NON-NEGATIVE MATRIX FACTORISATION INCORPORATING GREEDY HELLINGER SPARSE CODING APPLIED TO POLYPHONIC MUSIC TRANSCRIPTION

Ken O'Hanlon, Mark Sandler

Centre for Digital Music
Queen Mary University of London

Mark D. Plumbley

Centre for Vision, Speech and Signal Processing
University of Surrey

ABSTRACT

Non-negative Matrix Factorisation (NMF) is a commonly used tool in many musical signal processing tasks, including Automatic Music Transcription (AMT). However unsupervised NMF is seen to be problematic in this context, and harmonically constrained variants of NMF have been proposed. While useful, the harmonic constraints may be constrictive in mixed signals. We have previously observed that recovery of overlapping signal elements using NMF is improved through introduction of a sparse coding step, and propose here the incorporation of a sparse coding step using the Hellinger distance into a NMF algorithm. Improved AMT results for unsupervised NMF are reported.

Index Terms— Non-negative matrix factorisation, sparse coding, Hellinger distance, music transcription

1. INTRODUCTION

Given a matrix with all positive entries, $\mathbf{V} \in \mathbb{R}^{M \times N}$, Non-negative Matrix Factorisation (NMF) seeks the approximation

$$\mathbf{V} \approx \mathbf{WH} \quad s.t. \quad \mathbf{W}, \mathbf{H} \geq 0 \qquad (1)$$

where $\mathbf{W} \in \mathbb{R}^{M \times K}$ is a dictionary matrix with an atom in each column and $\mathbf{H} \in \mathbb{R}^{K \times N}$ is an activation matrix, with each row containing the activations for the corresponding atom. NMF is a popular tool used in many application areas, including face recognition[1], musical signal processing [2] [3] and hyperspectral imaging [4], amongst others.

A variety of different approaches to NMF problem exist, the earliest of which [5] used Non-negative Least Squares (NNLS) to alternately update $\mathbf{W}$ and $\mathbf{H}$. Multiplicative update (MU) algorithms, equivalent to fixed step gradient descent, were proposed by Lee & Seung [1] for the Euclidean and Kullback Leibler (KL) divergence cost functions, and have been extended to families of measures of fit such as the generalised β- divergence [6]. Problems with MU approaches include slow convergence and zero-locking, whereby a coefficient set to zero is maintained in that state. A general problem in NMF is the issue of separability, first noted in [7] which implies that the factors may need to be linearly independent to be recovered correctly.

Stepwise, sparse methods may help avoid some issues with local minima in NMF [8]. Non-negative K-SVD (NN-K-SVD) [9] and NMF-ℓ_0 [8], which seek to optimise an ℓ_0-penalised NNLS cost function, use methods such as Orthogonal Matching Pursuit (OMP) to estimate an active set. NN-K-SVD then uses a variant of the K-SVD sparse dictionary learning algorithm to update each dictionary atom and its activations simultaneously, while NMF-ℓ_0 uses NMF to update the dictionary and activations with the active set maintained due to zero-locking.

NMF is often used to factorise spectrograms in tasks such as Automatic Music Transcription (AMT) [2] [10]. However, overlap of musical signal elements due to harmonicity and the logarithmic structure of the musical scale presents difficulty to NMF. This separability issue in musical spectrograms was noted in the first paper on NMF for AMT [2]. Some of the more successful methods for NMF-based AMT use fixed dictionaries [11], or harmonically constrained dictionaries [10] [12]. While useful, these constrained methods may not be so effective when presented with data which is not amenable, and improved unsupervised NMF methods are desirable.

Incorporation of stepwise approaches into NMF presents one avenue of enquiry in the search for such methods. However, methods such as NN-K-SVD and NMF-ℓ_0 use Euclidean distance, generally considered to perform poorly for musical signals [2] [10]. A stepwise algorithm for KL-divergence, referred to as ASNA, was recently proposed [13]. However, we have observed [14] that the gradient of the KL-divergence, used as a selection criteria in ASNA, is badly conditioned, and unsuitable for sparse coding. For this case a nearest neighbour selection based on the Hellinger distance, which similar to KL is an α-divergence [6], was proposed [14]. We now propose a sparse NMF algorithm incorporating the Hellinger sparse approximation step. In the rest of this paper, NMF with α-divergence and ASNA are briefly outlined, before the proposed method is described. Experimental results show improved AMT using the proposed method relative to MU approaches.

