DRUM TRANSCRIPTION USING AUTOMATIC GROUPING OF EVENTS AND PRIOR SUBSPACE ANALYSIS

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

[1]Music Technology Center, Dublin Institute Of Technology, Rathmines Rd. Dublin, Ireland
[2]Dept. Of Electronic Engineering, National University of Ireland, Maynooth, Ireland

ABSTRACT

While Prior Subspace Analysis (PSA) has proved an effective tool for transcribing mixtures of snare, kick drum and hi-hat both in the "drums-only" case and in the presence of pitched instruments attempts to extend it to deal with increased numbers of drum types have met with mixed results. To overcome this an automatic modeling and grouping procedure has been developed which groups drum events on the similarity of their frequency content. Combining this procedure with PSA allows the extension of PSA to robustly handle greater numbers of drum types. The effectiveness of this approach is demonstrated in a drum transcription algorithm.

1. PRIOR SUBSPACE ANALYSIS

Prior Subspace Analysis (PSA) is a technique for sound source separation in single channel mixtures in cases where prior knowledge is available about the sources [1,2]. PSA represents sound sources as low dimensional independent subspaces in the time-frequency plane and is based on Independent Subspace Analysis (ISA) and the generalised sound classification techniques created by Casey [3,4]. It uses prior knowledge about the sources to overcome a number of problems associated with ISA not least of which is the problem of estimating the amount of information to be retained from the dimensional reduction stage of ISA.

The mixture signal is transferred to a time-frequency representation such as a spectrogram. PSA then assumes that the overall spectrogram ***Y*** results from the summation of l unknown independent spectrograms Y_j. This yields

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

These independent spectrograms Y_j are assumed to be represented by the outer product of an invariant frequency basis function f_j, and a corresponding invariant amplitude basis function t_j which describes the variations in amplitude of the frequency basis function over time. This gives

$$\mathbf{Y}_j = f_j t_j^T \qquad (2)$$

Summing over 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 which form a low dimensional subspace that represents the individual sounds in the time-frequency plane. The outer product assumption means that in practice no pitch change is allowed in the sound sources over the course of the spectrogram. This presents a problem when dealing with most musical instruments. However with drums sounds where the pitch does not change from one occurrence of a given drum to another this is a valid approximation.

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

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

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 amplitude basis functions returned are not independent. To make the basis functions independent, independent component analysis (ICA) is performed on the amplitude basis functions, yielding $\hat{t}_{ij}$. ICA attempts to separate linear mixtures of signals into the original source signals by making the signals as statistically independent as possible

[5]. These independent amplitude basis functions can then be used to obtain better estimates of the actual frequency subspaces, $\hat{f}_{ij}$. The independent spectrograms can then be estimated from

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

Phase information for resynthesis can be obtained via a method such as that described by Griffin and Lim [6].

Prior Subspace Analysis has proved an effective method for transcribing mixtures of snare, kick drum and hi-hat (or ride cymbal) both in the case of "drums-only" and in the presence of pitched instruments. The prior subspaces were obtained by analysing large numbers of each drum type. The drums were analysed using an ISA-type approach. First Principal Component Analysis was carried out on the spectrogram of the drum sample. The first three principal components were retained and these were then analysed using ICA. The independent component with the largest projected variance was then chosen to as the prior frequency subspace for the drum sample in question. K-means clustering was then carried out on the prior frequency subspaces for a given drum type. This yielded a single prior frequency subspace that characterised a given drum type.

2. LIMITATIONS OF PSA

Though successful in dealing with mixtures of three drums types attempts to extend PSA to deal with mixtures containing more drum types have met with mixed success. In particular the addition of toms to the mixture causes some difficulty. In some cases the mixtures of snare, kick drum, toms and hi-hat are separated correctly, but in other cases the analysis fails to handle the toms. This is partially due to the fact that there is a greater range of frequency overlap between snare drums and toms than there is between snare and kick drum. This can result in very similar initial estimates of the amplitude basis functions for both snares and toms. The similarity of these basis functions then causes the ICA algorithm to arrive at the wrong solution. Another contributing factor is due to the fact that there is a wider range of tunings for tom drums than for either snare or kick drum, making it harder for a single subspace to characterise the entire family of tom drums. However splitting the toms into smaller subgroups with similar tunings and

