2009-01-01

Using Tensor Factorisation Models to Separate Drums from Polyphonic Music

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

Matt Cranitch
Cork Institute of Technology, matt.cranitch@cit.ie

Eugene Coyle
Technological University Dublin, Eugene.Coyle@tudublin.ie

Audio Research Group

Articles

Dublin Institute of Technology *Year* 2009

Using Tensor Factorisation Models to Separate Drums from Polyphonic Music

Derry Fitzgerald* Matt Cranitch†

Eugene Coyle‡

*Dublin Institute of Technology, derry.fitzgerald@dit.ie

†Cork Institute of Technology, matt.cranitch@cit.ie

‡Dublin Institute of Technology, Eugene.Coyle@dit.ie

— Use Licence —

Attribution-NonCommercial-ShareAlike 1.0

You are free:

- to copy, distribute, display, and perform the work
- to make derivative works

Under the following conditions:

- Attribution.
  You must give the original author credit.
- Non-Commercial.
  You may not use this work for commercial purposes.
- Share Alike.
  If you alter, transform, or build upon this work, you may distribute the resulting work only under a license identical to this one.

For any reuse or distribution, you must make clear to others the license terms of this work. Any of these conditions can be waived if you get permission from the author.

Your fair use and other rights are in no way affected by the above.

USING TENSOR FACTORISATION MODELS TO SEPARATE DRUMS FROM POLYPHONIC MUSIC

Derry FitzGerald, *

Audio Research Group
Dublin Institute of Technology
Kevin St, Dublin Ireland
derry.fitzgerald@dit.ie

Eugene Coyle,

Audio Research Group
Dublin Institute of Technology
Kevin St, Dublin Ireland
eugene.coyle@dit.ie

Matt Cranitch,

Dept. of Electronic Engineering
Cork Institute of Technology
Bishopstown, Cork Ireland
matt.cranitch@cit.ie

ABSTRACT

This paper describes the use of Non-negative Tensor Factorisation models for the separation of drums from polyphonic audio. Improved separation of the drums is achieved through the incorporation of Gamma Chain priors into the Non-negative Tensor Factorisation framework. In contrast to many previous approaches, the method used in this paper requires little or no pre-training or use of drum templates. The utility of the technique is shown on real-world audio examples.

1. INTRODUCTION

The separation of drum instruments from polyphonic music signals has numerous applications including remixing, DJing tools,and rhythm-driven effects. Other possible uses include automatic transcription of drum parts, beat tracking, rhythmic description, and style identification.

In the past few years, numerous approaches to this problem have been proposed. Some approaches are based on the idea of separating the harmonic portion of the signal from the non-harmonic portion of the signal, via techniques such as sinusoidal modelling [1] or via noise subspace projection techniques [2]. The utility of these approaches is that they require no pretraining on the audio signals, but they suffer from limitations in that some parts of the drum sounds can be modelled by sinusoidal coefficients and that the attacks of some instruments will also be captured as part of the noise signal.

A spectral modulation approach was used by Barry et al [3] whereby a transient detector was used to modulate an audio spectrogram to separate out the percussive parts of the spectrogram. While capable of giving good rejection of the pitched instruments, the timbre of the recovered drums was noticeably different from those of the original drums. However, it should be noted that this method did not require any pre-training.

Other approaches include that of Helen [4], which used non-negative matrix factorisation to decompose the input spectrogram into various components. The resultant components were then classified as either pitched or percussive based on a pre-trained support vector machine, and these components used to resynthesise the separated drum track. However, in some cases the components separated contained elements of the pitched sounds. Another matrix factorisation based approach was developed by Gillet et al [5], where they used a separate sets of templates to model the background music and the drum parts of the signals. The drum templates were augumented with templates derived from the signal obtained using the subspace projection drum separation method described in [2] which helped improve the quality of the separations.

Another commonly used approach is that of template adaptation. Zils et al used simple time-domain templates for snare and kick drum which were then adapted iteratively using an anlysis by synthesis approach to extract a drum track containing these two drums [6]. However this approach suffered from a limitation in that it could not deal with simultaneous occurrences of these drums.

