ONLINE HARMONIC/PERCUSSIVE SEPARATION USING SMOOTHNESS/SPARSENESS CONSTRAINTS

F. Canadas-Quesada, P. Vera-Candeas, N. Ruiz-Reyes
Univ. de Jaén, Spain
fcanadas@ujaen.es

P. Alonso
Univ. Politécnica de Valencia, Spain
palonso@dsic.upv.es

J. Ranilla
Univ. de Oviedo, Spain
ranilla@uniovi.es

ABSTRACT

The separation of percussive sounds from harmonic sounds in audio recordings remains a challenging task since it has received much attention over the last decade. In a previous work, we described a method to separate harmonic and percussive sounds based on a constrained Non-negative Matrix Factorization (NMF) approach. The approach distinguishes between percussive and harmonic bases integrating percussive and harmonic sound features, such as smoothness and sparseness, into the decomposition process. In this paper, we propose an online version of our previous work. Instead of decomposing the whole mixture, the online proposal decomposes a set of segments of the mixture selected by a sliding temporal window. Both percussive and harmonic bases of the next segment are initialized using the bases obtained in the decomposition of the previous segment. Results show that an online proposal can provide satisfactory separation performance but the sound quality of the separated signals depends inversely on the computation time of the system.

1. INTRODUCTION

Separating percussive sounds from harmonic sounds in music remains a challenging problem since it has received much attention over the last years. Percussive sounds, e.g. snare drum, are impulsive and have a structure that is vertically smooth in frequency and sparse in time. Harmonic sounds, e.g. bass, are quasi-stationary and have a structure that is horizontally smooth in time and sparse in frequency (see Fig 1). Several music information retrieval applications could benefit from this separation such as music transcription or onset detection.

Although many algorithms have been developed to separate percussive and harmonic sounds from monaural music [1] [2] [3] [4] [5], one of the trends in percussive and harmonic separation is based on the concept of anisotropic smoothness which is related to the difference in the directions of continuity between the spectrograms of harmonic and percussive sounds. Ono et al. [6] [7] separated harmonic and percussive sounds by exploiting the

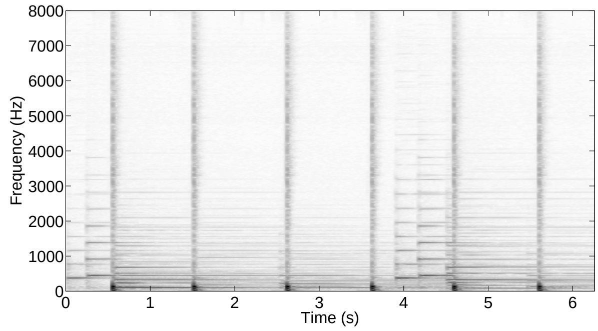

Figure 1. Magnitude spectrogram of a mixture composed of percussive and harmonic sounds. It can be seen that percussive sounds form vertical lines while the harmonic sounds form horizontal lines.

anisotropy of harmonic and percussive sounds in a maximum a posteriori (MAP) framework. Fitzgerald's system [8] extracted percussive sounds using the anisotropy smoothness by means of a median filtering. In this manner, the harmonics are considered to be outliers in a temporal slice. Recently, Canadas et. al [9] proposed a NMF approach that automatically distinguishes between percussive and harmonic bases by integrating spectro-temporal features, such as anisotropic smoothness or time-frequency sparseness, into the factorization process. Results were promising but the approach requires offline processing since it is necessary to decompose the whole mixture signal.

In this paper, we propose an online version of our previous work [9] where instead of decomposing the whole mixture, a set of segments of the mixture are decomposed using a sliding temporal window. Once a new segment is selected and decomposed by a constrained NMF, the sliding window is shifted by one segment. Using a small size of segment implies a faster NMF convergence. Percussive and harmonic bases are initialized randomly and updated in the decomposition of the first segment. However, the bases of the next segments are initialized using the bases obtained in the previous segment.