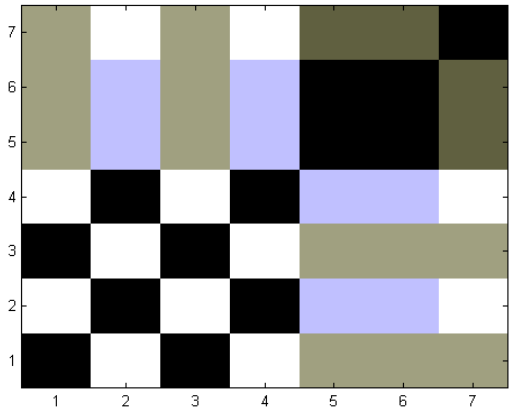

Figure 1. Similarity of events in a drum loop

then analysing each subgroup to obtain a prior frequency subspace for each subspace gave no noticeable improvement in performance. Similar problems occurred when trying to distinguish between hi-hats and ride cymbals.

3. AUTOMATIC GROUPING OF EVENTS

As can be seen from the above PSA cannot robustly deal with signals containing more than three drum types. Therefore an alternative approach is required to deal with mixtures of more than three drum types.

As the ISA-type analysis has proved successful in generating prior subspaces it is proposed to use this type of approach to automatically model the events that occur in a drum loop or drumming performance. To model each event individually it is necessary to identify when an event occurs. To this end a spectrogram of the input signal is multiplied by prior frequency subspaces for both snare and kick drum. The resulting amplitude basis functions are then normalised and all peaks above a set threshold are taken to be a drum event. This is sufficient to identify all skinned drum onsets including toms. Onset time of each event is then determined, and the sections of the spectrogram between each event analysed individually.

Principal Component Analysis is performed on each section and the first frequency principal component from each section retained. These are then normalised and the Euclidean distance is calculated between all pairs of principal components. For p events this results in a p x p symmetric matrix containing the distances between the events. The diagonal elements of the matrix are zero.

Figure 1 shows the similarity matrix obtained from analysing a drum loop containing snare, kick drum and two different types of tom-tom. Black indicates that the events are highly similar and white indicates regions of large dissimilarity. As can be seen events 1 & 3 are highly similar. These events correspond to occurrences of a kick drum. Events 2 & 4 correspond to occurrences of a snare drum Events 6 & 7 correspond to two occurrences of one of the toms, and event 5 is the other type of tom that occurs. Event 5 is closer to the other type of tom drum than to the snare and kick drums. It can be seen that the similarity matrix shows the correct grouping of the events.

To group the events the following procedure was used. Starting from the first event, all events with a Euclidean distance of less than one from the first event are grouped together and removed from the list of events remaining ungrouped. It is assumed that each event can belong to only one group. The next ungrouped event is then chosen and the procedure is repeated until all events have membership of a group. In cases where each event represented only a single drum this amounted to the correct transcription of the drum loop. However this is not usually the case. Typically a hi-hat or ride cymbal will occur with a skinned drum such as snare, kick or tom. In some cases the skinned drums will also occur simultaneously.

4. DRUM TRANSCRIPTION USING AUTOMATIC GROUPING

A drum transcription algorithm using automatic grouping was implemented in Matlab. The system assumes that at least snares, kick drums and hi-hats or ride cymbals are present. The initial stage of the analysis proceeds as described above, with the skinned drum events being grouped according to their similarity to other events. To overcome the most commonly occurring skinned drum overlap, that of snare and kick drum, the groups most likely to correspond to snare drum and kick are identified. The snare group is identified as the group that contains the largest peak found in the initial snare amplitude envelope obtained from multiplying the spectrogram with the snare prior subspace. The kick drum group is then identified as the group with the lowest spectral centroid. Any remaining groups are then identified as toms. Prior frequency subspaces are obtained for each of the groups, and all non-snare and kick events in the spectrogram are masked. PSA is then performed on the resulting spectrogram and the snare and kick drum events identified. The algorithm is still prone to errors from the overlap of toms with other skinned drums, but this is not a very common occurrence.

Power Spectral Density normalisation is then performed on the original spectrogram to eliminate the effects of the skinned drums as much as possible. The PSD normalised spectrogram is multiplied by a prior hi-hat subspace. This is sufficient to recover all metallic drum events, such as hi-hats and cymbals. However both snare and tom drum events will also appear in the resulting amplitude envelope which could be detected as a metallic drum where none is present. To overcome this overlap kick drum events are masked in the original spectrogram, and the resulting spectrogram is multiplied by a snare frequency subspace. ICA is then performed on the resulting amplitude envelope and that of the hi-hat subspace. All events above a threshold in the resulting hi-hat envelope are then taken as metallic drum events.

