Supplementary material for: Statistical modeling of intrinsic structures in impacts sounds

Sofia Cavaco

Computer Science Department, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213

Michael S. Lewicki

Computer Science Department and Center for the
Neural Basis of Cognition, Carnegie Mellon University
4400 Fifth Avenue, Pittsburgh, PA 15213

APPENDIX A: THE MODEL

This appendix gives a more detailed description of the model presented in Sec. II. Here we include illustrations of the matrices used in the model and details on matrix rearrangements.

The spectrograms can be defined as an ordered set of T bins or as a sequence of F frames. Model M_b considers the spectrogram of sound k, that is $\mathbf{S}^k$, as an ordered set of bins. In order to express the spectrogram in this way, we will assume we have a set of I basis functions $\mathbf{\Phi}$, which is represented as a matrix of size $(T \times I)$ whose column vectors are the temporal basis functions $\boldsymbol{\phi}_i$. We will also assume that for each sound k we have a set of *spectral source signals* represented as a matrix $\mathbf{C}^k$ of size $(I \times F)$, where the ith row contains vector $(\mathbf{c}_i^k)^T$. This spectral signal scales basis function $\boldsymbol{\phi}_i$ across frequencies. (Section IV.A shows how to find $\mathbf{\Phi}$ and $\mathbf{C}^k$.)

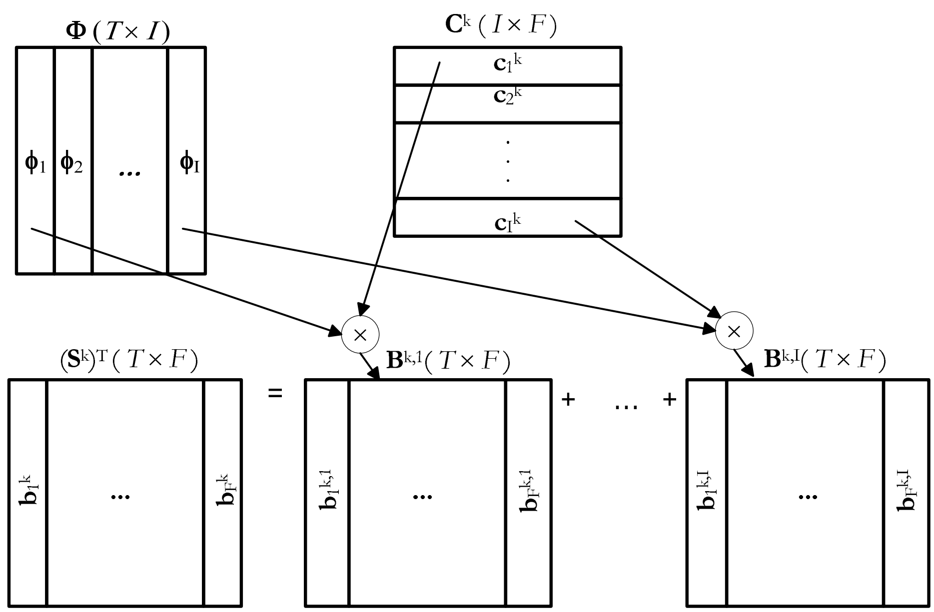


FIG. 14. Illustration of Eq. (A1). (For simplicity, here we did not mark row vectors with the transpose symbol T.)

By linearly combining a basis function with the corresponding spectral source signals, we obtain a $(T \times F)$ matrix $\mathbf{B}^{k,i}$ that represents part of the structure in $\mathbf{S}^k$. The whole structure in $\mathbf{S}^k$ is obtained by linearly combining matrices $\mathbf{B}^{k,i}$ (see Fig. 14):

$$(\mathbf{S}^k)^T = \sum_{i=1}^{I} \mathbf{B}^{k,i} = \sum_{i=1}^{I} \boldsymbol{\phi}_i \, (\mathbf{c}_i^k)^T \; = \boldsymbol{\Phi} \, \mathbf{C}^k \,. \tag{A1}$$

