2012

User Assisted Separation Using Tensor Factorisations

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

Recommended Citation

Fitzgerald, D. (2012) User assisted separation using tensor factorisations. *20th European Signal Processing Conference* (EUSIPCO-2012), Bucharest, Romania.

Funder: SFI

USER ASSISTED SEPARATION USING TENSOR FACTORISATIONS

Derry FitzGerald

Audio Research Group, School of Electrical Engineering Systems, Dublin Institute of Technology, Ireland

ABSTRACT

Recent research has demonstrated that user assisted techniques, where the user provides a "guide" version of the source to be separated, are capable of giving good sound source separation. Here the user sings or plays along with the target source, and the user input is used to guide the separation towards the source of interest. This is typically done in a factorisation framework, such as non-negative matrix factorisation. Here we extend such approaches to a tensor factorisation framework to deal with multichannel signals. Further, we demonstrate how this framework can be used to improve the output from other user assisted techniques, such as the Adress algorithm, where the user manually selects a region from the stereo space corresponding to a given source.

***Index Terms*—** Sound Source Separation, Non-negative Tensor Factorisation, User Assisted Separation

1. INTRODUCTION

Techniques such as Non-negative Matrix Factorisation(NMF) and related approaches have demonstrated considerable utility in the area of sound source separation due to their ability to give a part-based decomposition of audio spectrograms [1]. However, a notable shortcoming of the standard NMF model is that there are typically many more basis functions than sources, resulting in the need to cluster the basis functions to their associated sources.

In order to overcome this problem, more complicated models incorporating constraints such as source-filter modelling, harmonicity and temporal continuity have been proposed [2]. However, recent research has shown that providing user assistance to guide the factorisation can help overcome the clustering problem and give good separation results [3, 4]. Here, the user sings or plays along with the source to be separated. The frequency and time basis functions obtained from this guide source are used to push a subset of the mixture signal basis functions towards the targeted source, while the remaining mixture signal basis functions are free to adapt to the characteristics of the other sources.

Outside of the context of factorisation-based approaches to sound source separation, user assisted separation has also been performed on stereo recordings using a real-time version of the Adress algorithm [5], which uses gain scaling and phase cancellation to identify time-frequency bins associated with a given pan position in the stereo field. Here the user manually selects the desired pan position of the source to be separated, and all bins with a pan position within a user chosen distance from the desired pan position are assumed to belong to the source to be separated. As will be seen later, such an approach could be incorporated into a factorisation-based sound source separation framework.

In this paper, we extend user-assisted approaches NMF to deal with multichannel recordings through the use of Non-negative Tensor Factorisation (NTF). Section 2 describes briefly NTF, and introduces the user-assisted extension of NTF, as well as highlighting issues related to extending user-assisted algorithms to multichannel recordings. Section 3 evaluates the performance of the user assisted NTF algorithm. Following this, we give a short overview of the Adress algorithm, highlighting its potential use as a means of generating user-assisted information on the position of sources in the stereo field, as well as showing how the separations obtained via Adress can be improved by incorporating them into the user-assisted NTF framework. Finally, conclusions and areas for future work are highlighted.

2. NON-NEGATIVE TENSOR FACTORISATION

Non-negative Tensor Factorisation (NTF) is a generalisation of non-negative matrix factorisation to deal with multidimensional arrays [6]. It was first used for sound source separation in [7], where the signal model used was described as:

$$\mathcal{X} \approx \hat{\mathcal{X}} = \sum_{k=1}^{K} \mathbf{G}_{:k} \circ \mathbf{A}_{:k} \circ \mathbf{S}_{:k} \tag{1}$$

where $\mathcal{X}$ is a $c \times n \times m$ tensor containing the magnitude spectrograms of the c channel mixture with n the number of frequency bins, and m the number of frames. $\mathbf{G}$, $\mathbf{A}$ and $\mathbf{S}$ are matrices of size $c \times K$, $n \times K$ and $m \times K$ respectively, with K being the rank of the decomposition. $\mathbf{G}$ contains the gains of each factor in each channel, $\mathbf{A}$ contains the frequency basis functions, while $\mathbf{S}$ contains the time activations of these basis functions. Here $:k$ denotes the kth column of a given matrix. $\circ$ denotes outer product multiplication.