2. BACKGROUND

2.1. NMF with α-divergence

The α-divergence [6] is a generalised measure of similarity, given for two non-negative vectors

$$\mathcal{C}_\alpha(\mathbf{v}|\mathbf{z}) = \frac{1}{\alpha(1-\alpha)} \sum_m \alpha v_m + (1-\alpha)z_m - v_m^\alpha z_m^{1-\alpha}. \quad (2)$$

Special cases include the KL-divergence ($\alpha \to 1$)

$$\mathcal{C}_{KL}(\mathbf{v}|\mathbf{z}) = \sum_m v_m \log(\frac{v_m}{z_m}) - v_m + z_m \quad (3)$$

which is also a member of the better known β-divergence [15]. The inverse KL divergence $\mathcal{C}_{KL}(\mathbf{z}|\mathbf{v})$ is given as $\alpha \to 0$ and the Hellinger distance when $\alpha = \frac{1}{2}$:

$$\mathcal{C}_H(\mathbf{v}|\mathbf{z}) = 2\sum_m (\sqrt{v_m} - \sqrt{z_m})^2 \quad (4)$$

which is seen as the squared ℓ_2 norm of the square roots of two vectors. MUs for the α-divergence are given in [6] :

$$\mathbf{H} \longleftarrow \mathbf{H} \otimes \left[\frac{\mathbf{W}^T \left[\frac{\mathbf{V}}{\mathbf{WH}}\right]^{[\alpha]}}{\mathbf{W}^T \mathbf{1}_{M \times N}} \right]^{[\frac{1}{\alpha}]} \quad (5)$$

$$\mathbf{W} \longleftarrow \mathbf{W} \otimes \left[\frac{\left[\frac{\mathbf{V}}{\mathbf{WH}}\right]^{[\alpha]} \mathbf{H}^T}{\mathbf{1}_{M \times N} \mathbf{H}^T} \right]^{[\frac{1}{\alpha}]} \quad (6)$$

where $\otimes$ denotes elementwise multiplication, $\mathbf{X}^{[]}$ denotes elementwise exponentiation and all divisions are also elementwise. Other approaches such as coordinate descent can also be used for α-divergence [16].

2.2. Active Set Newton Algorithm

Recently a greedy active set method, employing the KL-divergence cost function has been proposed [13] for use with overcomplete dictionaries. This method, referred to as ASNA, is outlined in Algorithm 1. ASNA is similar to OMP [17], starting with an empty active set and iteratively adding an atom according to a selection criteria based on a signal estimate using the current active set. After each atom is added the current estimate is updated. In ASNA, the inactive atom with largest negative KL gradient

$$\hat{k} = \arg\min_k \mathbf{w}_k^T \left(\mathbf{1}_M - \frac{\mathbf{v}}{\mathbf{Wh}} \right) \quad (7)$$

is added to the active set, Γ, and the coefficient of the newly added atom is initialised to a small value (10^{-5}). The authors of [13] propose the use of two Newton steps

$$\mathbf{h} \longleftarrow \mathbf{h} + \mu \mathcal{H}^{-1}(\mathbf{h})\nabla(\mathbf{h}) \quad (8)$$

Algorithm 1 ASNA [13] and HSC

Input $\mathbf{v}, \mathbf{W}, \Gamma = \{\}$
Do
    Select atom, indexed by $\hat{k}$ using (7) (10)
    Add to active set and initialise
        $\Gamma \longleftarrow \Gamma \cup \hat{k}$;
        Set initial value of $h_{\hat{k}}$
    Re-estimate $\mathbf{h}$ using Newton step(s) (8)
Until stopping condition

