Audio Engineering Society
Convention Paper
Presented at the 114th Convention
2003 March 22–25 Amsterdam, The Netherlands

Prior Subspace Analysis for Drum Transcription

Derry FitzGerald[1], Bob Lawlor[2], and Eugene Coyle[3]

[1]Music Technology Center, Dublin Institute of Technology, Rathmines Rd. Dublin, Ireland

[2]Department of Electronic Engineering, National University of Ireland, Maynooth, Ireland

[3] Department of Electronic Engineering, Dublin Institute of Technology, Kevin St, Dublin Ireland

ABSTRACT

This paper introduces the technique of Prior Subspace Analysis (PSA) as an alternative to Independent Subspace Analysis (ISA) in cases where prior knowledge about the sources to be separated is available. The use of prior knowledge overcomes some of the problems associated with ISA, in particular the problem of estimating the amount of information required for separation. This results in improved robustness for drum transcription purposes. Prior knowledge is incorporated by use of a set of prior frequency subspaces that characterise features of the sources to be extracted. The effectiveness and robustness of PSA is demonstrated by its use in a simple drum transcription algorithm.

1. INDEPENDENT SUBSPACE ANALYSIS

Independent Subspace Analysis (ISA) provides a means of attempting sound source separation from single channel mixtures [1]. Based on redundancy reduction techniques, it represents sound sources as low dimensional independent subspaces in the time-frequency plane. To carry out ISA the single channel mixture signal is converted to a time-frequency representation such as a spectrogram. It is then assumed that the overall spectrogram $\mathbf{Y}$ results from the superposition of a number of unknown statistically independent spectrograms Y_j, yielding:

$$\mathbf{Y} = \sum_{j=1}^{l} Y_j \qquad (1)$$

Further it is assumed that each independent spectrogram can be represented as an outer product of an invariant frequency basis function and a corresponding invariant amplitude basis function, giving:

$$Y_j = f_j t_j^T \qquad (2)$$

The frequency basis function describes the relative strengths of the frequencies present in the independent spectrogram, and the amplitude basis

function describes the variations in amplitude of the frequency basis function over time. This is illustrated in Figure 1, which shows the frequency basis function and amplitude basis function of a snare drum. When multiplied together they produce a spectrogram which is a reasonable approximation to that of the original snare drum spectrogram.

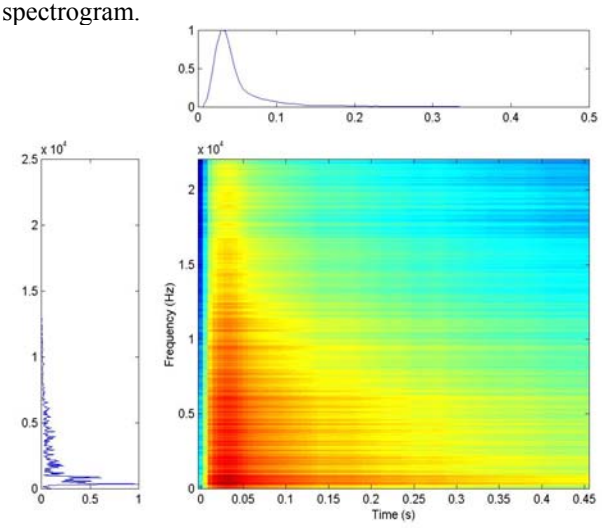

Figure 1. Basis functions and Spectrogram of a snare drum

Summing the $\mathbf{Y}_j$ yields:

$$\mathbf{Y} = \sum_{j=1}^{l} f_j t_j^T \qquad (3)$$

The basis functions represent features of the individual sound sources and each source is composed of a number of these basis functions. These basis functions form a low dimensional subspace that represents the individual sounds in the time-frequency plane.

Principal Component Analysis (PCA) provides a means of decomposing a spectrogram into a set of outer product basis functions and also provides a means of redundancy reduction. PCA takes a set of correlated variables and linearly transforms them into a number of uncorrelated variables that are termed principal components. These are ordered by the amount of variance of the original data they contain. As the principal components are ordered by decreasing variance PCA is used to reduce redundancy by discarding components of low variance.

However the principal components obtained from PCA are not statistically independent and so a technique known as Independent Component Analysis (ICA) is used to obtain a set of independent basis functions from the principal components retained from the PCA step. ICA attempts to separate a set of observed signals that are composed of linear mixtures of a number of independent non-gaussian sources into a set of signals that contain the independent sources [2]. It should be noted that the separation obtained by ICA is not perfect and so in some cases there will still be artefacts related to other sound sources in the independent basis functions. However these will be much reduced in comparison to the artefacts present before separation using ICA.