The author was supported by Science Foundation Ireland's Stokes Lectureship programme

The generalised Kullback-Leibler divergence was used as a cost function to factorise the spectrograms, and multiplicative update equations were derived. The algorithm assumed a linear instantaneous mixing model and that each source in the multichannel mixture occupied a unique point in the multichannel or stereo space. To this end, the recovered basis functions were then clustered according to their spatial position to yield the separated sources. However, clustering performance degraded with increased numbers of basis functions.

In order to incorporate user assistance into NTF we make use of gamma priors, proposed originally in [8], later adapted for user-assisted NMF in [4]. Taking the frequency basis functions in $\mathbf{A}$ as an example, it is assumed that each entry in the basis function is drawn independently from a Gamma distribution, yielding:

$$\begin{aligned} p(\mathbf{A}_{i,k}) &= \mathcal{G}(\mathbf{A}_{i,k} : \alpha_{i,k}, \beta_{i,k}^{-1}) \\ &= \mathbf{A}_{i,k}^{\alpha_{i,k}-1} \beta_{i,k}^{\alpha_{i,k}} e^{\mathbf{A}_{i,k}\beta_{i,k}} / \Gamma(\alpha_{i,k}) \end{aligned} \tag{2}$$

where $\mathcal{G}$ denotes the Gamma distribution defined for $\mathbf{A} > 0$, with hyperparameters $\alpha_{i,k}$ and $\beta_{i,k}^{-1}$ which can be chosen independently for each frequency basis function in $\mathbf{A}_{1:n,k}$. $\beta_{i,k}^{-1}$ can be interpreted as a set of weights which describe the typical or expected frequency spectrum of a given source, which are used to push the basis function towards a desirable set of frequency characteristics. Similar distributions can be defined for $\mathbf{S}$ and $\mathbf{G}$. The NTF cost function is then extended to:

$$D(\mathcal{X} \parallel \hat{\mathcal{X}}) + \lambda_{\mathbf{A}} \log(p(\mathbf{A})) + \lambda_{\mathbf{S}} \log(p(\mathbf{S})) + \lambda_{\mathbf{G}} \log(p(\mathbf{G})) \tag{3}$$

with $D(\mathcal{X} \parallel \hat{\mathcal{X}})$ the generalised Kullback-Leibler divergence. Here $\lambda_{\mathbf{A}}$,$\lambda_{\mathbf{S}}$ and $\lambda_{\mathbf{G}}$ are parameters used to control the influence of the prior on the factorisation. Initially these are set to 1 to ensure the factorisation is pointed towards the required source characteristics, but are then gradually reduced with each iteration to allow the factorisation to adapt to the characteristics of the actual source present, as opposed to those of the guide source.

In the user assisted framework, taking the frequency basis functions as an exemplar, $\beta_{i,k}^{-1}$ are obtained from a non-negative factorisation on the guide source spectrogram. Collecting the hyperparameters into a matrix, $\boldsymbol{\beta}_{\mathbf{A}}$ is then set as:

$$\boldsymbol{\beta}_{\mathbf{A}} = 1/\mathbf{A}_u \tag{4}$$

where $\mathbf{A}_u$ are the frequency basis functions recovered from the guide source. Typically $\alpha_{i,k}$ is set to 1. Multiplicative updates for $\mathbf{A}$ can then be derived yielding:

$$\mathbf{A} = \mathbf{A} \otimes \frac{\langle \mathcal{D}\mathcal{Q} \rangle_{\{[1,3],[1,2]\}}}{\langle \mathcal{O}\mathcal{Q} \rangle_{\{[1,3],[1,2]\}} + \lambda_{\mathbf{A}} \boldsymbol{\beta}_{\mathbf{A}}} \tag{5}$$

Here $\langle \bullet \rangle_{\{a,b\}}$ denotes contracted tensor product along the dimensions contained in a and b as per the conventions described in [9]. $\mathcal{D}$ is defined as $\mathcal{X}/\hat{\mathcal{X}}$, $\mathcal{O}$ is an all ones tensor the same size as $\mathcal{X}$ and $\mathcal{Q}$ is a tensor of size $c \times m \times K$ where $\mathcal{Q}(:,:,k) = \mathbf{G}_{:k} \circ \mathbf{S}_{:k}$. $\otimes$ denotes elementwise multiplication and all divisions are elementwise.