Consider the term latency as the time elapsed between receiving the input audio mixture and starting to perform separation in order to clarify the terms *offline*, *online* and *realtime*. The term *offline* indicates a latency equal to the duration of the whole input mixture because the whole input mixture is necessary to apply the constrained NMF [9]. The term *online* indicates a latency equal to the duration of the segment because only each segment is necessary to apply the constrained NMF [9] and obtain both percussive and harmonic signals related to the segment processed. However, none of the aforementioned terms provide a realtime separation. The term *realtime* indicates that the latency plus the computation time is about 30-40 milliseconds providing to the user the sense of an immediate output. The computation time is the time elapsed between starting to perform separation and obtaining each separated percussive and harmonic signals.

The remainder of the paper is organized as follows. Section 2 introduces NMF and its application to sound source separation. Section 3 describes briefly our previous offline harmonic/percussive separation work. Section 4 details the online harmonic/percussive separation proposal. Experimental results and performance analysis are shown in Section 5. Finally, conclusions are reported in Section 6.

2. NON-NEGATIVE MATRIX FACTORIZATION (NMF) FOR SOUND SOURCE SEPARATION

Non-negative Matrix Factorization (NMF) [10] is a technique for multivariate data analysis which aims to obtain a parts-based representation of objects, by imposing non-negative constraints. Given a matrix $\mathbf{X}$ of dimensions $F \times T$ with non-negative entries, it is possible to model it as linear combinations of K elementary non-negative spectra. Therefore, NMF is the problem of finding a factorization:

$$\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{WH}, \tag{1}$$

where $\hat{\mathbf{X}}$ is the estimated matrix, $\mathbf{W} \in R^{F\times K}$ is the matrix whose columns are the bases or components. These bases represent characteristic spectral patterns active in the input spectrogram. $\mathbf{H} \in R^{K\times T}$ is a matrix of component gains for all frames. These gains represent the temporal interval in which the spectral patterns are active. In typical audio applications, the matrix $\mathbf{X}$ is chosen as a time-frequency representation (e.g., magnitude or power spectrogram), $f = 1, ..., F$ denoting the frequency bin and $t = 1, ..., T$ the time frame.

In the case of magnitude spectra, the parameters are restricted to be non-negative, then, a common way to compute the factorization in eq. (1) is generally obtained by minimizing a cost function defined as

$$D(\mathbf{X}|\hat{\mathbf{X}}) = \sum_{f=1}^{F}\sum_{t=1}^{T} d(X_{ft}|\hat{X}_{ft}), \tag{2}$$

where $d(a|b)$ is a function of two scalar variables, d is typically non-negative and takes value zero if and only if $a = b$. Using the $\beta-$divergence cost [11], some of the most popular cost functions are the Euclidean distance (β=2), the generalized Kullback-Leibler divergence (β=1) and the Itakura-Saito divergence (β=0). The cost functions are non-increasing using $1 \leq \beta \leq 2$ [12]. In practice, Févotte et.al [11] observed that the criterion is still non-increasing for β <1 and β >2 but no proof is available. An iterative algorithm based on multiplicative update rules is proposed to obtain the model parameters that minimize the cost function. In general, the update rules can be defined as follows [11],

$$\mathbf{H} \leftarrow \mathbf{H} \odot \frac{\mathbf{W}^T((\mathbf{WH})^{\beta-2} \odot \mathbf{X})}{\mathbf{W}^T(\mathbf{WH})^{\beta-1}}, \tag{3}$$

$$\mathbf{W} \leftarrow \mathbf{W} \odot \frac{((\mathbf{WH})^{\beta-2} \odot \mathbf{X})\mathbf{H}^T}{(\mathbf{WH})^{\beta-1}\mathbf{H}^T}, \tag{4}$$

where $\mathbf{W}$ and $\mathbf{H}$ are initialized as random positive matrices, ^{T} is the transpose operator, $\odot$ represents the Hadamard (element-wise) multiplication and the division is also element-wise.

Considering that the source z is composed of a set of L components, the separated magnitude spectrogram of the source z can be reconstructed as follows,

$$X_z = \frac{\sum_{i=1}^{L} \mathbf{W_i H_i}}{\mathbf{WH}} \odot \mathbf{X}, \tag{5}$$

where the temporal signal $x_z(t)$ is computed inverting the spectrogram X_z to the time-domain using the phase of the original mixture.