Once obtained the independent basis functions can be used to generate the independent spectrograms. Phase information for resynthesis can be obtained by using the original phase information from the Short Time Fourier Transform used to generate the spectrogram or via a phase estimation method such as that described by Griffin and Lim [3].

However there are a number of problems associated with ISA. Firstly because the basis functions are invariant no pitch changes are allowed in the sound source spectrograms. This presents a problem when dealing with most musical instruments. However with drum sounds where the pitch does not change from one occurrence of a given drum to another this is a valid approximation. This makes ISA-type approaches well suited to drum transcription.

Secondly, due to the fact that all ICA algorithms are indeterminate with regards to ordering of the input components, it is necessary to identify a given source by some means such as their frequency characteristics or amplitude envelopes after ISA has been completed.

Thirdly the quality of separation also depends on the length of the signal input. For instance a signal containing just one hi-hat and snare played simultaneously will not separate correctly. For the hi-hat/snare separation 2-4 events are typically required, depending on the frequency and amplitude characteristics of the drums used.

Finally estimating the number of components to retain from the PCA stage represents a considerable difficulty. The number of components required for correct separation varies with the frequency and amplitude characteristics of the source sounds. There is also a trade-off between the number of components retained and the recognisability of the resulting basis functions. Keeping a large number of components results in basis functions that support small regions of the frequency spectrum. Using a small number of

components results in basis functions that contain recognisable features of the source sounds with support across the entire frequency spectrum.

As a result of this trade-off ISA works best on signals with less than five sources. This trade-off also means that it is necessary to choose carefully the number of components retained to achieve optimal source separation. Thresholding methods have not proved effective in obtaining the correct number of components, and as a result an observer is necessary to determine the required number of components. Methods such as sub-band ISA have been proposed in an effort to overcome this indeterminacy for the purposes of drum transcription [4].

2. PRIOR SUBSPACE ANALYSIS

As noted previously there are a number of problems inherent in the ISA method, in particular estimating the correct number of components to retain from the PCA step. While methods such as sub-band ISA go some way to overcoming this problems a more efficient method lies in the utilisation of prior knowledge about the sources to be separated.

ISA arose out of attempts to create a signal representation that could characterise and allow further manipulation of individual everyday sounds such as a coin hitting the floor [5]. The method looked for invariants that characterised sounds and involved performing PCA followed by ICA on a spectrogram of a sound in a manner similar to that of ISA. The technique was later used for generalised sound classification and incorporated into the MPEG7 specification [6]. Applying the same technique to a mixture of sounds resulted in ISA.

The success of this method in generalised sound classification suggests that it can be adapted to create a set of prior subspaces that can characterise a given sound source such as a snare drum. These prior subspaces can then be used to carry out an initial analysis of a mixture signal. This Prior Subspace Analysis (PSA) has the same underlying assumptions as ISA, namely that the overall mixture spectrogram results from the sum of a number of independent spectrogram, and that these independent spectrograms can be represented as the outer product of a frequency basis function and an amplitude basis function.

PSA then assumes that there exists known prior frequency subspaces or basis functions f_p that are good initial approximations to the actual subspaces. Substituting the f_j in equation 3 with these prior subspaces yields:

$$\mathbf{Y} \approx \sum_{j=1}^{l} f_p t_j^T \qquad (4)$$

Therefore multiplying the overall spectrogram by the pseudoinverse of a prior frequency subspace yields an estimate of the amplitude basis function $\hat{t}_j$. However the estimated amplitude basis functions returned are not independent, and so ICA is carried out on these amplitude basis functions to yield independent basis functions $\hat{t}_{ij}$. These independent basis functions can in turn be used to get improved estimates of the frequency basis functions $\hat{f}_{ij}$. The independent spectrograms can then be estimated from

$$\hat{Y}_j = \hat{f}_{ij}\hat{t}_{ij}^T \qquad (5)$$

Resynthesis can then be carried out in a similar manner to that of ISA.

Prior Subspaces for use with PSA are obtained by analysing large numbers of each of the sound sources of interest. An ISA-type approach is used to generate frequency prior subspaces for each sample of a particular sound source. PCA is carried out on the spectrogram of each sound source sample. The first three principal components are retained and these are then analysed using ICA to yield independent frequency subspaces. The independent subspace with the largest projected variance is then chosen to be the prior frequency subspace for the sound source sample in question. K-means clustering is then carried out on the prior subspaces for a given sound source to yield a single prior frequency subspace which characterises that sound source.