If we consider only the fth column (or bin) in $\mathbf{B}^{k,i}$ (i.e., $\mathbf{b}_f^{k,i}$), Eq. (A1) can be rewritten as follows:

$$\mathbf{b}_f^k = \sum_{i=1}^{I} \mathbf{b}_f^{k,i} = \sum_{i=1}^{I} \boldsymbol{\phi}_i \, c_{i,f}^k \,. \tag{A2}$$

where $\mathbf{b}_f^k$ is the transpose of the fth bin of $\mathbf{S}^k$, and the scalar $c_{i,f}^k$ is the value of $\mathbf{c}_i^k$ at frequency bin f. The basis $\boldsymbol{\Phi}$ can be used to describe a single sound, as in Eq. (A1), or the temporal regularities of a set of related sounds:

$$((\mathbf{S}^1)^T, (\mathbf{S}^2)^T, \ldots, (\mathbf{S}^K)^T) = \boldsymbol{\Phi} \, (\mathbf{C}^1, \mathbf{C}^2, \ldots, \mathbf{C}^K) \,. \tag{A3}$$

As explained in Sec. IV and Appendix C, the temporal basis functions $\boldsymbol{\Phi}$ and the spectral source signals $\mathbf{C}^k$ can be obtained by spectral ICA or PCA. (Note that Eq. (A1) does not correspond exactly to the implementation of the method, because the implementation uses the extended matrices, as mentioned in Sec. III. See Appendix C for exact details on the input and output to the function calls of spectral ICA and PCA.)

This model thus far describes the temporal structure, but not the spectral structure inherent in the spectral source signals $\mathbf{c}_i^k$. We can extend the model to describe the spectral structure in these signals.

We can first construct a new matrix $\mathbf{D}^i$ of size $(F \times K)$ where each column is the ith spectral source signal for a different sound, that is, $\mathbf{D}^i = (\mathbf{c}_i^1, \ldots, \mathbf{c}_i^K)$, for K sounds. In total there will be I $\mathbf{D}^i$ matrices, one for each temporal basis function $\boldsymbol{\phi}_i$. Let us rename the kth column of $\mathbf{D}^i$ to $\mathbf{d}_k^i$ (note that $\mathbf{d}_k^i = \mathbf{c}_i^k$).

We will assume we have a set of J basis functions $\boldsymbol{\Psi}^i$, represented as a matrix of size $(F \times J)$ where each column vector $\boldsymbol{\psi}_j^i$ is a spectral basis function. Also, we will assume we have a vector of coefficients $\mathbf{u}_j^i$ for each basis function, such that the jth vector scales basis function $\boldsymbol{\psi}_j^i$ across sounds. The set of these vectors, $\mathbf{U}^i$, is represented by a matrix of size $(J \times K)$, where the jth row is the vector $(\mathbf{u}_j^i)^T$. The linear combination of a basis