A more advanced template adaptation scheme was used by Yoshii et al in the context of both drum transcription [7] and drum sound remixing [8]. In these papers initial seed templates consisting of spectrograms of the various drums to be separated are iteratively adapted to provide a better match to the drums in the input signal. Harmonic component suppression is also used to further improve the results obtained. Finally, Itoyama et al. extended the idea of template adaptation to deal with both pitched and unpitched instruments, though this technique did require initialisation of the model parameters using a midi file of the piece to be separated [9]

In this paper, the method we propose for the separation of drum tracks from polyphonic audio is a version of the tensor factorisation models developed by the authors in [10] and [11] modified with the addition of gamma chain priors in a manner similar to that proposed by Virtanen in [12]. The model attempts to separate the harmonic parts of the signal from the non-harmonic parts of the signal, though unlike previous models which attempt this directly, it does so in the context of matrix and tensor factorisation techniques which have been shown to have good performance for the purposes of sound source separation. However, unlike previ-

* This author was supported by the Science Foundation Ireland Stokes Lectureship program

ous factorisation based methods, this technique does not require the use of pre-training, though prior information about the sources can easily be incorporated if required.

The remainder of the paper is organised as follows:, subsection 1.1 describes the notation conventions to be used throughout the paper, section 2 describes the separation model developed in this paper, section 3 presents separations obtained using the algorithm and 4 provides some conclusions and directions for future work.

1.1. Notation

For the remainder of this paper, all tensors, regardless of the number of dimensions, are signified by the use of caligraphic letters such as $\mathcal{A}$. $\langle \mathcal{AB} \rangle_{\{a,b\}}$ denotes contracted tensor multiplication of $\mathcal{A}$ and $\mathcal{B}$ along the dimensions a and b of $\mathcal{A}$ and $\mathcal{B}$ respectively. Outer product multiplication is denoted by $\circ$. Indexing of elements within a tensor is notated by $\mathcal{A}(i, j)$ as opposed to using subscripts. This notation follows the conventions used in the Tensor Toolbox for Matlab, which was used to implement the following algorithm [13]. For ease of notation, as all tensors are now instrument or source specific, the subscripts are implicit in all tensors within summations. Elementwise multiplication is denoted by $\otimes$ and all division is taken as elementwise.

2. TENSOR FACTORISATION MODEL

The model used in this paper is a simplified version of that proposed in [11], where shift-invariance in time has been eliminated for the model. Shift-invariance in time was removed for the purposes of separating pitched and unpitched sources as it was found that it allowed the unpitched section of the signal to capture aspects of the pitched part, which otherwise would belong in the pitched part of the model.

Given an r-channel mixture, magnitude spectrograms are obtained for each channel, resulting in $\mathcal{X}$, an $r \times n \times m$ tensor where n is the number of frequency bins and m is the number of time frames. The tensor is then modelled as:

$$\mathcal{X} \approx \hat{\mathcal{X}} = \sum_{k=1}^{K} \mathcal{G} \circ \langle \langle \langle \mathcal{FH} \rangle_{\{2,1\}} \mathcal{W} \rangle_{\{3,1\}} \mathcal{S} \rangle_{\{2,1\}} + \sum_{l=1}^{L} \mathcal{M} \circ \mathcal{B} \circ \mathcal{C} \quad (1)$$

where $\hat{\mathcal{X}}$ is an approximation to $\mathcal{X}$. The first right-hand side term models pitched sources, and the second unpitched or percussion sources. K denotes the number of pitched sources and L denotes the number of unpitched sources.