Figure 2 shows the prior subspaces obtained for snare, kick drum and hi-hat respectively. The general frequency characteristics of each drum type have been captured well. Kick drums have the vast majority of their energy in the lowest part of the spectrum, with very little energy outside this region. Snares contains most of their energy in the lower part of the spectrum, but the main resonance occurs at frequencies higher than that of kick drums. Snares also have some frequency energy spread across a large portion of the spectrum. Finally hi-hats have energy across the entire range of the spectrum. As can be seen these characteristics have been captured well in the prior subspaces obtained for these sources.

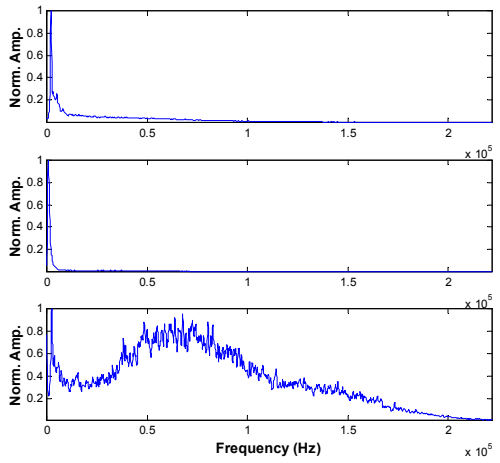

Figure 2. Prior Subspaces for snare, kick drum and hi-hat.

Figure 3 shows the independent amplitude basis functions obtained by carrying out PSA on a drum loop. As can be seen there is good separation of the events, with kick drums only showing up as tiny peaks in the snare basis function and vice-versa.

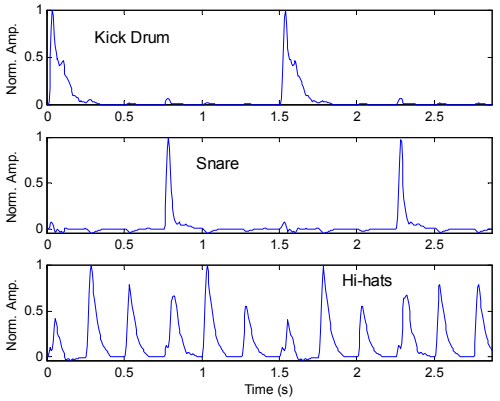

Figure 3. Drum loop separation using PSA

PSA needs prior knowledge of the sources present and a prior frequency subspace associated with each source. Significantly by using prior knowledge it only looks for as many sources as are known to be present and can find them efficiently. PSA is also faster than ISA or sub-band ISA due to the fact that it does not need to use PCA to obtain data reduction.

3. DRUM TRANSCRIPTION USING PSA

To demonstrate the utility of the PSA method a simple drum transcription algorithm was implemented. To allow direct comparison with sub-band ISA the same drum loops used in testing sub-band ISA were used in testing PSA. The 15 drum loops used contained hi-hats, snares and kick drums. These drums were chosen as they are the most commonly occurring drums in popular music. The drum patterns used were examples of commonly found patterns in rock and pop music as well as variations on these patterns. The drums were taken from various sample CDs and were chosen to cover the wide variations in sound within each type of drum. Tempos ranging from 80 bpm to 150 bpm. were used and different meters such as 4/4, 3/4 and 12/8 were used. Relative amplitudes between the drums were varied between 0 dBs to –24 dBs to cover a range of situations so as to make the tests as realistic as possible.

In order to overcome the source signal ordering problem inherent in ICA a number of assumptions were made to allow identification of the sources. Firstly it is assumed that hi-hats occur more frequently than the other drums present. This assumption holds for most drum patterns containing hi-hats in popular music. Secondly it is assumed that the kick drum has a lower spectral centroid than the snare drum. As snare drums are perceptually brighter than kick drums, and the brightness of sounds has been found to correlate well with the spectral centroid, this is a reasonable assumption [8].

As a result of imperfect separation from the ICA stage the recovered independent amplitude basis functions are normalised and all peaks over a set threshold are taken as an occurrence of a given drum. The same threshold was used for all the test signals in both PSA and sub-band ISA to allow for direct comparison of results. Onset times were calculated using a variation on the onset detection algorithm developed by Klapuri [7].