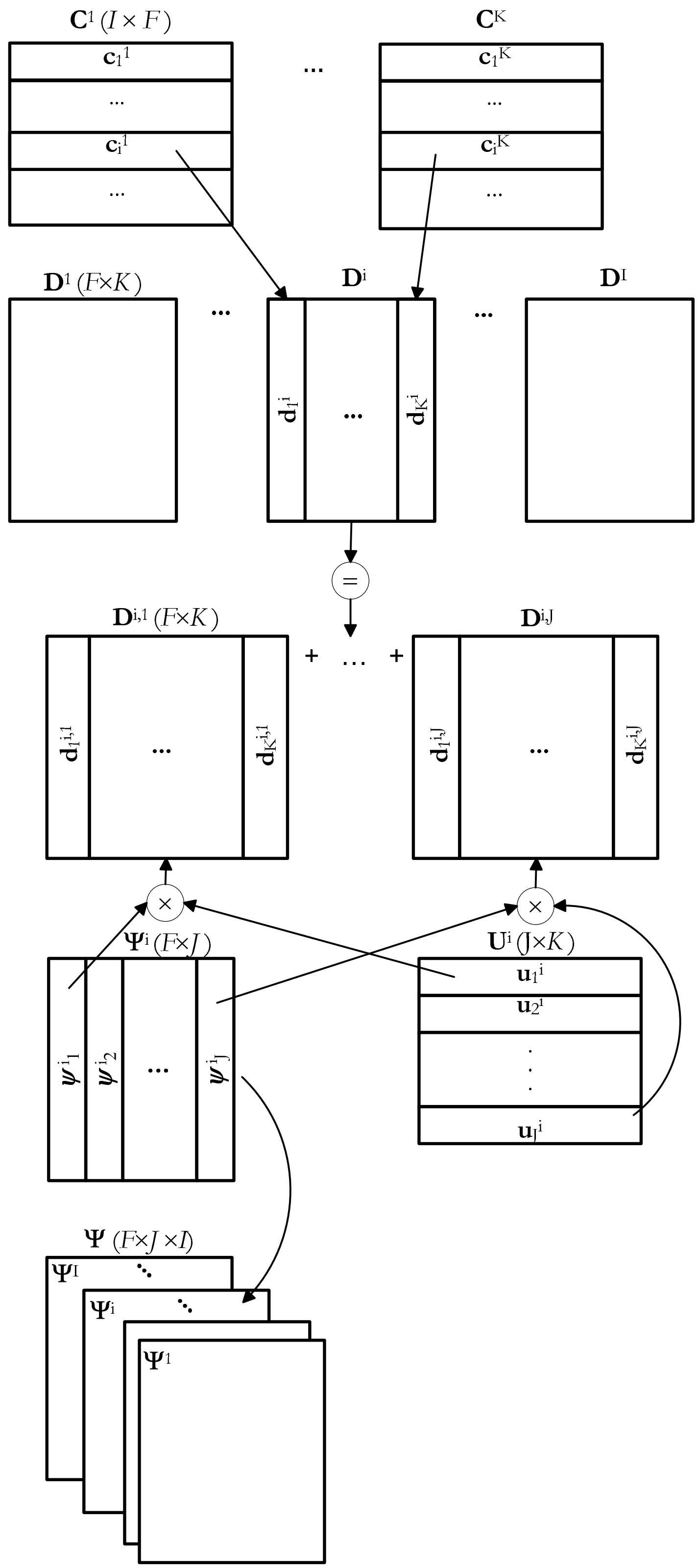


FIG. 15. Illustration of Eq. (A4). (For simplicity, here we did not mark row vectors with the transpose symbol T.)

function $\boldsymbol{\psi}_j^i$ with the corresponding vector of coefficients produces an $(F \times K)$ matrix $\mathbf{D}^{i,j}$ that represents part of the structure in $\mathbf{D}^i$. The whole structure in $\mathbf{D}^i$ is a linear combination of matrices $\mathbf{D}^{i,j}$ (see Fig. 15):

$$\mathbf{D}^i = \sum_{j=1}^{J} \mathbf{D}^{i,j} = \sum_{j=1}^{J} \boldsymbol{\psi}_j^i \, (\mathbf{u}_j^i)^T \; = \boldsymbol{\Psi}^i \, \mathbf{U}^i \, . \tag{A4}$$

As explained in Sec. IV and Appendix C, the spectral basis functions $\boldsymbol{\Psi}$ and matrices of coefficients $\mathbf{U}^i$ can be obtained by ICA or PCA.

When we consider only one column in $\mathbf{D}^i$, we can rewrite the previous equation as follow:

$$\mathbf{d}_k^i = \sum_{j=1}^{J} \mathbf{d}_k^{i,j} = \sum_{j=1}^{J} \boldsymbol{\psi}_j^i \, u_{j,k}^i \, , \tag{A5}$$

where $\mathbf{d}_k^{i,j}$ is the kth column of $\mathbf{D}^{i,j}$, and the scalar $u_{j,k}^i$ is the value of $\mathbf{u}_j^i$ for sound k.

We can now consider the Eq. (A5) at a given frequency bin f and express $d_{k,f}^i$ as follows:

$$d_{k,f}^i = \sum_{j=1}^{J} d_{k,f}^{i,j} = \sum_{j=1}^{J} \psi_{j,f}^i \, u_{j,k}^i \, , \tag{A6}$$

where the scalars $d_{k,f}^i$, $d_{k,f}^{i,j}$, and $\psi_{j,f}^i$ are, respectively, the values of $\mathbf{d}_k^i$, $\mathbf{d}_k^{i,j}$,and $\boldsymbol{\psi}_j^i$ at frequency bin f.