$\mathcal{G}$ is a tensor of size r, containing the gains of a given pitched source in each channel. $\mathcal{F}$ is of size $n \times n$, where the diagonal elements contain a filter which attempts to model the formant structure of an instrument, thus allowing the timbre of the instrument to alter with frequency. $\mathcal{H}$ is a tensor of size $n \times z_k \times h_k$ where z_k and h_k are respectively the number of allowable notes and the number of harmonics used to model the kth instrument, and where $\mathcal{H}(:, i, j)$ contains the frequency spectrum of a sinusoid with frequency equal to the jth harmonic of the ith note. $\mathcal{W}$ is a tensor of size h_k containing the harmonic weights for the kth source. $\mathcal{S}$ is a tensor of size $z_k \times m$ which contains the activations of the z_k notes associated with the kth source, and in effect contains a transcription of the notes played by the source. For the separation of signals containing pitched instruments only, best results were obtained when the lowest note played by each instrument was used as the lowest note in the source harmonic dictionary $\mathcal{H}$.

For unpitched instruments, $\mathcal{M}$ is a tensor of size r containing the gains of an unpitched source in each channel. $\mathcal{B}$ is of size n and contains a frequency basis function which models the timbre of the unpitched instrument. $\mathcal{C}$ is a tensor of size m which contains the activations of the lth unpitched instrument.

It can be seen that to obtain an estimate of the pitched sources only the first right hand side term of eqn 1 needs to be recontructed, and for the unpitched sources, only the second right hand side term needs to be used. The model can also be collapsed to the single channel case by eliminating both $\mathcal{G}$ and $\mathcal{M}$ from the model.

The generalised Kullback-Leibler divergence is used as a cost function to measure reconstruction of the original data as it has been shown to be effective for audio sound source separation [10]:

$$D\left(\mathcal{X} \parallel \hat{\mathcal{X}}\right) = \sum \mathcal{X} \log \frac{\mathcal{X}}{\hat{\mathcal{X}}} - \mathcal{X} + \hat{\mathcal{X}} \quad (2)$$

where summation takes place over all dimensions of $\hat{\mathcal{X}}$. Using this measure, iterative multiplicative update equations can be derived for each of the model variables in a manner similar to that described in [10]. Non-negativity is achieved through initialising the variables to non-negative values and the use of multiplicative updates. From these, separation of pitched and unpitched instruments can be attempted. However, considerably improved results can be obtained by the incorporation of gamma priors into the model in a manner similar to that proposed by Virtanen in the context of non-negative matrix factorisation (NMF) [12]. This is discussed in the following subsection.

2.1. Gamma Priors

Virtanen introduces two types of gamma priors, the first encourages temporal continuity to recover source activations which vary slowly in time, while the second allows the incorporation of prior knowledge about the frequency characteristics of sources in a simple and intuitive manner. In the context of separating pitched and unpitched instruments, incorporating temporal continuity in the pitched part of the model should favour the recovery of pitched sources as these will be more slowly varying in time than the unpitched sources which are typically more transient in nature, while incorporating prior knowledge of the characteristics of the unpitched sources could potentially aid recovery of these sources.

Temporal continuity is encouraged through the use of a gamma-chain. Adapting the notation used by Virtanen, the gamma distribution is defined for $y > 0$ as:

$$G(y; a, b) = y^{a-1} b^{-a} e^{-y/b} / \Gamma(a) \quad (3)$$

where $\Gamma(a)$ is the gamma (generalised factorial) function The gamma chain is then constructed through the use of an auxilary tensor $\mathcal{Z}$ of size $z_k \times m + 1$, which is defined as follows:

$$\begin{aligned} \mathcal{Z}(i,1) &\sim G(\mathcal{Z}(i,1); a+1, (ab)^{-1}) \\ \mathcal{S}(i,\tau) \mid \mathcal{Z}(i,\tau) &\sim G(\mathcal{S}(i,\tau); a, (\mathcal{Z}(i,\tau)a)^{-1}) \\ \mathcal{Z}(i,\tau) \mid \mathcal{S}(i,\tau) &\sim G(\mathcal{Z}(i,\tau+1); a+1, (\mathcal{S}(i,\tau)a)^{-1}) \end{aligned} \quad (4)$$

where τ indicates the time index in frames and lies between 1 and n, and i indexes over $1 : z_k$. In this context a acts as a coupling parameter between frames, and larger values of a result in more strongly coupled adjacent frames.