4. DRUM TRANSCRIPTION RESULTS

The results obtained for transcription using PSA are summarised in Table 1 below. Table 2 shows the results obtained using sub-band ISA to allow comparison between the two methods. The percentage of correctness was obtained from the following formula:

$$correct = \frac{total - missing - incorrect}{total} * 100$$

Type	Total	Missing	Incorrect	%
Snare	21	0	2	90.5
Kick	33	0	0	100
Hats	79	2	6	89.9
Overall	133	2	8	92.5

Table 1: Drum Transcription Results using PSA

Type	Total	Missing	Incorrect	%
Snare	21	0	2	90.5
Kick	33	0	0	100
Hats	79	6	6	84.8
Overall	133	6	8	89.5

Table 2: Drum Transcription Results using sub-band ISA

As can be seen from the tables the results for snares and kicks are identical. However it should be noted that the extra snares detected using PSA were as a result of amplitude modulation rather than identifying kick drums as snares, as was the case with sub-band ISA. A change to the PSA transcription algorithm to take amplitude modulations into account would possibly eliminate these errors.

PSA correctly detected more of the hi-hats than sub-band ISA. The fact that PSA correctly identified a greater number of hats suggests that using prior subspaces provides a better means to detect hi-hats than the blind separation methods of sub-band ISA. In both methods the undetected hats were separated correctly but fell below the threshold for detection. A number of snares were also identified as hi-hats in both PSA and sub-band ISA. This is due to the high frequency energy present in snare drums which can make the separation between snares and hats difficult. It should be noted that there is a trade-off in setting the threshold level between detecting low amplitude occurrences of a drum and between incorrectly detecting drums due to imperfect separation. In the case of hi-hat/snare separation setting the threshold too low results in extra snares being detected as hi-hats, while too high a threshold results in increased numbers of undetected hi-hats. The threshold used was found to represent a good balance between the two.

The average error in detecting onsets was 10 ms for both PSA and sub-band ISA. This is due mainly to smearing of the onsets as a result of the overlapping windows used in calculating the spectrogram and also due to the limitations of time resolution in the STFT used to calculate the spectrogram.

Drum transcription using PSA is considerably faster than using sub-band ISA. When both algorithms are implemented in Matlab PSA is approximately ten times faster than sub-band ISA. This is as a result of the elimination of the PCA step which results in an increase in the speed of the algorithm. It should also be noted that sub-band ISA needs two passes through the data, resulting in ISA being performed twice, compared to the single pass required for PSA, making PSA more efficient than sub-band ISA.

5. CONCLUSIONS

This paper has introduced the technique of Prior Subspace Analysis as a tool for drum transcription and sound source separation. It has proved itself to be a viable method for the transcription of drums, overcoming some of the problems associated with ISA. It has also proved to be more effective and efficient in transcribing drums than sub-band ISA.

Future work includes the extension of PSA to transcribe drums in the presence of pitched instruments, and to extend PSA to handle increased numbers of drums.

6. REFERENCES

[1] Casey, M.A. & Westner, A., “Separation of Mixed Audio Sources By Independent Subspace Analysis” in Proc. Of ICMC 2000, pp. 154-161, Berlin, Germany.

[2] A. Hyvärinen and E. Oja. “Independent Component Analysis: Algorithms and Applications”. *Neural Networks*, 13(4-5): pp 411-430, 2000.

[3] Griffin, D., & Lim, J. S. “Signal Estimation from Modified Short-Time Fourier Transform”, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-32, pp. 236-243, 1984.

[4] FitzGerald, D., Coyle E, Lawlor B. “Sub-band Independent Subspace Analysis for Drum Transcription”, Proceedings of the. Digital Audio Effects Conference (DAFX02), Hamburg, pp. 65-69, 2002.

[5] Casey, M., “Auditory Group Theory: with Applications to Statistical Basis Methods for Structured Audio”, Ph.D. Thesis, MIT Media Lab, February 1998.

[6] Casey, M., “Generalized Sound Classification and Similarity in MPEG-7”, Organized Sound, 6:2, 2002

[7] Klapuri, A., “Sound Onset Detection by Applying Psychoacoustic Knowledge”. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 1999.

[8] Gordon, J., and Grey, J. M., "Perceptual Effects of Spectral Modifications on Orchestral Instrument Tones." Computer Music Journal, Vol. 2, N° 1, pp. 24-31, 1978