If instead of considering the rows of $\mathbf{U}^i$, we consider its columns, where $\mathbf{v}_i^k$ is the kth column of $\mathbf{U}^i$, we can define a new matrix $\mathbf{V}^k$ of size $(J \times I)$ with all vectors that refer to sound k from all $\mathbf{U}^i$ matrices: each column in $\mathbf{V}^k$ is the kth column of a different $\mathbf{U}^i$, that is $\mathbf{V}^k = (\mathbf{v}_1^k, \ldots, \mathbf{v}_I^k)$ (see Fig. 16). Considering $v_{i,j}^k$ as the jth value of $\mathbf{v}_i^k$, we have that $v_{i,j}^k = u_{j,k}^i$, and since $c_{i,f}^k = d_{k,f}^i$, we can rewrite Eq. (A6) as follows:

$$c_{i,f}^k = \sum_{j=1}^{J} \psi_{j,f}^i \, v_{i,j}^k \, . \tag{A7}$$

Finally, combining Eqs. (A2) and (A7) it follows that the bins of $\mathbf{S}^k$ can be expressed as

$$\mathbf{b}_f^k = \sum_{i=1}^{I} \sum_{j=1}^{J} \boldsymbol{\phi}_i \, \psi_{j,f}^i \, v_{i,j}^k \, , \tag{A8}$$

which shows how $\mathbf{S}^k$ can be modeled by a set of temporal basis functions $\boldsymbol{\Phi}$, a set of spectral basis functions $\boldsymbol{\Psi}$ (where $\boldsymbol{\Psi}$ is a three-dimensional matrix of size $(F \times J \times I)$ whose slices are the matrices $\boldsymbol{\Psi}^i$), and a set of coefficients $\mathbf{V}^k$, that is, $\mathbf{S}^k = M_b(\boldsymbol{\Phi}, \boldsymbol{\Psi}, \mathbf{V}^k)$.

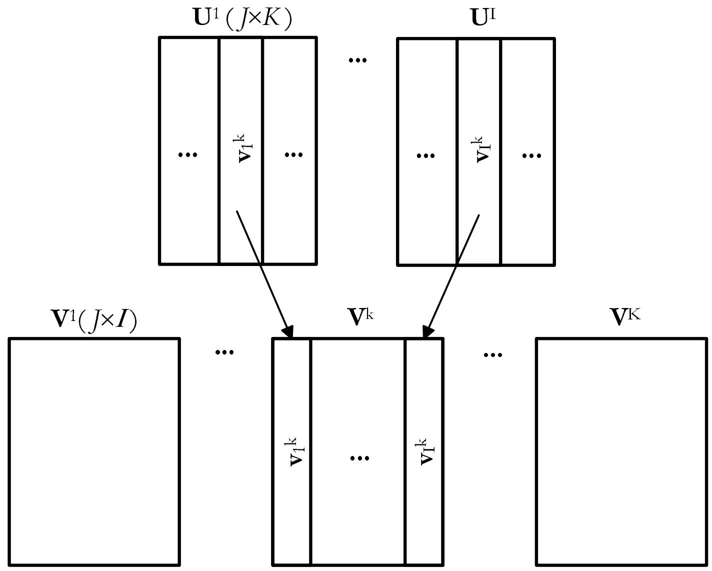


FIG. 16. Illustration of matrices $\mathbf{V}$.

APPENDIX B: TEMPORAL AND SPECTRAL ANALYSIS

ICA and PCA can be used to model the structure in the spectrogram frames or the structure in the spectrogram bins. In this section, we describe the differences between these two types of analysis.

Spectral analysis considers the frames (or power spectra) of $\mathbf{S}^k$ as spectral signal mixtures, i.e., the spectral signal in each frame is considered to be 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 instance, in order to learn the temporal basis functions $\mathbf{\Phi}$ and find the sets of spectral source signals $\mathbf{C}^1, \mathbf{C}^2, \ldots, \mathbf{C}^K$ (where $\mathbf{C}^k$ has one spectral source signal associated with each temporal basis function), 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 this data matrix has one row (or transposed frame) that corresponds to the start of all K impacts.