3. OFFLINE HARMONIC/PERCUSSIVE SEPARATION

Unconstrained NMF cannot discriminate between percussive and harmonic bases. To overcome this problem, we proposed [9] an unsupervised system that can separate percussive and harmonic sounds in monaural music integrating percussive and harmonic sound features into the NMF decomposition. For that purpose, an objective function is defined to decompose a mixture spectrogram X into two separated spectrograms, X_P (a percussive spectrogram) and X_H (a harmonic spectrogram). Each separated spectrogram exhibits specific spectro-temporal features for percussive or harmonic sounds. The factorization model is given in eq. (6),

$$X \approx X_P + X_H = W_P H_P + W_H H_H, \tag{6}$$

where X_P, X_H, W_P, H_P, W_H and H_H are non-negative matrices.

The percussive constraints used to model percussive sounds assume smoothness in frequency (the energy slowly decreases in frequency) and sparseness in time (most of the signal energy is concentrated over short time intervals). Two constraints, spectral smoothness SSM and temporal sparseness TSP, are associated to the percussive matrix W_P.

The harmonic constraints used to model harmonic sounds assume smoothness in time (amplitudes that vary slowly in time) and sparseness in frequency (spectral peaks). Two constraints, spectral sparseness SSP and temporal smoothness TSM, are associated to the harmonic matrix W_H.

The global cost function D uses the β-divergence cost d_β, the percussive constraints (SSM, TSP) and the harmonic constraints (SSP, TSM),

$$\begin{aligned} D = d_\beta(X|(X_P + X_H)) + K_{SSM}SSM+ \\ +K_{TSP}TSP + K_{TSM}TSM + K_{SSP}SSP, \end{aligned} \tag{7}$$

where the parameters $K_{SSM}, K_{TSP}, K_{TSM}, K_{SSP}$ determine the degree of control of each constraint in the NMF procedure. However, the system requires offline processing since it is necessary to decompose the whole mixture signal X. More details can be found in [9].

4. ONLINE HARMONIC/PERCUSSIVE SEPARATION

We extend our offline harmonic/percussive separation work to the case online. In the online proposal, a constrained NMF [9] is not applied to the whole mixture spectrogram X. The whole mixture signal of duration T seconds is segmented in $L = \left\lceil \frac{T}{T_i} \right\rceil$ segments S_i using a non-overlapped sliding window of duration T_i seconds as can be seen in Fig. 2.

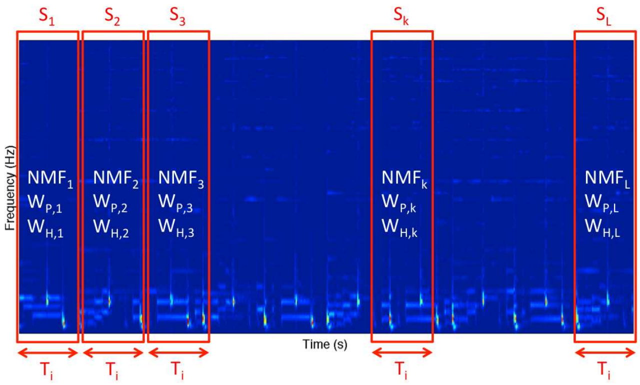

Figure 2. Online percussive/harmonic proposal. It can be seen that each segment S_i is decomposed using a constrained NMF [9] obtaining the percussive matrix $W_{P,i}$ and the harmonic matrix $W_{H,i}$ related to the magnitude spectrogram of the segment S_i.