Again, using the method proposed by Virtanen, priors over the unpitched frequency basis functions $\mathcal{B}$ are derived assuming each entry of each prior is independently drawn from a Gamma distribution:

$$p(\mathcal{B}(v)) = G(\mathcal{B}(v); \alpha_v, \beta_v^{-1}) = \mathcal{B}(v)^{\alpha_v - 1} \beta_v^{\alpha_v} e^{-\mathcal{B}(v)\beta_v} / \Gamma(\alpha_v) \quad (5)$$

The hyperparameters α_v and β_v can be chosen independently for each source, and a simple interpretation of β_v^{-1} is as a set of weights which describe the typical frequency content of a given source. Therefore, in this context β_v^{-1} could be a typical frequency spectrum of an unpitched instrument such as a snare drum. For example, the priors used for drum transcription using Prior Subspace Analysis (PSA) [14] can be used as β_v^{-1}. However, unlike PSA, the frequency basis functions are still free to adapt during analysis.

2.2. Update Equations

Incorporating the above gamma priors into a cost function with the generalised Kullback Liebler divergence results in a extended cost function given by:

$$D\left(\mathcal{X} \parallel \hat{\mathcal{X}}\right) + \log(p(\mathcal{Z}, \mathcal{S})) + \log(p(\mathcal{B})) \quad (6)$$

From this iterative multiplicative update equations can be derived in a manner similar to those obtained in [12]. Defining $\mathcal{D} = \mathcal{X}/\hat{\mathcal{X}}$ and $\mathcal{O}$ as an all ones tensor the same size as $\mathcal{X}$, the update equations are given below for all the parameters in the model:

$$\mathcal{G} = \mathcal{G} \otimes \frac{\langle \mathcal{D} \langle \langle \mathcal{F}\mathcal{H} \rangle_{\{2,1\}} \mathcal{W} \rangle_{\{3,1\}} \mathcal{S} \rangle_{\{2,1\}} \rangle_{\{2:3,1:2\}}}{\langle \mathcal{O} \langle \langle \mathcal{F}\mathcal{H} \rangle_{\{2,1\}} \mathcal{W} \rangle_{\{3,1\}} \mathcal{S} \rangle_{\{2,1\}} \rangle_{\{2:3,1:2\}}} \quad (7)$$

$$\mathcal{F} = \mathcal{F} \otimes \frac{\langle \langle \mathcal{G}\mathcal{D} \rangle_{\{1,1\}} \langle \langle \mathcal{T}\mathcal{W} \rangle_{\{3,1\}} \mathcal{S} \rangle_{\{2,1\}} \rangle_{\{2,2\}}}{\langle \langle \mathcal{G}\mathcal{O} \rangle_{\{1,1\}} \langle \langle \mathcal{T}\mathcal{W} \rangle_{\{3,1\}} \mathcal{S} \rangle_{\{2,1\}} \rangle_{\{2,2\}}} \quad (8)$$

$$\mathcal{W} = \mathcal{W} \otimes \frac{\langle \langle (G \circ \langle \mathcal{F}\mathcal{H} \rangle_{\{2,1\}}) \mathcal{D} \rangle_{\{1:2,1:2\}} \mathcal{S} \rangle_{\{[1,3],1:2\}}}{\langle \langle (G \circ \langle \mathcal{F}\mathcal{H} \rangle_{\{2,1\}}) \mathcal{O} \rangle_{\{1:2,1:2\}} \mathcal{S} \rangle_{\{[1,3],1:2\}}} \quad (9)$$

$$\mathcal{S} = \mathcal{S} \otimes \frac{(2a/\mathcal{S}) + \langle (G \circ \langle \langle \mathcal{F}\mathcal{H} \rangle_{\{2,1\}} \mathcal{W} \rangle_{\{3,1\}}) \mathcal{D} \rangle_{\{1:2,1:2\}}}{(a\mathcal{Q}) + \langle (G \circ \langle \langle \mathcal{F}\mathcal{H} \rangle_{\{2,1\}} \mathcal{W} \rangle_{\{3,1\}}) \mathcal{O} \rangle_{\{1:2,1:2\}}} \quad (10)$$