Hyperparameters can be derived similarly for $\mathbf{S}$ and $\mathbf{G}$:

$$\boldsymbol{\beta}_{\mathbf{S}} = 1/\mathbf{S}_u, \quad \boldsymbol{\beta}_{\mathbf{G}} = 1/\mathbf{G}_u \tag{6}$$

The update equations for $\mathbf{G}$ and $\mathbf{S}$ are then given by

$$\mathbf{G} = \mathbf{G} \otimes \frac{\langle \mathcal{D}\mathcal{P} \rangle_{\{[2,3],[1,2]\}}}{\langle \mathcal{O}\mathcal{P} \rangle_{\{[2,3],[1,2]\}} + \lambda_{\mathbf{G}} \boldsymbol{\beta}_{\mathbf{G}}} \tag{7}$$

$\mathcal{P}$ is a tensor of size $n \times m \times K$ with $\mathcal{P}(:,:,k) = \mathbf{A}_{:k} \circ \mathbf{S}_{:k}$.

$$\mathbf{S} = \mathbf{S} \otimes \frac{\langle \mathcal{D}\mathcal{R} \rangle_{[1,2],[1,2]}}{\langle \mathcal{O}\mathcal{R} \rangle_{[1,2],[1,2]} + \lambda_{\mathbf{S}} \boldsymbol{\beta}_{\mathbf{S}}} \tag{8}$$

$\mathcal{R}$ is a tensor of size $c \times n \times K$ with $\mathcal{R}(:,:,k) = \mathbf{G}_{:k} \circ \mathbf{A}_{:k}$.

In the user assisted NTF (UA-NTF) algorithm, the priors are applied to a subset of the basis functions. These basis functions should then capture the source targeted by the guide signal. The remaining basis functions are free to adapt to the other sources, and their updates can be obtained by setting λ to 0 in the above equations. Spectrograms for the target source and the remaining sources can then be estimated by applying eqn 1 to the required subset of the basis functions. These are used to generate a Wiener-type filter mask to apply to the original spectrogram before inversion to the time domain.

There are several issues with extending user-assisted approaches to deal with multi-channel signals. The first is that it is reasonable to assume that the user-assisted guide signal will be a single channel recording, and therefore contains no spatial information about the source to be separated. Even if a stereo guide was made, it would be difficult for the user to replicate the stereo placement properly. This leaves open the issue of how to handle the spatial components of the user assisted source. The first option is to leave the basis functions free to adapt to any spatial position as determined by the optimisation. The second is to assume that the source to be separated is a point source, and so all associated basis functions should share a common spatial position. The final option is to assume that the spatial position can be estimated by the user through other means, such as via the Adress algorithm (see section 4), which can then be incorporated either via the use of a prior or by fixing the gains to this position. The effectiveness of these options will be evaluated in the next section.

3. UA-NTF PERFORMANCE EVALUATION

In order to test how UA-NTF operated under the different configurations described above, a test set of 10 signals was created using material taken from [10]. Here, both the stereo instrumental backing tracks and the acapella vocals for these tracks were available separately. Mono excerpts of solo vocals were mixed with the stereo backing track to create stereo mixes where the vocal was positioned in the centre channel. Then a user was recorded singing along to the lead vocal, and these recordings were used to guide the separations. All recordings and mixes were done at a sampling frequency of 44.1 kHz.

Magnitude spectrograms were then obtained for each channel of the mixes and used to create a spectrogram tensor describing the mixture. A spectrogram was also created for the guide vocal recording. In all cases, a window/FFT size of 4096 samples and a hopsize of 1024 samples was used. 100 basis functions were used to capture the information related to the guide source, while another 100 basis functions were used to capture the remaining sources. The total number of iterations used was 20, with the influence of the user-assisted prior reducing linearly to zero over the course of these iterations. 20 iterations was chosen as tests on user assisted separation using mono recordings suggest that 20 iterations was optimal in terms of separation performance [11].

Several versions of the algorithm were tested. Firstly, the gain parameters of the user-assisted basis functions were allowed to adapt freely to independent positions in the stereo space (denoted Free in Table 1). Secondly, the gain parameters were allow adapt freely under the influence of a prior on the gains (FreeP). In the third version, all the user-assisted basis functions were constrained to have a common spatial position which adapted over the optimisation (Group).

The third test was the same as the second test, but with the addition of a prior on the common spatial position which was provided by the user (GroupP). The fourth again used prior information on the spatial position, but in this case the position was used directly as the gain for the user-assisted basis functions and was held fixed throughout the optimisation (Force).

