) pp. 345–349.

Virtanen, T. (**2004**). "Separation of sound sources by convolutive sparse coding," Proceedings of the ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing, 3 Oct., Jeju, Korea.

Xiong, Z., Radhakrishnan, R., Divakaran, A., and Huang, T. (**2003**). "Audio events detection based highlights extraction from baseball, golf and soccer games in a unified framework," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).