where

$$\mathcal{Q} = \mathcal{Z}(i, 1 : m) + \mathcal{Z}(i, 2 : m + 1) \quad (11)$$

and

$$\mathcal{Z}(i, \tau) = \begin{cases} 1/(\mathcal{S}(i, 1) + b), & \tau = 1 \\ 2/(\mathcal{S}(i, \tau) + \mathcal{S}(i, \tau - 1)), & 1 < \tau < m + 1 \\ 1/\mathcal{S}(i, \tau) & \tau = m + 1 \end{cases} \quad (12)$$

In line with Virtanen b is set to zero in the above equation, as this results in the gain prior becoming independent of the overall level of the gains, while a was set to 100 as this appeared to give good results as suggested in [12].

$$\mathcal{M} = \mathcal{M} \otimes \frac{\langle \mathcal{D}(\mathcal{B} \circ \mathcal{C}) \rangle_{\{2:3,1:2\}}}{\langle \mathcal{O}(\mathcal{B} \circ \mathcal{C}) \rangle_{\{2:3,1:2\}}} \quad (13)$$

$$\mathcal{B} = \mathcal{B} \otimes \frac{(\alpha_v - 1)/\mathcal{B} + \langle \langle \mathcal{M}\mathcal{D} \rangle_{\{1,1\}} \mathcal{C} \rangle_{\{2,1\}}}{\beta_v + \langle \langle \mathcal{M}\mathcal{O} \rangle_{\{1,1\}} \mathcal{C} \rangle_{\{2,1\}}} \quad (14)$$

Again, following Virtanen, α_v is set to one and so the update equation for $\mathcal{B}$ only utilises the prior frequency characteristics provided.

$$\mathcal{C} = \mathcal{C} \otimes \frac{\langle (\mathcal{M} \circ \mathcal{B}) \mathcal{D} \rangle_{\{1:2,1:2\}}}{\langle (\mathcal{M} \circ \mathcal{B}) \mathcal{O} \rangle_{\{1:2,1:2\}}} \quad (15)$$

3. EXPERIMENTS

In the context of separating pitched instruments from unpitched instruments, the separation of individual pitched instruments is not a priority, and so the harmonic dictionaries for the sources do not have to be set as accurately as for the separation of individual instruments. To this end, the main focus is on covering a large number of pitches and so the number of pitched sources was set to 3, and each source covered a pitch range of 2 octaves or z_k=24, starting from 55 Hz for the lowest source. No overlap in allowable notes or pitch occurred between sources and so in total a 6 octave range was covered. The number of harmonics per note of each source was set to $h_k = 15$.

In many cases the number of pitched instruments, which often included singing voice, was greater than 3, but the algorithm showed sufficient flexibility to represent the overall set of pitched instruments using just a combination of 3 sources. In most cases, the bass line predominates in the lowest pitched source, and a small adjustment in z_k for this source is often sufficient to recover the bass line independently of other sources.

Similarly, the number of unpitched sources was set to 3, and in the majority of cases tested, the results obtained when using no priors were equivalent to those obtained using priors with regards to the overall recovery of the unpitched sources. Further investigation of the cases where using priors improved results showed that it was typically the kick drum that was recovered better when using priors, and that equivalent results could be obtained by just using a prior of a kick drum only, with no priors on the other drum sources. This highlights the utility of the model proposed in this paper over previous factorisation and template based approaches in that little or no training is required to separate the sources.

Rather than directly resynthesising the sources using the estimates obtained from the model in conjunction with the original mixture phase information, more natural sounding results were obtained by using a type of Wiener filtering, where the model estimates were used to filter the original complex spectrogram as shown in the equations below:

$$\hat{\mathcal{Y}}_p = \mathcal{Y} \otimes \frac{\hat{\mathcal{X}}_p}{\hat{\mathcal{X}}} \quad (16)$$

where $\mathcal{Y}$ is a tensor containing the original complex spectrograms, $\hat{\mathcal{Y}}_p$ contains the estimated overall pitched source spectrograms, and $\hat{\mathcal{X}}_p$ contains the overall pitched source spectrograms estimated from the model.

