2003-01-01

Drum Transcription using Automatic Grouping of Events and Prior Subspace Analysis

Derry Fitzgerald
Technological University Dublin, derry.fitzgerald@tudublin.ie

Robert Lawlor
Technological University Dublin

Eugene Coyle
Technological University Dublin

Recommended Citation

Fitzgerald, D., Lawlor, R. & Coyle, E. (2003) Drum transcription using automatic grouping of events and prior subspace analysis. Proceedings of the *Digital Audio Effects Conference* (DAFX03) pp.13-17, London, England, 2003

DRUM TRANSCRIPTION USING AUTOMATIC GROUPING OF EVENTS AND PRIOR SUBSPACE ANALYSIS

D. FITZGERALD, R LAWLOR AND E. COYLE
Dublin Institute of Technology.
Rathmines Road, Dublin, Ireland
E-mail: derry.fitzgerald@dit.ie, rlawlor@eeng.may.ie, eugene.coyle@dit.ie

While Prior Subspace Analysis (PSA) has proved an effective tool for transcribing mixtures of snare, kick drum and hi-hat, attempts to extend it to increased numbers of drum types have met with mixed results. To overcome this an automatic grouping method has been developed to group drum events on their similarity in frequency content. Combined with PSA this creates a system able to handle robustly greater numbers of drum types. The effectiveness of this method is demonstrated in a drum transcription algorithm.

1. Prior Subspace Analysis

Based on Independent Subspace Analysis (ISA) [1], Prior Subspace Analysis (PSA) is a technique for sound source separation in single channel mixtures where prior models of the sources are available [2].

A spectrogram **Y** is created from the mixture signal. PSA then assumes that **Y** results from the summation of l unknown independent spectrograms Y_j. It is assumed that the outer product of an invariant frequency subspace f_j, and a corresponding amplitude subspace t_j represents these Y_j. Summing the Y_j yields:

$$\mathbf{Y} = \sum_{j=1}^{l} Y_j = \sum_{j=1}^{l} f_j t_j^T \tag{1}$$

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

$$\mathbf{Y} \approx \sum_{j=1}^{l} f_p t_j^T \tag{2}$$

Multiplying the overall spectrogram by the pseudoinverse of a prior frequency subspace yields an estimate of the amplitude subspace, $\hat{t}_j$. Independent Component Analysis (ICA) is performed on the amplitude subspaces to make them independent, 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 [3]. These independent amplitude subspaces can then be used to obtain better estimates of the actual frequency subspaces, $\hat{f}_{ij}$. The independent spectrograms can then be estimated from

$$\hat{Y}_j = \hat{f}_{ij}\hat{t}_{ij}^T \tag{3}$$

Prior subspaces are obtained by analysing large numbers of each drum type using an ISA-type approach such as in [4]. First Principal Component Analysis (PCA) is carried out on the spectrogram of each sample. The first three principal components are analysed using ICA. The independent component with the largest projected variance is then used as the prior frequency subspace for that sample. K-means clustering is then carried out on the prior subspaces for a given drum type to yield a single prior frequency subspace for the drum type.

Prior Subspace Analysis has proved an effective method for transcribing mixtures of snare, kick drum and hi-hat/ride cymbal achieving a 92% success rate. However attempts to deal with further drum types have met with mixed success. This is due to the amount of frequency overlap between some drums such as toms and snares. This can result in similar amplitude subspaces for these drums, which causes the ICA algorithm to arrive at the wrong solution.

2. Automatic Grouping of Events

As PSA cannot robustly deal with signals containing more than three drum types, another approach is required. It is proposed to use ISA-type methods to model each event in a drum loop. Skinned drum events are detected by:

$$a_j = f_{ppj}\mathbf{Y} \tag{4}$$

where f_{ppj} is the pseudoinverse of skinned drum prior subspace j. The a_j are normalised and all peaks above a set threshold taken as events. The onset times of each event are determined, and the sections of the spectrogram between each event analysed individually.