The outputs from each of these tests were then evaluated using the PEASS toolbox version 2 [12], and the results obtained, averaged across the 10 mixes are presented in Table 1 for both the user-assisted source and the separated backing track. It can be seen that in all cases UA-NTF is capable of giving reasonable results, with the best results obtained for Overall Perceptual Score (OPS) when the basis functions for the target source are constrained to having a common spatial position (Force), followed by when the source basis functions are guided by a prior, but are free to adapt individually (FreeP). Grouping the basis functions in conjunction with the use of a prior on the source position gives the worst results, suggesting that in this case, the factorisation is overconstrained. This is borne out on listening to these separations, where elements from other sources positioned at the same point are more noticeable than with other separation configurations.

	OPS	TPS	APS	IPS	SDR	SIR	SAR	ISR
Free	24.8	19.3	25.7	59.0	2.8	4.7	14.1	6.6
Free P	25.1	19.7	27.8	57.8	2.9	4.6	14.2	6.8
Group	25.6	18.7	26.2	47.9	1.0	1.7	16.2	7.3
Group P	22.2	37.7	52.0	21.3	0.0	0.1	18.8	12.3
Force	24.9	15.5	22.1	63.6	3.8	6.6	14.0	6.4

Table 1. Performance Evaluation of NTF. Free denotes spatial gains are free to adapt individually, Free P is as free, but with the addition of priors on the spatial gains. Group denotes that the spatial gain is the same for all user-assisted basis functions. Group P denotes the use of a prior on this spatial gain, while Force indicates that the spatial gain was held constant for all user-assisted basis functions

4. THE ADRESS ALGORITHM AND NTF

In this section we highlight the fact that user-assisted versions of NTF have another potential use. This is that that outputs from another sound source separation technique can be used as an input to these techniques in order to obtain improved separations. This is particularly beneficial in cases where the initial separation technique is not based on global properties of the mixture, as is the case with NTF based algorithms, but is instead based on more local characteristics of the mixture signal. Source separation algorithms in this category include the Adress algorithm [5], where the separation is estimated on a frame by frame basis. We will use Adress as an exemplar to illustrate the use of user-assisted NTF algorithms in bootstrapping the performance of other separation techniques.

The Adress algorithm performs real-time sound source separation on linear instantaneous stereo mixtures and separates sources based on their pan position through the use of frequency domain gain scaling. A Short Time Fourier Transform (STFT) is carried out on each of the mixture signals. A frequency-azimuth plane is obtained for each channel by scaling and subtraction of each channel for a range of azimuth gain values. At azimuth positions where a source is present, the energy in the frequency bins associated with a given source will be cancelled out resulting in a minimum at that position in the azimuth frequency plane. This minimum represents the residual energy present due to other sources in the mixture. The energy of the source present can then be estimated as the difference between this minimum and the actual energy present at that time frequency bin.

However, due to frequency overlap between sources, the position of a frequency minimum can drift away from that of the actual source position. As a result, all bins which have minima within a given distance of the chosen azimuth position are taken as belonging to a source at that azimuth position, and their energies estimated. All bins outside this distance are set to zero. The estimated source spectrogram is then inverted to the time-domain. The interested reader is referred to [5] for a more detailed explanation of the Adress algorithm.

The Adress algorithm typically requires user assistance to achieve separation of the sources, where the user selects the source position in the azimuth plane, as well as the distance around the azimuth position. Therefore, the user-selected source position could potentially be used to guide the spatial position of the target source in UA-NTF. However as seen above, this may not be required in many cases.

With regards to the separations obtained via Adress, there is a trade-off in selecting the azimuth distance, with wider distances capturing more of the bins associated with a given source, while at the same time allowing increased bleed from other sources. Any bin which has energy associated with a given source but which falls outside the azimuth subspace will have no energy in the reconstructed source spectrogram and so separations obtained using Adress typically have gaps in otherwise well recovered harmonics due to smearing in the azimuth-frequency plane. This is often audible in the form of artifacts in the recovered source. Further, the separation obtained with Adress varies locally with time as the frequencies at which overlap occur change with time. This is in contrast to NTF based approaches which are based on global characteristics of the mixture. Therefore, it is hoped that the global nature of NTF-based separation will compensate for local errors in separation associated with Adress.

We therefore propose to take the source spectrogram recovered from Adress and use it to guide a factorisation of the original stereo mixture signal. Here, instead of the user singing or playing along, the user provides assistance through choosing the parameters to separate a source using Adress. This provides two pieces of information, firstly an estimated source spectrogram, and secondly, an estimate of the source position in stereo space.