---

to perform the coefficient estimation step where $\mathcal{H}(\mathbf{h})$ is the Hessian matrix, $\nabla(\mathbf{h})$ denotes the gradient and μ is a step size equal to 1 in the general case. A small valued diagonal matrix is added to the Hessian in order to better condition the inversion [13]. In ASNA the stepsize μ may be set to a smaller value than 1 in order to maintain the non-negative constraint $\mathbf{h} \geq 0$. This decreased stepsize is given by $\mu = \min_{v_k > 0} v_k$ where $\mathbf{v} = \mathbf{h} \oslash \mathcal{H}^{-1}(\mathbf{h})\nabla(\mathbf{h})$ where $\oslash$ denotes elementwise division. When used, this smaller stepsize results in at least one atom, presenting a zero coefficient, being removed from the active set. Faster convergence is reported using the Newton approach in comparison to multiplicative updates [13].

3. PROPOSED METHOD

We have observed difficulty in recovering toy harmonic atoms [18] using NMF with MUs, a problem that was ameliorated by incorporation of greedy pursuits, and consider that such modifications to NMF may improve AMT performance. Furthermore, we have recently observed improved non-negative sparse coding using a Hellinger distance related selection criteria [14] relative to the ASNA (7) and OMP [17] criteria in the case of musical signals. In light of these observations, an NMF algorithm incorporating Hellinger Sparse Coding (HSC) is proposed.

3.1. Hellinger Sparse Coding

A minimum gradient selection criteria similar to (7)

$$\hat{k} = \arg\min_k \mathbf{w}_k^T \mathbf{1} - \mathbf{w}_k^T \left[\frac{\mathbf{v}}{\mathbf{Wh}}\right]^{[\frac{1}{2}]} \quad (9)$$

may be used for the Hellinger distance. However, a nearest neighbour approach based on the cosine distance, $\frac{\mathbf{y}^T \mathbf{x}}{\|\mathbf{y}\|_2 \|\mathbf{y}\|_2}$, similar to OMP is proposed. As the residual $\mathbf{r} = \mathbf{v} - \mathbf{Wh}$ may contain negative values, an alternative vector given by $\bar{r}_m = \text{sgn}(r_m)\sqrt{|r_m|}$ is used, leading to the selection criteria

$$\hat{k} = \arg\max_k \frac{\mathbf{w}_k^{[0.5]T} \bar{\mathbf{r}}}{\sqrt{\mathbf{w}_k^T \mathbf{1}}} \quad (10)$$

which considers an extension of Hellinger distance for mixed signs described in [14] along with other extensions.

Otherwise, HSC proceeds similar to ASNA, with both described in Algorithm 1. Initialisation of the coefficient of the newly selected atom is performed using

$$h_{\hat{k}} = \left(\frac{\mathbf{w}_k^{[0.5]T} \bar{\mathbf{r}}}{\mathbf{w}_k^T \mathbf{1}} \right)^2. \tag{11}$$

The Hessian matrix for the Hellinger distance, given by

$$\mathcal{H} = \mathbf{W}_\Gamma^T \mathrm{diag} \left(\frac{\mathbf{v}^{[0.5]}}{[\mathbf{Wh}]^{[1.5]}} \right) \mathbf{W}_\Gamma \tag{12}$$

is used to perform a Newton update (8) with $\nabla(\mathbf{h}) = 2\mathbf{W}_\Gamma^T \left(1 - \frac{\mathbf{v}^{[0.5]}}{[\mathbf{Wh}]^{[0.5]}} \right)$ using a similar mechanism to ASNA to enforce the non-negative constraint. It is found that the use of a reasonable initialisation (11) for a newly added atom in HSC results in one full ($\mu = 1$) step of the Newton method (8) being sufficient.

3.2. NMF-HSC Algorithm

The HSC approach is incorporated into an NMF algorithm outlined in Algorithm 2. After initialisation of $\mathbf{W}$ as a positive matrix random matrix and $\mathbf{H}$ as a uniform valued matrix an iterative loop is entered. This main loop consists of a number, L, of iterations of the NMF updates (5) (6) using the Hellinger distance ($\alpha = \frac{1}{2}$) followed by application of HSC to reset the activation matrix $\mathbf{H}$ and ascertain the active set. After J iterations, the main loop exits at which point the dictionary, $\mathbf{W}$ is finalised. A further, final regression is performed with this fixed dictionary. In this way it is hoped to counter errors that may have occurred during the sparse coding step.