Temporal analysis considers the frequency bins of $\mathbf{S}^k$ as temporal signal mixtures, i.e., the time varying signal in each frequency bin is considered to be a linear combination of independent (for ICA) or uncorrelated (for PCA) time varying source signals. Like the temporal basis functions illustrated in Sec. IV.A.1, these source signals are vectors that range over the time space (see Sec. II for the definition of time space). Therefore, here we call them *temporal source signals*. The goal of temporal analysis is to decompose $\mathbf{S}^k$ into this set of temporal source signals, which here we call $\mathbf{P}^k$. $\mathbf{P}^k$ has one temporal source signal associated with each spectral basis function. For instance, a temporal analysis on matrix $(\mathbf{S}^1, \mathbf{S}^2, \ldots, \mathbf{S}^K)$ would result in learning a set of spectral basis functions (which here we call $\boldsymbol{\Theta}$) and finding the sets of temporal source signals $\mathbf{P}^1, \mathbf{P}^2, \ldots, \mathbf{P}^K$.

Considering the case of one sound, temporal PCA and spectral PCA give equivalent results, where the role of source signals and basis functions switch: when we consider a subset of I basis functions from the set of F spectral basis functions and a subset of I basis functions from the set of T temporal basis functions, learned by temporal or spectral PCA, respectively, the two types of analysis give equivalent results. Temporal PCA learns a set of spectral basis functions represented by an $(F \times I)$ matrix $\mathbf{A}^t$ (with $I \leqslant F$) and it finds the corresponding temporal source signals represented by an $(I \times T)$ matrix $\mathbf{Y}^t$, while spectral PCA learns a set of temporal basis functions represented by an $(T \times I)$ matrix $\mathbf{A}^s$ (with $I \leqslant T$) and it finds the corresponding spectral source signals represented by an $(I \times F)$ matrix $\mathbf{Y}^s$. The spectral basis functions learned by temporal PCA are the spectral source signals found by spectral PCA, that is, $\mathbf{A}^t = (\mathbf{Y}^s)^T$, and the temporal basis functions learned by spectral PCA are the temporal source signals found by temporal PCA, that is, $\mathbf{A}^s = (\mathbf{Y}^t)^T$.

Therefore, considering the case of one sound only and without extending the matrices (as discussed at the end of Sec. III), using temporal PCA to learn the spectral basis functions $\boldsymbol{\Theta}$ and find temporal source signals $\mathbf{P}^k$ is equivalent to using spectral PCA to learn the temporal basis functions $\boldsymbol{\Phi}$ and find spectral source signals $\mathbf{C}^k$. If the signal mixture matrices consist of the extended matrices $\mathbf{X}_1 = (-\mathbf{S}, \mathbf{S})$ for temporal analysis, and $\mathbf{X}_2 = (-\mathbf{S}^T, \mathbf{S}^T)$ for spectral analysis, in theory the results are different because since $\mathbf{X}_1$ and $\mathbf{X}_2$ have different sizes, $(F \times 2T)$ and $(T \times 2F)$, respectively, the results consist of matrices of different sizes.

However, if we ignore the results due to the negative parts in the extended matrices, that is, if we ignore that the results include $-\mathbf{P}^k$ and $-\mathbf{C}^k$, we can consider them equivalent because $\boldsymbol{\Theta} = (\mathbf{C}^k)^T$ and $\boldsymbol{\Phi} = (\mathbf{P}^k)^T$. (With more than one sound the results would not be equivalent because the signal mixture matrices would have different sizes, due to the spectrograms being concatenated in different ways for temporal and spectral analysis.)

The same does not happen with temporal and spectral ICA. ICA looks for correlations in the joint statistics of the data: spectral ICA looks for correlations (or structures) across the frames, while temporal ICA looks for correlations across the bins. Due to ICA's underlying assumptions of the statistical model and to the differences in the joint distribution of frames and bins (Fig. 17), the results obtained by temporal ICA and spectral ICA are not equivalent. Figure 18 shows that the source signals found by spectral and temporal ICA have different types of distributions.