Similarly, the unpitched or drum sources are obtained from:

$$\hat{\mathcal{Y}}_d = \mathcal{Y} \otimes \frac{\hat{\mathcal{X}}_d}{\hat{\mathcal{X}}} \quad (17)$$

where $\hat{\mathcal{Y}}_d$ contains the estimated overall pitched source spectrograms, and $\hat{\mathcal{X}}_d$ contains the overall pitched source spectrograms estimated from the model.

Figure 1 shows an excerpt from "Rosanna" by Toto, together with the separated pitched instruments and the separated unpitched

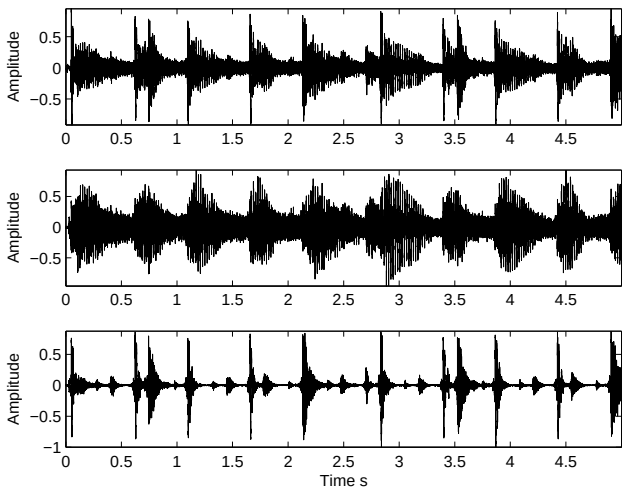

Figure 1: *(a) Original excerpt from "Rosanna" by Toto, (b) Separated pitched instruments, (c) Separated drums*

instruments. It can be seen from the separated waveforms that the prominent transients associated with the drums are absent from the pitched sounds, and that there is little evidence of pitched sounds in the separated drum track. On listening, the presence of the drums has been very much reduced in the separated pitched instruments, while the only evidence of the pitched instruments in the separated drums is the presence of the attack of the piano, while the timbre of the drums is a reasonable approximation to the original.

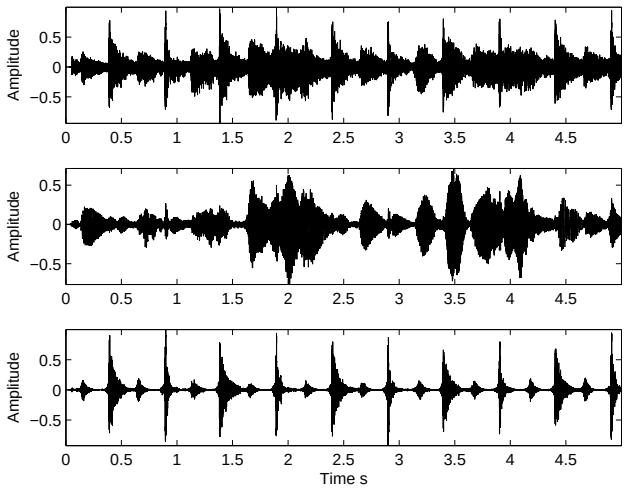

Figure 2: *(a) Original excerpt from "Like a Virgin" by Madonna, (b) Separated pitched instruments, (c) Separated drums*

Figure 2 shows an excerpt from "Like a Virgin" by Madonna, as well as the separated pitched instruments and the separated drums. Again, the prominent transients have been removed from the pitched instruments and no evidence of pitched instruments can be heard in the separated drum track, and the drum timbres have been preserved reasonably well.