Algorithm 2 NMF-HSC

Input $\mathbf{V}, \mathbf{W}, \mathbf{H}, \alpha = 0.5$
For 1:J
  For 1:L
    Perform multiplicative updates (6) (5)
  End For
  Reset $\mathbf{H}$ by performing HSC
End For
Re-initialise $[H]_{k,n} = 0.01 \forall \{k, n\}$
Re-estimate $\mathbf{H}$ by iterating (5)

3.3. Post-processing for AMT

Some common steps are required to convert a NMF factorisation into a pitch-time representation, or piano roll.

3.3.1. Pitch Estimation

First, pitch estimation needs to be performed on each atom of the dictionary. Given the fundamental frequency for the pth

pitch on the piano scale note with MIDI number $p + 20$

$$f_o^p = 2^{\frac{p-49}{12}} \times 440$$

the expected frequency of the rth harmonic partial is given by $f_r = r f_0$. Pitch estimation is performed through weighted addition of the coefficients at the expected locations, and sidelobes, of harmonic partials of a note. Given that f_r is most closely associated with the ρth dimension of the frequency spectrum the strength of the pth pitch in the kth atom is

$$S_k^p = \frac{1}{R^p} \sum_{r=1}^{R^p} \frac{1}{\sqrt{r}} \sum_{e=\rho-1:\rho+1} [W]_{e,k} \tag{13}$$

where $R^p = f_s/(2 \times f_0)$ is the number of partials of the pth pitch that can be expected to be found in the spectrum and f_s is the sampling frequency. In practice R^p us set to a maximum of 10. The strongest pitch is then assigned to the kth atom

$$\hat{p}^k = \arg\max_p S_k^p. \tag{14}$$

3.3.2. Pitch Salience

Similar to [19], the pitch salience matrix $\mathbf{X}$ is calculated by gathering the contributions of a collection of atoms $\mathbf{W}(p)$, and their activations $\mathbf{H}(p)$ that share the pth pitch label

$$x_{p,n} = \|\mathbf{W}(p)\mathbf{H}_n(p)\|_2 \tag{15}$$

Thresholding is applied to $\mathbf{X}$ [19] [10] with a thresholding parameter, δ applied to the maximum value of the pitch salience matrix in order to determine the threshold

$$\lambda = \delta \times \max_{p,n} [X]_{p,n} \tag{16}$$

with all elements of $\mathbf{X}$ with values less than λ set to zero.

3.4. Relationship to Prior Work

NMF-HSC approach bears a similarity to other NMF, or Nonnegative Sparse Dictionary Learning (SDL) [9] approaches utilising stepwise methods, particularly the NMF-ℓ_0, [8]. A subtle difference in the algorithm flow is found in the fact that NMF-HSC performs better when NMF, with random initialisation, is performed prior to the initial sparse coding, while the SDL approaches stipulate sparse coding as the initial step. The most obvious difference, to NMF-ℓ_0 [8] is the use of the Hellinger distance cost function, both in active set selection (HSC) and in NMF gradient steps. We have recently found HSC to outperform ASNA and OMP in the context of AMT experiments using fixed dictionaries [14]. HSC is similar to the ASNA algorithm [13], which uses the KL-divergence, employing a Newton-based projection step after each active set addition. A notable modification relative to ASNA is the use of a nearest neighbour-based selection approach (with a modification for negative coefficients in the residual) rather than a minimum gradient selection.

4. EVALUATION

AMT experiments were performed to compare the proposed method, NMF-HSC, with standard multiplicative update based NMF using both the KL and Hellinger cost functions. A standard dataset consisting of the first $30s$ of 30 classical piano pieces, recorded live on a Disklavier piano, from the MAPS database [20] was used. Each piece was transformed into a time-frequency representation using ERB transforms, using code supplied by the authors of [10]. ERBTs of varying dimensions were used, similar to [21]; two ERBTs with a sampling frequency of $22.05kHz$ and dimensions $M = \{250, 512\}$ respectively, and an ERBT with dimension $M = 1024$ using a sampling frequency of $44.1kHz$. For each transform a random dictionary of $K = 88$ atoms, the number of keys on a grand piano, was generated from a flat distribution. These dictionaries were used as the initial dictionary for all pieces and for all algorithms.