The distribution of the frames (Fig. 17a) approximates the distribution assumed by ICA better than the distribution of the bins (Fig. 17b). Since spectral ICA looks for correlations across the frames, it matches the statistics of the data better than temporal analysis. Therefore, one can expect spectral ICA to lead to better results than temporal ICA. (This property applies to all sounds that have the same type of bin and frame joint distribution as in Fig. 17.) In fact, the temporal basis functions $\boldsymbol{\Phi}$ and spectral source signals $\mathbf{C}^k$ obtained by spectral ICA are smother and more easily interpretable than the results obtained by temporal ICA, which look rougher or noisier (results not shown).

The reason for this is that there is structure that is not represented by the spectral basis functions learned by temporal ICA. While most plots obtained by spectral ICA (Fig. 18a) have clear directions that define the data, the same does not happen in the plots obtained by temporal ICA (Fig. 18b). The lack of ability of the spectral basis functions (from temporal ICA) to explain the whole structure in the spectrogram results from the difference between ICA's underlying assumptions of the statistical model and the joint distribution of the bins (Fig. 17).

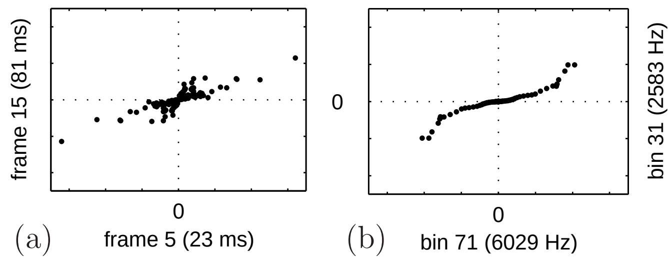
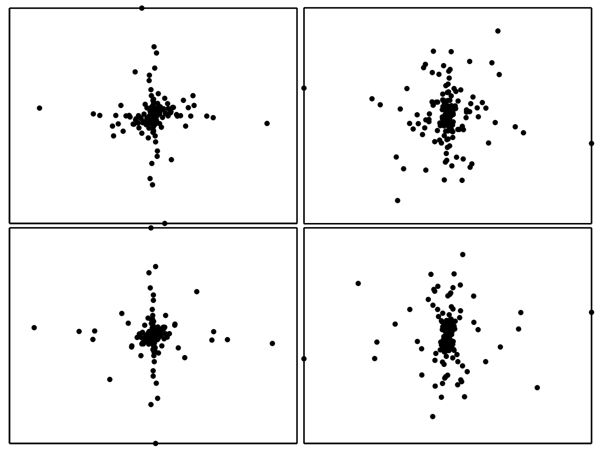
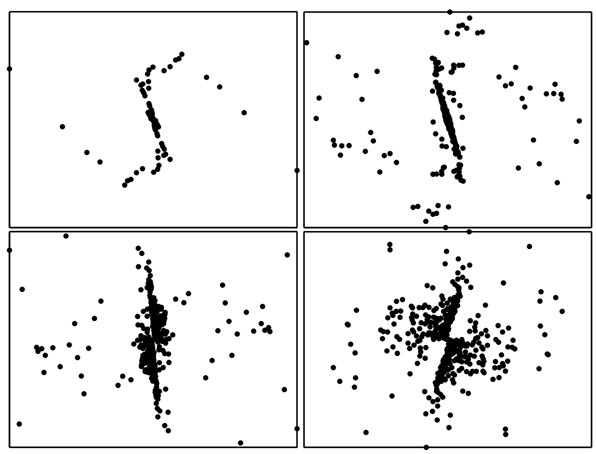


FIG. 17. Scatter plots of frames and bins from a sound from an aluminum rod (Al1). (a) Distribution of two frames: This scatter plot shows the values of frame 5 (23 ms) plotted against the values of frame 15 (81 ms), where each point in the plot corresponds to a different frequency bin. These frames have high energy in some frequency bins, which correspond to points far from zero, and low energy in other frequency bins, which correspond to points close to zero. Since frame 5 occurs earlier in the impact sound, it has more energy than frame 15. (b) Distribution of two bins: this scatter plot shows the values of bin 71 (6 kHz) plotted against the values of bin 31 (2.6 kHz), where each point in the plot corresponds to a different time frame. These bins have high energy in the first time slots (see the points in the upper right, or lower left) and then the energy decreases until they reach zero. Frames and bins with high energy were chosen in both plots to avoid having all the points falling into the neighborhood of zero. The points in the third quadrant of both plots are due to the negative part of the extended matrices (see the end of Sec. III for details).