Informal listening tests suggest that the incorporation of temporal continuity gives considerably improved results over those obtained without temporal continuity and that the method gives improved results over those obtained using the techniques described in [3, 4, 5]. Subjective listening tests will be completed in the very near future. However, in common with other techniques that attempt to separate harmonic from non-harmonic components, traces of the attacks of certain instruments such as pianos can be heard in the recovered unpitched spectrogram. Nonetheless the timbre of the recovered drum sounds is considerably closer to the original timbres in many cases over previous methods.

4. CONCLUSIONS

A tensor factorisation based technique for separating unpitched instruments from polyphonic music has been proposed. The technique imposes harmoncity constraints on the pitched instruments and extends previous tensor factorisation based techniques through the incorporation of temporal continuity constraints on the pitched instruments in the form of gamma chain priors.

An advantage of this technique over previous factorisation based and template based techniques is that the method requires little or no pretraining, with only a kick drum prior required in some cases, while in most cases good separation is obtained with no prior knowledge of the signals. However, it still suffers from some of the problems associated with methods that separate harmonic from non-harmonic portions of the signal, in that percussive attacks of pitched instruments also get separated in the non-harmonic portions of the signal.

Informal listening tests suggest that the proposed method gives better resynthesis quality and separation than previous approaches, and subjective listening tests will be completed in the very near future. Future work will also concentrate on trying to remove the problem of percussive attacks of pitched instruments.

5. REFERENCES

[1] X. Serra, *Musical Signal Processing*, chapter Musical Sound Modeling with Sinusoids plus Noise, pp. 91–122, G. D. Poli, A. Picialli, S. T. Pope, and C. Roads Eds. Swets & Zeitlinger, Lisse, Switzerland, 1996.

[2] O. Gillete and G. Richard, "Extraction and remixing of drum tracks from polyphonic music signals," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2005.

[3] D. Barry, D. FitzGerald, and E. Coyle, "Drum source separation using percussive feature detection and spectral modulation," in *Proc. IEE Irish Signals and Systems Conference*, 2005.

[4] M. Helen and T. Virtanen, "Separation of drums from polyphonic music using non-negative matrix factorisation and support vector machine," in *Proc. European Signal Processing Conference*, Anatalya, Turkey, 2005.

[5] O. Gillet and G. Richard, "Transcription and separation of drum signals from polyphonic music," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 3, pp. 529–540, 2008.

[6] A. Zils, F. Pachet, O. Delerue, and F. Gouyon, "Automatic extraction of drum tracks from polyphonic music signals," in *Proc. Wedelmusic*, Darmstadt, Germany, 2002.

[7] K. Yoshii, M. Goto, and H. Okuno, "Drum sound recognition for polyphonic audio signals by adaptation and matching of spectrogram templates with harmonic structure suppression," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 1, pp. 333–345, 2007.

[8] K. Yoshii, M. Goto, and H. Okuno, "Inter:d a drum sound equaliser for controlling volume and timbre of drums," in *Proc. European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies*, 2005.

[9] K. Itoyama et al, "Integration and adaptation of harmonic and inharmonic models for separating polyphonic musical signals," in *Proc. IEEE Conference on Acoustics, Speech and Language Processing*, 2007.

[10] D. FitzGerald, M. Cranitch, and E. Coyle, "Extended nonnegative tensor factorisation models for musical sound source separation," *Computational Intelligence and Neuroscience*.

[11] D.FitzGerald, M. Cranitch, and E. Coyle, "Musical source separation using generalised non-negative tensor factorisation models," in *Workshop on Music and Machine Learning, International Conference on Machine Learning*, Helsinki, Finland, 2008.

[12] T. Virtanen, A. Cemgil, and S. Godsill, "Bayesian extensions to non-negative matrix factorisations for audio signal modelling," in *Proc. IEEE Conference on Acoustics, Speech and Language Processing*, 2008.

[13] B. Bader and T. Kolda, "Matlab tensor toolbox version 2.2," Available at http://csmr.ca.sandia.gov/ tgkolda/TensorToolbox/.

[14] D. FitzGerald, M. Cranitch, and E. Coyle, "Generalised prior subspace analysis for polyphonic pitch transcription," in *Proc. 8th Digital Audio Effects Conference (DAFX05)*, 2005.