PCA is performed on each section to yield a single frequency vector that is then normalised. The distance is then calculated between all pairs of vectors, where the distance between vectors a and b of size n is calculated from:

$$d_{ab} = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2} \tag{5}$$

For p events this results in $\boldsymbol{D}$**,** a p x p symmetric distance matrix. Starting from the first event, all events with $d_{1b} < 1$ are grouped and removed from the list of events. The next ungrouped event is then chosen and the above repeated until all events are grouped. If each event represents a single drum, this represents correct transcription of the excerpt, however this is not usually the case.

3. Drum Transcription using Automatic Grouping

A drum transcription algorithm using automatic grouping was implemented in Matlab. It is assumed that at least snares, kick drums and hi-hats/ride cymbals are present. The initial stage of the analysis proceeds as above, with the skinned drum events grouped by their similarity in frequency to other events. To overcome the most common skinned drum overlap, that of snare and kick drum, the groups corresponding to these drums are identified. The snare group is identified as the group that contains the largest peak in the initial estimate of the snare amplitude envelope. The kick drum group is identified as the group with the lowest spectral centroid. Any remaining groups are identified as toms. All tom events in the spectrogram are masked and PSA 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 overlap is a less common occurrence than that of snare and kick drum.

Power Spectral Density normalisation is carried out on the original spectrogram to eliminate the effects of the skinned drums as much as possible. The normalised spectrogram is multiplied by a prior hi-hat subspace to recover the metallic drum events. However both snare and tom drum events can also appear in the resulting amplitude envelope. To overcome this kick drum events are masked in the original spectrogram, and the resulting spectrogram is multiplied by a snare frequency subspace. The resulting subspace will also contain the tom events. ICA is then performed on the resulting amplitude subspace and that of the hi-hat subspace. All events above a threshold in the resulting hi-hat subspace are taken as metallic drum events.

The metallic drum events are then grouped. Due to interference from other drums no simple threshold suffices for grouping the drums. To overcome this 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. If two large groups occur that do not overlap in time then both hi-hat and ride cymbal are taken to occur, and these groups are kept separated. Otherwise all events are grouped together. This is justified in that most drummers tend to stay on either hi-hat or ride cymbal for long periods, changing only when the piece changes from one section to another, such as from verse to chorus. As a result overlapping groups are most likely to occur as a result of 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, of which both can occur in at any point. As the algorithm has no means of distinguishing between hi-hats and ride cymbals the groups are labeled metallic drums 1 and 2.

4. Results

The drum transcription algorithm was tested on 25 drum loops. A wide variety of drum sounds, drum patterns, tempos and meters were used to make the tests as realistic as possible. The same analysis parameters were used on all test signals. The results are summarised in Table 1.

Table 1. Drum Transcription Results.

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

The majority of the missing drums resulted from events that fell below the threshold for detection, though some were as a result of kick drums and toms being confused. The extra detections were mainly as a result of incorrect separation of the metallic and snare/tom subspaces. The automatic grouping performed remarkably well with all detected events being grouped correctly. All errors occurred elsewhere in the algorithm. The success rate of 89% compares well with the 92% achieved by PSA in more restricted conditions. It should also be noted that the algorithm works without any form of rhythmic modeling.

5. Conclusions and Future Work

Automatic grouping in conjunction with PSA has been shown to be an effective method for drum transcription. Future work will concentrate on increasing the generality of the system, in particular extending it to identify groups as hi-hats or ride cymbals and to deal with crash cymbals and open hi-hats.

References

1. M. Casey, & A. Westner,, “Separation of Mixed Audio Sources By Independent Subspace Analysis”, Proceedings Of ICMC 2000, pp. 154-161
2. D. FitzGerald, E. Coyle, B. Lawlor. “Prior Subspace Analysis for Drum Transcription”, 114th AES Conference Amsterdam March 22nd–25th 2003
3. A. Hyvärinen & E. Oja, “Independent Component Analysis: Algorithms and Applications”. Neural Networks, 13(4-5): pp. 411-430, 2000.
4. M. Casey, “Auditory Group Theory: with Applications to Statistical Basis Methods for Structured Audio”, ch. 4, pp. 81-124, Ph.D. thesis, MIT 1998.