The proposed NMF-HSC was run for $J = 10$ iterations of the main loop each with $L = 50$ inner iterations of NMF. When HSC was employed for sparse approximation, the stopping condition was set to 11 iterations allowing selection of no more than 1/8 of atoms in the dictionary. The KL-NMF and H-NMF algorithms were run until convergence, which was typically seen to take of the order of 300 iterations. Pitch estimation was subsequently performed on all learnt atoms.

In order to measure the AMT performance related to the various approaches, a standard thresholding setup [10] [19] [21] is used. For a variety of values of $\delta \in \{15, ...45\}$, thresholding is performed on each piece as per (16). Comparison of the thresholded activation matrix and the ground truth activity, supplied with the MAPS dataset, is performed and true positives, $\mathcal{TP}$, false positives, $\mathcal{FP}$ and false negatives, $\mathcal{FN}$, are denoted, and the common pattern recognition measures

$$\mathcal{P} = \frac{|\mathcal{TP}|}{|\mathcal{TP}| + |\mathcal{FP}|}$$

$$\mathcal{R} = \frac{|\mathcal{TP}|}{|\mathcal{TP}| + |\mathcal{FN}|}$$

$$\mathcal{F} = \frac{2|\mathcal{TP}|}{2|\mathcal{TP}| + |\mathcal{FP}| + |\mathcal{FN}|}$$

are then derived. Recorded results for each approach relate the maximum $\mathcal{F}$-measure at one value of δ across all pieces and the corresponding precision, $\mathcal{P}$, and Recall, $\mathcal{R}$.

4.1. Results

The results for the experiments are given in Table 1. First it is seen that the performance for the H-NMF and KL-NMF are almost identical, validating the choice of the Hellinger distance for AMT tasks. A large difference is seen between the performance using the multiplicative NMF approaches across

M	Alg.	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$
250	KL-NMF	56.2	55.5	55.9
	H-NMF	56.5	55.2	55.9
	NMF-HSC	66.1	62.4	64.2
512	KL-NMF	59.7	60.8	60.2
	H-NMF	59.2	61.5	60.3
	NMF-HSC	69.0	64.0	66.4
1024	KL-NMF	60.8	65.1	62.9
	H-NMF	63.0	62.4	62.7
	NMF-HSC	67.0	66.0	66.5
	SS-β-NMF [10]	70.3	65.5	67.7

Table 1. Precision, Recall and $\mathcal{F}$-measures in AMT experiments for various ERB dimensions, M, using NMF with KL and Hellinger distances and the proposed NMF-HSC. Results also given for Hamonic NMF [10]

the various transforms, with the difference between the smallest and largest dimension ERBTs being of the order of 7%. We note that the corresponding improvement, using supervised NMF with fixed dictionaries, reported in [21] was of the order of 3%.

The proposed NMF-HSC method improves over the multiplicative update approaches by between 4% and 8% depending on the transform, with the largest improvements seen with the ERB of smallest dimension and worst performance. In doing this the difference between the different transforms is muted, with a maximum difference in $\mathcal{F}$-measure relative to transform of only 2.3%. The results for the Harmonic NMF method of Vincent et al [10] for the same dataset are also shown in Table 1 and this method is seen to perform better than NMF-HSC, however the difference is small, less than 1.5% and the proposed method is unsupervised.

5. CONCLUSIONS

A large improvement in AMT using unsupervised NMF has been presented through incorporation of a sparse coding step using a cost function that is generally overlooked in musical signal processing. The proposed approach was applied to a live piano dataset, with AMT performance close to harmonically constrained methods recorded. The use of transforms which have previously been observed to lead to less correlated dictionaries was also seen to enhance the performance. However, separability issues are still evident. Close inspection of the results reveals atoms containing energy from two, or possibly more, note spectra, often corresponding to notes that are played few times in a piece. A harmonic filtering postprocessing step may be introduced in order to further improve the AMT performance. Optimisation strategies other than the Newton method may also be explored. Furthermore, a coarse sparse coding strategy was used, selecting a fixed number of atoms. Other stopping criteria may be considered.