When the magnitude spectrogram X_1 of the first segment S_1 is computed, the matrices $W_{P,1}$, $W_{H,1}$, $H_{P,1}$ and $H_{H,1}$ are initialized randomly and then are updated using [9]. In this manner, we obtain the matrices of the percussive $W_{P,1}$ and harmonic $W_{H,1}$ bases related to the first segment S_1. Next, the sliding window is shifted by one segment and a new segment is selected to apply the method [9]. Taking into account the next segment S_i, the gains matrices are randomly initialized but the bases are initialized using the bases obtained from the previous segment S_{i-1}, that is $W_{P,i} = W_{P,i-1}$, $W_{H,i} = W_{H,i-1}$ for $i = 2...L$. The idea is to use a better initialization in the next segment from the values of the bases obtained in the previous segment. The reason is because these previous bases reflect properties of percussive and harmonic sounds and it could find a better minimum local in the NMF decomposition [13]. A small number of iterations is sufficient to the NMF convergence due to the size of the spectrogram related to the segment S_i is relatively small compared to the whole spectrogram X of the mixture.

Instead of reconstructing the whole mixture as occurs in [9], each percussive $x_{p,i}(t)$ or harmonic $x_{h,i}(t)$ temporal signal related to the segment S_i is reconstructed. The separated percussive signal $x_{p,i}(t)$, composed of $L_{p,i}$ percussive components, can be synthesized inverting into the time domain via inverse Short Time Fourier Transform (STFT) using the phase of the segment S_i of the mixture. In a similar way, the harmonic signal $x_{h,i}(t)$ is synthesized taking into account the harmonic components $L_{h,i}$.

Considering a segment S_i, the time T_r defines the computation time to obtain both separated percussive $x_{p,i}(t)$ and harmonic $x_{h,i}(t)$ signals associated with the segment S_i.

5. EXPERIMENTAL RESULTS

5.1 Data set, metrics and State-of-the-art methods

A data set, composed of nine monaural real-world music excerpts, taken from the Guitar Hero game [14] [15], has been created to evaluate the performance of the proposed method as can be seen in Table 1. Each music excerpt contains percussive and harmonic instruments and a duration about T=30 seconds. All of the signals were converted from stereo to mono and sampled at 16 kHz.

Identifier	Title	Artist
$M1$	Hollywood Nights	Bob Seger & The Silver Bullet Band
$M2$	Hotel California	Eagles
$M3$	Hurts So Good	John Mellencamp
$M4$	La Bamba	Los Lobos
$M5$	Make It Wit Chu	Queens Of The Stone Age
$M6$	Ring of Fire	Johnny Cash
$M7$	Rooftops	Lost prophets
$M8$	Sultans of Swing	Dire Straits
$M9$	Under Pressure	Queen

Table 1. Identifier, Title and Artist of the files of the database

The assessment of the performance of the online proposal has been performed using the metrics Signal to Distortion Ratio (SDR), Signal to Interference Ratio (SIR) and Signal to Artifacts Ratio (SAR) [16] [17] widely used in the field of sound source separation. Higher values of these ratios indicate better separation quality.

The separation performance of the online proposal is evaluated using different durations T_i (seconds) of the segment: Online-1 ($T_i = 1$), Online-2 ($T_i = 2$), Online-3 ($T_i = 3$), Online-5 ($T_i = 5$), Online-10 ($T_i = 10$) and Online-15 ($T_i = 15$). Moreover, these online proposals are compared with the offline method [9] (the offline method assumes that the whole mixture has a duration of T seconds) and the two recent state-of-the-art percussive and harmonic sound separation methods. The first one is the method HPSS [7] and the second one is the method MFS [8].

5.2 Parameters

The normalization process [9] is applied taking into account not the size of the whole mixture but the size of the segment. Most of the parameters used in [9] are also used in the online proposal. Specifically, a frame size N =1024 samples, a time shift J =512 samples and the optimum values β = 1.5, $K_{TSP} = K_{SSP}$ = 0.1 and $K_{TSM} = K_{SSM}$ = 0.2. The convergence of the online proposal is empirically observed using a number of iterations $MaxIter$ = 50. Compared to the offline version, the online proposal needs a lower number of iterations to converge due to the lower size of the spectrogram of the segment to decompose.