Rather than use a gamma-prior approach, we use instead a partial cofactorisation technique, another method for incorporating prior knowledge into the factorisation process [13]. This is because preliminary testing has shown this approach to perform better in this application. In this case, a set of temporal and frequency basis functions are held in common between the factorisation of the Adress estimate and the original mixture signal, with the remaining basis functions associated with the mixture signal free to adapt to the characteristics of other sources in the mixture. We assume that the bins estimated by Adress are accurate estimations and so their influence should be maintained throughout the factorisation process, though this influence could also be reduced in a manner similar to that described in Section 2 if desired. Further, we assume that the estimate of the source position obtained from Adress is also accurate and this is passed directly to the factorisation process. This results in the following signal model:

$$\mathcal{X} \approx \hat{\mathcal{X}} = \sum_{k=1}^{K} \mathbf{G}_{:k} \circ \mathbf{A}_{:k} \circ \mathbf{S}_{:k} + \mathbf{P} \circ (\mathbf{WH}) \qquad (9)$$

$$\mathbf{Y} \approx \hat{\mathbf{Y}} = \mathbf{WH} \qquad (10)$$

where $\mathbf{W}$ is an $n \times q$ matrix containing a set of frequency basis functions shared between the mixture and the source estimate spectrogram obtained from Adress with q denoting the number of shared basis functions. $\mathbf{H}$ is a $q \times m$ matrix containing the associated time basis functions. $\mathbf{P}$ is a 2×1 matrix containing a common gain in each channel for the shared basis functions, which is obtained via the Adress algorithm. Finally, $\mathbf{Y}$ is the estimated source spectrogram obtained via Adress. This results in an extended cost function:

$$D(\mathcal{X} \parallel \hat{\mathcal{X}}) + D(\mathbf{Y} \parallel \hat{\mathbf{Y}}) \qquad (11)$$

Multiplicative update equations can then be derived for the additional model variables as:

$$\mathbf{W} = \mathbf{W} \otimes \frac{\langle\langle \mathbf{P}\mathcal{D}\rangle_{\{1,1\}}\mathbf{H}\rangle_{\{2,2\}} + \mathbf{Q}\mathbf{H}^{\mathsf{T}}}{\langle\langle \mathbf{P}\mathcal{O}\rangle_{\{1,1\}}\mathbf{H}\rangle_{\{2,2\}} + \mathbf{R}\mathbf{H}^{\mathsf{T}}} \qquad (12)$$

where $\mathbf{Q} = \mathbf{Y}/\hat{\mathbf{Y}}$, $\mathbf{R}$ is an all ones matrix the same size as $\mathbf{Y}$ and T is matrix transpose.

$$\mathbf{H} = \mathbf{H} \otimes \frac{\langle \mathcal{D}(\mathbf{P} \circ \mathbf{W})\rangle_{\{[1,2],[1,2]\}} + \mathbf{W}^{\mathsf{T}}\mathbf{Q}}{\langle \mathcal{O}(\mathbf{P} \circ \mathbf{W})\rangle_{\{[1,2],[1,2]\}} + \mathbf{W}^{\mathsf{T}}\mathbf{R}} \qquad (13)$$

The updates for $\mathbf{G}$,$\mathbf{A}$ and $\mathbf{S}$ can be obtained from eqns (5),(7) and (8) by setting $\lambda = 0$.

5. ADRESS-UA-NTF EVALUATION

In order to test the effectiveness of the Adress-UA-NTF algorithm, a set of test signals was created from the Sisec 2011 developmental set for professionally produced music recordings [14]. This consisted of snippets from 5 songs from a number of genres. Here the stereo source tracks were all converted to mono to generate point sources as Adress assumes linear intantaneous mixing of mono sources. The resulting mono tracks were then spread equally across the stereo space to yield stereo mixtures.The only exception to this was the first snippet from the dev1 set, where the various drum tracks were mixed to a single source. This results in a total of 22 sources to separate from the 5 mixures. The maximum distance between the sources was limited to an azimuth width of 0.5, thereby ensuring that mixtures with low numbers of sources were not trivial to separate. The user then separated the sources using Adress, and a source width H of 0.2 was used for all sources. All snippets were at a samplerate of 44.1 kHz, and again an FFT/window size of 4096 samples and a hopsize of 1024 was used. The UA-NTF algorithm was then ran for 100 iterations. To make a direct comparison with Adress, resynthesis was obtained by applying the original phase information of the channel in which the source was dominant to the spectrogram obtained from UA-NTF. Again, 100 basis functions were allocated to the guide source,and another 100 basis functions to deal with the remaining sources.