6. REFERENCES

[1] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems (NIPS 14)*, Denver, 2000, pp. 556–562.

[2] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, 2003, pp. 177–180.

[3] C. Fevotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, March 2009.

[4] N. Dobigeon and C. Fevotte, "Robust nonnegative matrix factorization for nonlinear unmixing of hyperspectral images," in *Proc. IEEE Workshop Hyperspectral image and signal processing: Evolution in remote sensing (WHISPERS)*, 2013.

[5] P. Paatero and U. Tapper, "Positive matrix factorisation: A non-negative factor model with optimal utilization of error," *Environmetrics*, vol. 5, pp. 111–126, 1994.

[6] A. Cichocki, S. Amari, R. Zdunek, R. Kompass, G. Hori, and Z. He, "Extended smart algorithms for non-negative matrix factorization," *Lecture notes in Artificial Intelligence*, vol. 4029, pp. 548–562, 2006.

[7] D. Donoho and V. Stodden, "When does non-negative matrix factorisation give a correct decomposition into parts?," in *Advances in Neural Information Processing Systems 16*, 2004.

[8] R. Peharz, M. Stark, and F. Pernkopf, "Sparse nonnegative matrix factorisation using ℓ_0 constraints," in *IEEE International Workshop on Machine Learning for Signal Processing*, Kittila, 2010, pp. 83–88.

[9] M. Aharon, M. Elad, and A. M. Bruckstein, "K-SVD and its non-negative variant for dictionary design," in *Proceedings of the SPIE conference (Wavelets XI)*, Baltimore, 2005, pp. 327–339.

[10] E. Vincent, N. Bertin, and R. Badeau, "Adaptive harmonic spectral decomposition for multiple pitch estimation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 3, pp. 528–537, March 2010.

[11] E. Benetos and S. Dixon, "A shift-invariant latent variable model for automatic music transcription," *Computer Music Journal*, vol. 36, no. 4, pp. 81–94, Winter 2012.

[12] S. Raczynski, N. Ono, and S. Sagayama, "Extending non-negative matrix factorisation - a discussion in the context of multiple frequency estimation of musical signals," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Glasgow, 2009, pp. 934–938.

[13] T. Virtanen, J. F. Gemmeke, and B. Raj, "Active set newton algorithm for overcomplete non-negative representations of audio," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 11, pp. 2277–2289, November 2013.

[14] K. O'Hanlon and M. D. Plumbley, "A greedy active set approach using hellinger distance applied to spectrogram decompositions," In preparation.

[15] A. Cichocki, R. Zdunek, and S. Amari, "Csiszar's divergences for non-negative matrix factorization: Family of new algorithms," in *6th International Conference on Independent Component Analysis and Blind Signal Separation (ICA)*, Florida, 2006, pp. 32–39.

[16] A. Cichocki and A. H. Phan, "Fast local algorithms for large scale non-negative matrix and tensor factorizations," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E92-A(3), pp. 708–721, March 2009.

[17] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proceedings of the 27th Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, 1993, vol. 1, pp. 40–44.

[18] K. O'Hanlon and M. D. Plumbley, "Structure-aware dictionary learning with harmonic atoms," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Barcelona, 2011, pp. 1761–1765.

[19] E. Vincent, N. Bertin, and R. Badeau, "Harmonic and inharmonic non-negative matrix factorisation for polyphonic music transcription," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas, 2008, pp. 109–112.

[20] V. Emiya, R. Badeau, and B. David, "Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle," *IEEE Transactions on Audio Speech and Language Processing*, vol. 18, no. 6, pp. 1643–1654, August 2010.

[21] K. O'Hanlon and M. D. Plumbley, "Row-weighted decompositions for automatic music transcription," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, 2013.