(a)



(b)

FIG. 18. Scatter plots of source signals. Each plot shows one source signal plotted against another source signal, where both source signals are associated with the same basis function. The source signals have been found by either spectral or temporal analysis of a spectrogram from a sound from an aluminum rod (Al1). (a) Results obtained by spectral ICA. (b) Results obtained by temporal ICA.

APPENDIX C: PCA AND ICA FUNCTION CALLS

This appendix gives details of the implementation of the model presented in Sec. II. As explained in Sec. III, we use PCA and ICA to learn the sets of basis functions $\mathbf{\Phi}$ and $\mathbf{\Psi}$. In order to learn $\mathbf{\Phi}$, the implementation of the first part of model M_b uses spectral analysis, namely spectral PCA and spectral ICA of spectrograms (see Sec. IV or Appendix B for the definition of spectral analysis). Afterwards, it learns $\mathbf{\Psi}$ by applying PCA and ICA to the spectral source signals $\mathbf{C}^k$ associated with $\mathbf{\Phi}$.

The implementation of model M_b with spectral PCA applies MATLAB's built in `princomp` function to $(-\mathbf{X}, \mathbf{X})^T$, where $\mathbf{X}$ is the horizontal concatenation of transposed spectrograms $((\mathbf{S}^1)^T, (\mathbf{S}^2)^T, \ldots, (\mathbf{S}^K)^T)$, and $(-\mathbf{X}, \mathbf{X})^T$ is the vertical concatenation of $-\mathbf{X}^T$ and $\mathbf{X}^T$. This produces matrices $\mathbf{\Phi}$ and $\mathbf{C}^T$, where $\mathbf{\Phi}(:, i) = \boldsymbol{\phi}_i$, and $\mathbf{C}$ is a matrix that contains $\mathbf{C}^1, \mathbf{C}^2, \ldots \mathbf{C}^K$ (where $\mathbf{C}^k$ contains vectors $\mathbf{c}_1^k, \ldots, \mathbf{c}_I^k$ as well as the results relative to $-\mathbf{X}$), or more specifically, $\mathbf{C}(i, :)$ is a vector that contains $\mathbf{c}_i^k$ (and $-\mathbf{c}_i^k$) for $k \in 1, ..., K$.

The implementation of the model with spectral ICA applies `fastica` (Hyvärinen et al., 2001) to matrix $(-\mathbf{X}, \mathbf{X})$. The results shown here were obtained with option `g`, which specifies the nonlinearity used in the fixed-point algorithm, set to `tanh` and option `lastEig`, which specifies the number of eigenvectors used in the computation, set to `50`. This produces matrices $\mathbf{\Phi}$ and $\mathbf{C}$ as defined earlier. Note that here a signal mixture is the horizontal concatenation of one transposed frame from each of the K spectrograms, and there are T signal mixtures. Therefore $I \leqslant T$ in Eqs. (2)$-$(7). (See Sec. III for the definition of signal mixture and for details on the size of the matrices.)

Once it finds the matrices of spectral signals $\mathbf{C}^1, \mathbf{C}^2, \ldots \mathbf{C}^K$, the implementation of model M_b rearranges them into matrices $\mathbf{D}^i$, where $\mathbf{D}^i = (\mathbf{c}_i^1, ..., \mathbf{c}_i^K)$ (see the top of Fig. 15). For each matrix $\mathbf{D}^i$, it applies function `princomp` to $(\mathbf{D}^i)^T$ or `fastica` (with the same options as for spectral ICA) to $\mathbf{D}^i$. These calls produce matrices $\mathbf{\Psi}^i$, and $\mathbf{U}^i$. Finally, it obtains matrices $\mathbf{V}^k$ by rearranging matrices $\mathbf{U}^i$ (see Fig. 16). Here a signal mixture is a row of $\mathbf{D}^i$. Since $\mathbf{D}^i$ is a matrix of size $(F \times K)$ there are F signal mixtures. Therefore, $J \leqslant F$ in Eqs. (5)$-$(7).