	OPS	TPS	APS	IPS	SDR	SIR	SAR	ISR
Ad	19.6	28.9	15.4	66.7	2.5	7.1	9.9	2.0
Ad-UA-NTF	23.7	41.6	23.4	55.1	2.8	5.3	11.6	2.8

Table 2. Performance Evaluation of Adress-UA-NTF. Ad denotes the results obtained from the Adress algorithm directly, while Ad-UA-NTF denotes the results from using UA-NTF on the estimated source spectrograms obtained via Adress.

The results are shown in Table 2, and it can be seen that the use of UA-NTF improves the separations obtained directly from Adress, though at the cost of some extra bleed from the surrounding sources. Particularly noticeable on listening to the separated sources is the improvement in the sharpness of the transients on the percussion instruments. Audio examples of the separations obtained for the algorithms described in this paper can be found at [15].

6. CONCLUSIONS

Having discussed the concept of user-assisted separation using NMF, we then extended the approach to deal with multichannel signals through the use of NTF. We then demonstrated the effectiveness of the approach using a set of test signals, as well as highlighting issues related to the spatial position of the source to be recovered. We show that ensuring the user-assisted basis functions come from a common position gives the best results for UA-NTF. We then proposed the use of the Adress algorithm as a means of capturing the spatial position of the desired source, should that be desired, as well as demonstrating that the separations obtained from Adress can be improved by incorporating them into a user-assisted NTF framework. Future work will concentrate on the effects of using multiple guide sources simultaneously.

7. REFERENCES

[1] P. Smaragdis, "Discovering auditory objects through non-negativity constraints," in *Proc. of ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing*, 2004.

[2] D. FitzGerald, M. Cranich, and E. Coyle, "Extended non-negative tensor factorisation models for musical sound source separation," *Computational Intelligence and Neuroscience, vol. 2008, Article ID 872425*, 2008.

[3] P. Smaragdis and G. Mysore, "Separation by humming: User-guided sound extraction from monophonic mixtures," in *IEEE Workshop on Applications Signal Processing to Audio and Acoustics*, 2009.

[4] D. FitzGerald, "User assisted source separation using non-negative matrix factorisation," in *IET Irish Signals and Systems Conference*, Dublin, 2011.

[5] D. Barry, B. Lawlor, and E. Coyle, "Real-time sound source separation using azimuth discrimination and resynthesis," in *Proc.of 117th Audio Engineering Society Convention*, 2004.

[6] A. Shashua and T. Hazan, "Non-negative tensor factorization with applications to statistics and computer vision," in *International Conference of Machine Learning (ICML)*, 2005.

[7] D. FitzGerald, M. Cranitch, and E. Coyle, "Non-negative tensor factorisation for sound source separation," in *Proceedings of the Irish Signals and Systems Conference*, 2005.

[8] A. T. Cemgil T. Virtanen and S. J. Godsill, "Bayesian extensions to non-negative matrix factorisation for audio signal modelling," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008.

[9] B.W. Bader and T.G. Kolda, "Algorithm 862: Matlab tensor classes for fast algorithm prototyping," *ACM Transactions on Mathematical Software*, 2006.

[10] "The pet sounds sessions," The Beach Boys, 1997, Capitol 72438 37662 2 2.

[11] D. FitzGerald and R. Jaiswal, "On the use of masking filters in sound source separation," in *accepted for publication at 15th International Conference on Digital Audio Effects DAFx 2012*, 2012.

[12] V. Emiya, E. Vincent, N. Harlander, and V. Hohmann, "Subjective and objective quality assessment of audio source separation," *IEEE Transactions on Audio, Speech and Language Processing*, 2011.

[13] Jiho Yoo, Minje Kim, Kyeongok Kang, and Seungjin Choi, "Nonnegative matrix partial co-factorization for drum source separation," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, march 2010, pp. 1942 –1945.

[14] Sisec 2011, "http://sisec.wiki.irisa.fr/," .

[15] Examples, "http://eleceng.dit.ie/derryfitzgerald/index.php?uid=489&menu_id=61," .