Automatic grouping is then carried out on the metallic drum events. However due to interference from other drums no simple threshold suffices for grouping the drums. To overcome this and set an approximate threshold for the drums a histogram of the distances is obtained. The lower edge of the first histogram bin with no entry is taken as the threshold. Events are then grouped as before using this threshold. If two large groups occur that do not overlap in time then both hi-hat and ride cymbal are taken to occur within the loop, and these groups are kept separated. If not then all events are grouped together. The justification for this is that most drummers tend to stay on either hi-hat or ride cymbal for long periods, usually only changing when the piece or song changes from one section to another, such as from verse to chorus. It is rare to hear a drummer alternating between hi-hat and ride events in the course of a bar of music. As a result if overlapping groups occur it is most likely to be the same metallic drum that has been grouped into a number of groups due to interference from skinned drums. However as a result of this grouping strategy the algorithm is unable to detect the presence of either crash cymbals or open hi-hats.

At present the transcription algorithm has no means of distinguishing between hi-hats and ride

cymbals, and so the groups are labeled metallic drums 1 and 2.

5. RESULTS

The drum transcription algorithm was tested on 25 drum loops, with the number of different drums (including different types of tom) in the loops ranging from three to seven drums. The drums were obtained from sample CDs and were chosen to cover as wide a spread of drum sounds within a given drum type as possible. A wide variety of different drum patterns and drum fills were used. The tempos used ranged from 150bpm to 80 bpm and different meters were used, including 4/4, 3/4 and 12/8. The relative amplitudes between the drums varied between 0 dBs to –24 dBs to make the tests as realistic as possible. The same analysis parameters were used on all the test signals. The results are summarised in Table 1.

Type	Total	Missing	Incorrect	%
Snare	40	0	0	100
Kick	64	3	1	93.8
Toms	31	3	4	77.4
Metallic	165	9	12	87.3
Overall	300	15	16	89.3

Table 1. Transcription Results

As can be seen all the snare drums were correctly identified. The three missing kick drums and the extra kick drum all come from the same drum loop. The three missing kick drums were in fact correctly grouped together. However in the loop in question an unusually low tuned tom was mistakenly identified as the kick drum, leading to the kick drums being identified as kick drums. Three of the extra toms come from this misidentification also. The remaining extra tom came from an unusually loud hi-hat being detected as a skinned drum. The three missing toms fell below the threshold for detection as a skinned drum. The missing nine metallic drums also all fell below the threshold for detection. The twelve extra metallic drums were as a result of incorrect separation of the metallic and snare/tom subspaces. In cases where both hi-hat and ride cymbal were present in the same loop the drums were grouped correctly together.

The automatic grouping performed remarkably well on the skinned drums. All events passed to the grouping stage were in fact correctly grouped, with any errors in the transcription process occurring elsewhere in the algorithm. This demonstrates the effectiveness of the grouping methodology as a tool for drum transcription.

It should also be noted that these results were achieved without any form of rhythmic modeling or incorporating models of common drum patterns.

7. CONCLUSIONS AND FUTURE WORK

Automatic grouping in conjunction with PSA has been shown to be an effective tool for drum transcription, extending the range of circumstances in which robust "drums-only" transcription is possible. However there are a number of limitations on the system. Future work will concentrate on removing these limitations to allow the algorithm to work in even more generalised situations. In particular it is proposed to extend the system to identify groups as hi-hats or ride cymbals and to allow the algorithm to deal with crash cymbals and open hi-hats. It is proposed to do this by incorporating a drum classification system into the algorithm. It is also proposed to remove the assumption that at least snare, kick drum and hi-hat or ride cymbal are present in the drum loop.

6. REFERENCES

[1] FitzGerald, D., Coyle E, Lawlor B. "Prior Subspace Analysis for Drum Transcription", submitted to 114th AES Conference

[2] FitzGerald, D., Lawlor B., Coyle E., "Drum Transcription in the presence of pitched instruments using Prior Subspace Analysis" submitted to ICASSP 2003

[3] Casey, M. & Westner, A., "Separation of Mixed Audio Sources By Independent Subspace Analysis" ,*Proceedings Of ICMC 2000*, pp. 154-161, Berlin, Germany, 2000.

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

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

[6] 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.