Considering the optimum number of percussive $R_{p_{offline}}$ and harmonic $R_{h_{offline}}$ components used in [9], the number of percussive $R_{p_{online}}$ and harmonic $R_{h_{online}}$ components used in the online proposal is computed as,

$$R_{p_{online}} = \left\lfloor \frac{T_i \cdot R_{p_{offline}}}{T} \right\rfloor, \quad (8)$$

$$R_{h_{online}} = \left\lfloor \frac{T_i \cdot R_{h_{offline}}}{T} \right\rfloor, \quad (9)$$

It seems that a lower number of percussive and harmonic components will be necessary to decompose a segment of shorter duration since a lower number of sources will be active. More details can be found in [9].

5.3 Results

Fig. 3 shows SDR (top figure) and SIR (bottom figure) percussive results evaluating the database for the online proposals and the offline method in function of the duration of the segment. Percussive SIR results improve using a very long duration ($T_i \geq 5$) of segment. Initially the percussive SDR and SIR improves using short segments but this improvement cannot be compared to the other online proposals with a long segment (see percussive SIR results). The initial SDR improvement of the Online-1 reports that using bases that contains typical features of percussive and harmonic sounds achieves to find a better minimum local in the NMF decomposition. This minimum local obtains a set of bases that reflect higher musical sense similar as percussive and harmonic sounds are perceived in the nature. The improvement in SIR between the proposal Online-1 and Online-2 is about 4dB in average so, SIR results improve significantly using a segment of duration $T_i > 1$. A drawback of the proposal Online-1 is that it captures a high amount of harmonic sounds in the separated percussive signal but these harmonic sounds are attenuated if a longer segment is used. The reason is because a longer segment provides more useful information to model correctly the sounds active in the mixture. The improvement in SIR between the proposal Online-2 and the others is approximately 2.5dB in average and approximately 1dB between the proposal Online-3 and the other online methods. Finally, the percussive performance using a segment of duration $T_i > 3$ is similar.

Fig. 4 shows SDR (top figure) and SIR (bottom figure) harmonic results evaluating the database for the online proposals and the offline method in function of the duration of the segment. Unlike in the percussive separation (as can be seen in Fig. 3), the online proposals show a slight negative slope in the harmonic SDR and SIR results. This behavior could indicate that some harmonic bases that correctly model the harmonic sounds of the previous segment are replaced by the new bases obtained in the update process of the next segment providing a fluctuation of the harmonic

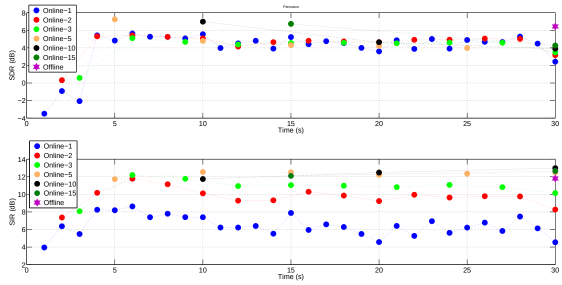

Figure 3. SDR (top) and SIR (bottom) percussive separation performance related to the offline method and online proposals.

SDR and SIR along the segments. This effect of replacement seems to be more critical in the separation of harmonic sounds compared to percussive sounds (see Fig. 3 and Fig. 4).

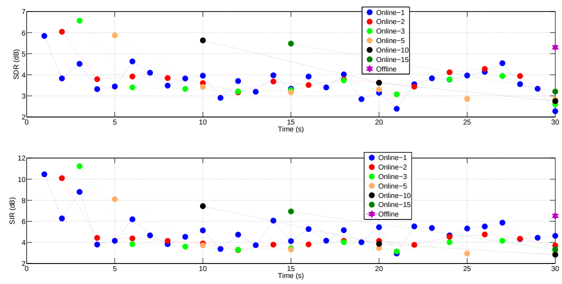

Figure 4. SDR (top) and SIR (bottom) harmonic separation performance related to the offline method and online proposals.

Fig. 5 shows SDR and SIR separation performance related to the offline method, online proposals and the two state-of-the-art percussive and harmonic sound separation methods. Each percussive bar is computed using the mean of all of the separated percussive signals of the database evaluated. In a similar way, each harmonic bar is computed taking into account the separated harmonic signals. Each group of bars in the left figure refers to percussive results and each group of bars in the right figure refers to harmonic results. Comparing the offline method and the online proposals, the best percussive and harmonic SDR and SIR results are obtained by the offline method. As can be seen, a longer duration of the segment shows a better separation providing higher quality of the separated signals. It means that a constrained NMF improves the separation performance using sufficient information of the mixture in order to model correctly the sounds active. Although the percussive SDR and the harmonic SDR and SIR are similar using the method Online-1 and Online-2, the method Online-2 achieves a significant percussive SIR improvement of about 4dB compared to the method Online-1. This improvement reports that a segment of duration equal to one second cannot model correctly harmonic sounds. For this reason, a high amount of harmonic sounds are captured in the separated percussive signal by the method Online-1 minimizing the percussive SIR as shown in Fig. 5 (bottom).

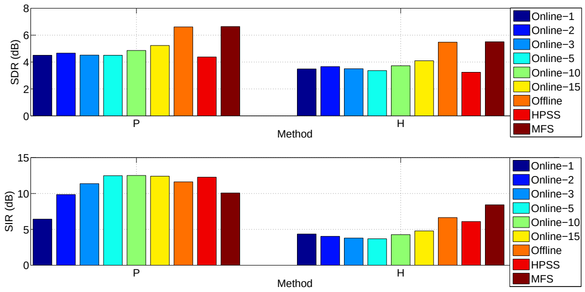

Figure 5. SDR (top) and SIR (bottom) percussive and harmonic separation performance related to the online, offline and state-of-the-art percussive and harmonic sound separation methods. The letter P in the *x-axis* refers to percussive results while the letter H in the *x-axis* refers to harmonic results.

The computation time of the online proposals is shown in Table 2 which has been computed using Matlab on a PC with Intel Core i5 CPU of 2.5 GHz and 4 GB of RAM. It can be observed that the computation time of the online proposals increases with the size of the segment. Moreover, the processing factor P_F defined as the ratio between the computation time and the duration of the segment also increases with the size of the segment. Specifically, it ranges approximately from a $1/5$ to $2/3$ of the duration of the segment processed. Although the method Online-15 obtains the best SDR and SIR results comparing the online proposals, the method Online-3 can be considered the best choice because it provides the best trade-off between sound quality and computation time.

Online proposal	Computation time T_r (sec)	$P_F = \frac{T_r}{T_i}$
Online-1 (1sec)	0.25	0.25
Online-2 (2sec)	0.43	0.22
Online-3 (3sec)	0.64	0.22
Online-5 (5sec)	1.12	0.22
Online-10 (10sec)	3.02	0.30
Online-15 (15sec)	5.85	0.39
Offline (30sec)	19.82	0.66

Table 2. The computation time of each online proposal

6. CONCLUSIONS

We extend our offline harmonic/percussive separation work [9] to the case online to separate harmonic and percussive sounds in monaural music. Instead of decomposing the whole mixture, a set of segments of the mixture are decomposed using a sliding temporal window. Once a new segment is selected and decomposed by a constrained NMF, the sliding window is shifted by one segment. Using a small size of segment implies a faster NMF convergence. Percussive and harmonic bases are initialized using the bases obtained in the NMF decomposition of the previous segment. The idea is to use a better initialization in the next segment with bases that reflect properties of percussive and harmonic sounds.

Percussive and harmonic SDR and SIR results show that a longer duration of the segment shows a better separation providing higher quality of the separated signals. It means that a constrained NMF improves the separation performance using sufficient information of the mixture in order to model correctly the sounds active.

The initialization of percussive and harmonic bases using spectral patterns that model the energy distribution of these types of sounds seems to find a better minimum local in the NMF decomposition. This better minimum local means that it provides bases that reflect higher musical sense similar as percussive and harmonic sounds are perceived in the nature. A drawback of the online proposals is that some bases that correctly model the percussive or harmonic sounds of the previous segment are replaced by the new bases obtained in the update process of the next segment providing a fluctuation of the separation performance. This replacement of bases with “good properties” is more critical in the separation of harmonic sounds compared to percussive sounds.

In the future, we plan to work on two extensions to improve the sound quality of the online proposal. The first extension is based on measuring the similarity between consecutive segments. In this manner, the percussive and harmonic bases of a segment will be initialized with random values if a low similarity is obtained regarding to the previous segment. However, the percussive and harmonic bases of a segment will be initialized with the harmonic and percussive bases computed in the previous segment if a high similarity is obtained regarding to the previous segment. The second extension is based on a new update of percussive and harmonic bases. The idea is to keep fixed along the segments those bases that have correctly model percussive or harmonic sounds in previous segments.

Acknowledgments

This work was supported by the Andalusian Business, Science and Innovation Council under project P2010- TIC-6762 (FEDER) and the Spanish Ministry of Economy and Competitiveness under Projects TEC2012-38142-C04-01, TEC2012-38142-C04-03 and TEC2012-38142-C04-04.

7. REFERENCES

[1] L. Daudet, “Review on techniques for the extraction of transients in musical signals,” in *Proceedings of the Third international conference on Computer Music Modeling and Retrieval*, 2005, pp. 219–232.

[2] M. Helen and T. Virtanen, “Separation of drums from polyphonic music using non-negative matrix factorisation and support vector machine,” in *Proceedings of European Signal Processing Conference*, Anatalya, Turkey, 2005, pp. –.

[3] O. Gillet and G. Richard, “Transcription and separation of drum signals from polyphonic music,” in *IEEE*

Transactions on Audio, Speech, and Language Processing, vol 3, no. 16, 2008, pp. 529–540.

[4] A. Ozerov, E. Vincent, and F. Bimbot, "A general flexible framework for the handling of prior information in audio source separation," in *IEEE Transactions on Audio, Speech, and Language Processing, vol 20, no. 4*, 2012, pp. 1118–1133.

[5] J. Yoo, M. Kim, K. Kang, and S. Choi, "Nonnegative matrix partial co-factorization for drum source separation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010, pp. –.

[6] N. Ono, K. Miyamoto, H. Kameoka, and S. Sagayama, "A real-time equalizer of harmonic and percussive components in music signals," in *Proceedings of the Ninth International Conference on Music Information Retrieval (ISMIR)*, 2008, pp. 139–144.

[7] N. Ono, K. Miyamoto, J. L. Roux, H. Kameoka, and S. Sagayama, "Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram," in *Proceedings of the European Signal Processing Conference*, 2008, pp. 25–29.

[8] D. Fitzgerald, "Harmonic/percussive separation using median filtering," in *Proceedings of Digital Audio Effects (DAFX)*, 2010, pp. –.

[9] F. Canadas, P. Vera, N. Ruiz, J. Carabias, and P. Cabanas, "Percussive/harmonic sound separation by non-negative matrix factorization with smoothness/sparseness constraints," in *Journal on Audio, Speech, and Music Processing, vol. 2014, no. 26*, 2014, pp. 1–17.

[10] D. Lee and H. Seung, "Algorithms for non-negative matrix factorization," in *Proc. of Advances in Neural Inf. Process. System*, 2000, pp. 556–562.

[11] C. Févotte, N. Bertin, and J. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence. with application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, 2009.

[12] R. Kompass, "A generalized divergence measure fon nonnegative matrix factorization," *Neural Computation*, vol. 19, no. 3, pp. 780–791, 2007.

[13] B. Zhu, W. Li, R. Li, and X. Xue, "Multi-stage non-negative matrix factorization for monaural singing voice separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 10, pp. 2096–2107, 2013.

[14] Activision, "Guitar hero world tour," in *http://en.wikipedia.org/wiki/Guitar_Hero_World_Tour*.

[15] ——, "Guitar hero 5," in *http://en.wikipedia.org/wiki/Guitar_hero_5*.

[16] E. Vincent, C. Févotte, and R. Gribonval, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.

[17] C. Févotte, R. Gribonval, and E. Vincent, "Bss_eval toolbox user guide - revision 2.0, technical report 1706," in *IRISA, April*, 2005, pp. –.